

# **Der Nutzen informetrischer Analysen und nicht-textueller Dokumentattribute für das Information Retrieval in digitalen Bibliotheken**

Philipp Schaer

26. Mai 2013

Vom Promotionsausschuss des Fachbereichs 4: Informatik der Universität Koblenz-Landau zur Verleihung des akademischen Grades Doktor der Naturwissenschaften (Dr. rer. nat.) genehmigte Dissertation.

Tag der wissenschaftlichen Aussprache: 15. Mai 2013

Vorsitzender des Promotionsausschusses: Prof. Dr. Rüdiger Grimm

Promotionskommission

Vorsitzender: Prof. Dr. Klaus Diller  
Berichterstatter: Prof. Dr. York Sure-Vetter  
Prof. Dr. Jürgen Krause

Veröffentlicht als Dissertation an der Universität Koblenz-Landau.



# Inhaltsverzeichnis

<b>Zusammenfassung</b>	<b>1</b>
<b>Abstract</b>	<b>3</b>
<b>Danksagung</b>	<b>5</b>
<b>1 Einleitung</b>	<b>7</b>
1.1 Zielsetzung . . . . .	9
1.2 Forschungsfragen . . . . .	11
1.3 Aufbau der Arbeit . . . . .	13
<b>I Grundlegende Modelle, Technologien und Konzepte</b>	<b>15</b>
<b>2 Digitale Bibliotheken</b>	<b>17</b>
2.1 Aufbau, Charakteristika und Begrifflichkeit . . . . .	17
2.2 Beispielsysteme . . . . .	19
2.2.1 Kommerzielle Systeme und Verlagskataloge . . . . .	20
2.2.2 Öffentliche Systeme und Metakataloge . . . . .	20
2.2.3 Open-Access-Repositoryen . . . . .	22
2.2.4 Suchmaschinen- und Web-Crawl-Systeme . . . . .	22
2.3 Polyzentrische Informationslandschaft . . . . .	23
2.4 Metadaten in digitalen Bibliotheken . . . . .	25
2.4.1 Zeitschriften . . . . .	27
2.4.2 Monografien . . . . .	28
2.4.3 Autorennamen . . . . .	30
2.4.4 Kontrollierte Schlagwörter und Thesauri . . . . .	31
2.5 Nutzerverhalten und -bedürfnisse . . . . .	31
2.5.1 Die Rolle digitaler Bibliotheken . . . . .	32
2.5.2 Relevanzbeurteilung durch Nutzer . . . . .	33
2.5.3 Disziplinarität digitaler Angebote . . . . .	35
2.6 Dokumentenretrieval . . . . .	35
2.7 Zusammenfassung . . . . .	38
<b>3 Anfragerweiterung und Recommendersysteme</b>	<b>39</b>
3.1 Sprachproblem des IR . . . . .	40
3.2 Polyrepräsentation . . . . .	42

3.3	Textuelle Anfrageerweiterung . . . . .	45
3.4	Nicht-textuelle Anfrageerweiterungen . . . . .	48
3.5	Zusammenfassung . . . . .	50
<b>4</b>	<b>Ranking und Relevanz</b>	<b>53</b>
4.1	Sortierung vs. Ranking . . . . .	53
4.2	Der Relevanzbegriff . . . . .	54
4.3	Textuelles Ranking . . . . .	58
4.3.1	Boolesches Modell . . . . .	58
4.3.2	Erweitertes boolesches Ranking . . . . .	59
4.3.3	Vektorraummodell . . . . .	60
4.3.4	Probabilistisches Modell . . . . .	61
4.3.5	Vergleich der Modelle . . . . .	62
4.4	Nicht-textuelles Ranking . . . . .	63
4.4.1	PageRank und HITS . . . . .	64
4.4.2	Google Scholar Citation-Rank . . . . .	65
4.4.3	Wissenschaftliches Ranking: Journal Impact Factor und h-Index . . . . .	66
4.4.4	Zentralität in Autorennetzwerken . . . . .	67
4.4.5	Web 2.0-Retrieval und Social-Ranking . . . . .	68
4.4.6	Sonstige Rankingverfahren . . . . .	70
4.5	Automatisiertes vs. interaktives Retrieval . . . . .	71
4.5.1	Historischer Hintergrund: Agentensysteme vs. direkte Manipulation . . . . .	73
4.5.2	Learning-to-Rank . . . . .	77
4.5.3	Human-Computer Information Retrieval . . . . .	79
4.6	Zusammenfassung . . . . .	80
<b>5</b>	<b>IR-Evaluation</b>	<b>83</b>
5.1	Das Cranfield-Paradigma . . . . .	83
5.2	Kennzahlen in der IR-Evaluation . . . . .	87
5.3	Vergleich von gerankten Listen . . . . .	91
5.4	Zusammenfassung . . . . .	92
<b>6</b>	<b>Informetrie</b>	<b>93</b>
6.1	Power-Law-Verteilungen . . . . .	97
6.1.1	Pareto-Verteilung . . . . .	97
6.1.2	Lotkas Gesetz . . . . .	99
6.1.3	Zipfs Gesetz . . . . .	101
6.1.4	Bradfords Gesetz . . . . .	103
6.1.5	Lotkaische Informetrie und der Informationsproduktionsprozess . . . . .	104
6.1.6	Anwendung bibliometrischer Gesetze . . . . .	106
6.1.7	Eigenschaften von Power-Laws . . . . .	107
6.1.8	Vergleich von empirisch ermittelten und formalen Power-Laws . . . . .	107
6.1.9	Ermittlung des Power-Law-Exponenten . . . . .	111

6.2	Kookkurrenzmodelle . . . . .	112
6.2.1	Jaccard-Index . . . . .	114
6.2.2	Likelihood-Ratio . . . . .	114
6.2.3	Machine-Learning und automatische Klassifikationssysteme . . . . .	115
6.3	Zusammenfassung . . . . .	116
<b>II</b>	<b>Verfahren und Materialien</b>	<b>117</b>
<b>7</b>	<b>Nicht-textuelle Mehrwertdienste für das Dokumentenretrieval</b>	<b>119</b>
7.1	Anfrageerweiterung . . . . .	119
7.1.1	Nicht-textuelle Anfrageerweiterung . . . . .	119
7.1.2	Pretest: Anfrageerweiterung mit Thesaurustermen . . . . .	120
7.1.3	Pretest: Anfrageerweiterung mit Autorennamen . . . . .	121
7.2	Alternatives Ranking – Reranking . . . . .	123
7.2.1	Ausgangslage: TF*IDF-Implementation der Suchmaschine Solr . . . . .	123
7.2.2	Erweiterung des Standardrankings . . . . .	125
7.2.3	Pretest: Vergleich von zonenbasierten und -losem Bradfordizing . . . . .	129
7.3	Zusammenfassung . . . . .	130
<b>8</b>	<b>Materialien</b>	<b>131</b>
8.1	IR-Testkollektionen . . . . .	131
8.1.1	Testkorpora . . . . .	131
8.1.2	Topics . . . . .	136
8.1.3	Relevanzurteile . . . . .	138
8.2	Ermittlung der Power-Law-Parameter in empirischen Daten . . . . .	141
8.3	Zusammenfassung . . . . .	143
<b>III</b>	<b>Auswertung und Diskussion</b>	<b>145</b>
<b>9</b>	<b>Auswertung der Experimente zur Anfrageerweiterung</b>	<b>147</b>
9.1	Aufbau des Experimentes für die Anfrageerweiterung . . . . .	147
9.2	Effekte der Anfrageerweiterung . . . . .	149
9.2.1	Einfluss der Kombination unterschiedlicher Dokumentattributen . . . . .	149
9.2.2	Einfluss der Art der Anfrageformulierung: Strukturierte vs. unstrukturierte Anfragen . . . . .	151
9.2.3	Einfluss der Dokumentfilter . . . . .	154
9.3	Zusammenfassung der Ergebnisse . . . . .	156
<b>10</b>	<b>Auswertung der Experimente zum Reranking</b>	<b>159</b>
10.1	Aufbau der Power-Law-Analyse und der Reranking-Experimente . . . . .	159
10.2	Power-Law-Verteilungen in GIRT und iSearch . . . . .	160
10.3	Effekte des Rerankings . . . . .	165
10.3.1	Reranking mit Autorennamen in GIRT . . . . .	165

10.3.2	Reranking mit Klassifikationscodes in GIRT . . . . .	166
10.3.3	Reranking mit ISSN-Codes in GIRT . . . . .	170
10.3.4	Reranking mit Ortsangaben in GIRT . . . . .	173
10.3.5	Reranking mit Forschungsmethoden in GIRT . . . . .	175
10.3.6	Reranking mit Verlagsnamen in GIRT . . . . .	177
10.3.7	Reranking nach Erscheinungsjahr in GIRT . . . . .	179
10.3.8	Reranking mit Thesaurustermen in GIRT . . . . .	181
10.3.9	Reranking mit Autorennamen in iSearch . . . . .	183
10.3.10	Reranking mit Publikationsquellen in iSearch . . . . .	185
10.4	Zusammenfassung der Ergebnisse . . . . .	187
<b>11</b>	<b>Diskussion</b>	<b>193</b>
11.1	Produktivität, Kookkurrenz und Relevanz . . . . .	193
11.2	Bekannte Einschränkungen der evaluierten Verfahren . . . . .	194
11.3	Die Konsequenz von informetrischer Analyse für das IR . . . . .	197
<b>12</b>	<b>Ausblick</b>	<b>205</b>
	<b>Literaturverzeichnis</b>	<b>209</b>
	<b>Abkürzungsverzeichnis</b>	<b>237</b>
	<b>Anhang</b>	<b>239</b>
<b>A</b>	<b>Ermittelte Entitäten der Kookkurrenzanalyse</b>	<b>241</b>
A.1	Ermittelte Entitäten für GIRT . . . . .	241
A.2	Ermittelte Entitäten für iSearch . . . . .	252
<b>B</b>	<b>Ermittelte Werte der Power-Law-Analyse</b>	<b>257</b>
B.1	Produktivität der Autoren in GIRT . . . . .	258
B.2	Produktivität der Klassifikationen in GIRT . . . . .	270
B.3	Produktivität der Zeitschriften in GIRT . . . . .	282
B.4	Produktivität des Erscheinungslandes in GIRT . . . . .	294
B.5	Produktivität der Forschungsmethode in GIRT . . . . .	306
B.6	Produktivität der Verlage in GIRT . . . . .	318
B.7	Produktivität des Erscheinungsjahres in GIRT . . . . .	330
B.8	Produktivität der Thesaurusterme in GIRT . . . . .	342
B.9	Produktivität der Autorennamen in iSearch . . . . .	354
B.10	Produktivität der Autorennamen in iSearch . . . . .	360
<b>C</b>	<b>Publikationen im Rahmen der Dissertation</b>	<b>367</b>
<b>D</b>	<b>Software im Rahmen der Dissertation</b>	<b>369</b>

**E Lebenslauf**

**371**





# Zusammenfassung

**Forschungshintergrund:** Die Suche nach wissenschaftlicher Literatur ist eine Forschungs-herausforderung für das Information Retrieval im besonderen Umfeld der digitalen Bibliotheken. Aktuelle Nutzerstudien zeigen, dass im klassischen IR-Modell zwei typische Schwächen auszumachen sind: das Ranking der gefundenen Dokumente und Probleme bei der Formulierung von Suchanfragen. Gleichzeitig ist zu sehen, dass traditionelle Retrievalsysteme, die primär textuelle Dokument- und Anfrage-merkmale nutzen, bei IR-Evaluationskampagnen wie TREC und CLEF in ihrer Leistung seit Jahren stagnieren.

**Forschungsmethode:** Zwei informatrisch-motivierte Verfahren zur Suchunterstützung werden vorgestellt und mittels einer Laborevaluation mit den beiden IR-Test-kollektionen GIRT und iSearch sowie 150 und 65 Topics evaluiert. Die Verfahren sind: (1) eine auf der Kookkurrenz von Dokumentattributen basierende Anfra-geerweiterung und (2) ein Rankingansatz, der informatrische Beobachtungen zur Produktivität von Informationserzeugern ausnutzt. Beide Verfahren wurden mit einer Referenzimplementierung auf Basis der Suchmaschine Solr verglichen.

**Ergebnisse:** Beide Verfahren zeigen positive Effekte beim Einsatz von zusätzlichen Doku-mentattributen wie Autorennamen, ISSN-Codes und kontrollierten Schlagwörtern. Bei der Anfrageerweiterung konnte ein positiver Effekt in Form einer Verbesserung der Precision (bpref +12%) und des Recall (R +22%) erzielt werden. Die alternati-ven Rankingansätze konnten beim Ansatz von Autorennamen und ISSN-Codes die Baseline erreichen bzw. diese beim Einsatz der kontrollierten Schlagwörter über-treffen (MAP +14%). Einen negativen Einfluss auf das Ranking hatten allerdings die Einbeziehung von Faktoren wie Verlagsnamen oder Erscheinungsorten. Für beide Verfahren konnte eine substantiell andere Sortierung der Ergebnismenge, gemessen anhand von Kendalls  $\tau$  ( $\leq 0,8$ ), beobachtet werden. Zusätzlich zu der verbesserten Relevanz der Ergebnisliste kann der Nutzer so eine neue Sicht auf die Dokumentenmenge gewinnen.

**Neuigkeitswert:** Die Anfrageerweiterung mit Autorennamen, ISSN-Codes und Thesau-rustermen zeigt das bisher ungenutzte Potential, das sich in digitalen Bibliotheken durch die Datenfülle und -qualität ergibt. Die Rankingverfahren konnten die Leis-tung des Baseline-Systems übertreffen, nachdem eine Überprüfung auf Vorliegen einer Power Law-Verteilung und eine anschließende Filterung durchgeführt wur-de. Dies zeigt, dass die Rankingverfahren nicht universell für alle Suchanfragen anwendbar sind, sondern ein Vorhandensein bestimmter Häufigkeitsverteilungen

## *Zusammenfassung*

voraussetzen. So wird die enge Verbindung der Verfahren zu informetrischen Gesetzmäßigkeiten wie Bradfords, Lotkas oder Zipfs Gesetz deutlich.

**Praktische Auswirkungen:** Die beiden in der Arbeit evaluierten Verfahren sind als interaktive Suchunterstützungsdienste in der sozialwissenschaftlichen digitalen Bibliothek Sowiport implementiert. Die Verfahren lassen sich über entsprechende Web-Schnittstellen auch in anderen Anwendungskontexten einsetzen.

# Abstract

**Purpose:** The search for scientific literature in scientific information systems is a discipline at the intersection between information retrieval and digital libraries. Recent user studies show two typical weaknesses of the classical IR model: ranking of retrieved and maybe relevant documents and the language problem during the query formulation phase. At the same time traditional retrieval systems that rely primarily on textual document and query features are stagnating for years, as it could be observed in IR evaluation campaigns such as TREC or CLEF. Therefore alternative approaches to surpass these two problem fields are needed.

**Methodology and approach:** Two different search support systems are presented in this work and evaluated with a lab evaluation using the IR test collection GIRT and iSearch with 150 and 65 topics, respectively. These two systems are (1) a query expansion that is based on the analysis of co-occurrences of document attributes and (2) a ranking mechanism that applies informetric analysis of the productivity of information producers in the information production process. Both systems were compared to a baseline system using the Solr search engine.

**Findings:** Both methods showed positive effects when applying additional document attributes like author names, ISSN codes and controlled terms. The query expansion showed an improvement in precision (bpref +12%) and in recall (R +22%). The alternative ranking methods were able to compete with the baseline for author names and ISSN codes and were able to beat the baseline by using controlled terms (MAP +14%). A clear negative influence was seen when using entities like publishers or locations. Both methods were able to generate a substantially different sorting of the result set, measured using Kendall's  $\tau$  ( $\leq 0.8$ ). So, additional to the improved relevance in the result list, the user can get a new and different view on the document set.

**Originality and value:** Query expansion using author names, ISSN codes and thesaurus terms showed great potential that lies within the rich metadata sets of digital library systems. The proposed ranking methods could outperform standard relevance ranking methods after they were filtered by the existence of a so-called power law. This showed that the proposed ranking methods cannot be used universally in any case but require specific frequency distributions in the metadata. A connection between the underlying informetric laws of Bradford, Lotka and Zipf is made clear.

**Practical implications:** The evaluated methods were implemented as interactive search supporting systems that can be used in an interactive prototype and the social

## *Abstract*

science digital library system Sowiport. Besides that, the methods are adaptable to other systems and environments using a free software framework and a web API.

# Danksagung

„Two parts of a dissertation are more frequently read than others: the acknowledgments and the reference list. These sections place the dissertation in its social and intellectual context, showing that a doctoral thesis— although single authored—is a collaborative effort.“ (Hammarfelt, 2012, S. 9)

Obwohl den Worten von Björn Hammarfelt wenig hinzuzufügen ist, möchte ich die Gelegenheit nutzen und den Menschen danken, die mich bei der Erstellung der Dissertation und der „gemeinsamen Anstrengung“ unterstützt haben.

Zunächst gilt mein Dank meinen beiden Doktorvätern York Sure-Vetter und Jürgen Krause. Herr Krause hat mich seit meiner Zeit an der Uni Koblenz als seine studentische Hilfskraft immer wieder menschlich wie wissenschaftlich inspiriert und in langen Gesprächen während meiner Zeit als sein Mitarbeiter in meinen Forschungsvorhaben bestärkt. York Sure-Vetter hat nach der Pensionierung von Herrn Krause dankenswerterweise die Rolle des Motivators fortgeführt. Beide haben mir in meinen Jahren bei GESIS bestmöglich den Rücken freigehalten und mich ermutigt, diese Dissertation erfolgreich zu beenden.

Philipp Mayr war über die gemeinsamen Jahre bei GESIS, in den Projektzeiten von IRM I und II sowie als mein interner Betreuer eine große Stütze, Inspirationsquelle und Mentor. Hierfür gilt ihm meine tiefe Dankbarkeit. Mit ihm und Peter Mutschke, dem ich für seine Beharrlichkeit und den Einsatz für meine Zeit nach der Abgabe dieser Dissertation danke, hatte ich die ersten umfangreichen Paper-Sessions („absolute Arbeitsstille“) an die ich mich gerne erinnere. Danken möchte ich Thomas Lücke und Wilko van Hoek, die den Löwenanteil an der Kookkurrenzanalyse programmiert haben sowie Oliver Hopt, der mich in einer langen Kaffee-Diskussion auf die entscheidende Idee zur Implementierung des Rerankings brachte. Julia Achenbach gilt mein Dank für ihre unermüdliche Arbeit bei der Korrektur dieser Dissertation. Ihr ist zu verdanken, dass ich in der Endphase der Arbeit noch das Eine oder Andere über die deutsche Rechtschreibung gelernt habe. Mit Benjamin Zapilko, Daniel Hienert und Timo Wandhöfer, die gemeinsam mit mir in den letzten Zügen ihrer Dissertation lagen, verbindet mich nicht nur der gemeinsame Weg, sondern zusammen mit den anderen wunderbaren Kollegen der FuE, IPS und späteren WTS eine Menge Lebensfreude, die mich gerne auf die vergangenen Jahre zurückblicken lässt.

Meine Familie und enge Freunde haben mich in den letzten Monaten leider weniger gesehen und gehört, als mir selber lieb gewesen ist. Umso mehr habe ich mich über ihren stetigen Zuspruch und die moralische Unterstützung gefreut. Janine hat dabei eine besondere Rolle gespielt und für ihr unermüdliches Verständnis und ihren Zuspruch zu allen Zeiten gilt ihr mein ganz besonderer Dank.



# 1 Einleitung

Die Suche in wissenschaftlichen Informationssystemen ist eine Herausforderung mit der sich der Fachbereich des Information Retrieval (IR) im besonderen Umfeld der digitalen Bibliotheken (*Digital Libraries*, DL) seit vielen Jahren beschäftigt. Bereits grundlegende Papiere der IR-Evaluation wie die frühen Cranfield-Experimente (Cleverdon, 1960) führten ihre Untersuchungen vor dem Hintergrund des wissenschaftlichen Dokumentenretrievals durch. Auch heutzutage ist die Suche in digitalen Bibliotheken ein breites Forschungsfeld mit internationalen Konferenzen wie z. B. der ACM/IEEE Joint Conference on Digital Libraries (JCDL) oder der Theories and Practices in Digital Libraries (TPDL), die immer auch einen starken IR-spezifischen Hintergrund haben. Auch in Zeiten des World Wide Web, der sozialen Netzwerke und von Multimedia-Datenbanken wie YouTube oder Flickr und deren im Vergleich zur Literatursuche besonderen Anforderungen, scheint es immer noch Innovationspotential und -bedarf im Bereich des klassischen Dokumentenretrievals in digitalen Bibliotheken zu geben.

Das Bedürfnis nach Innovation in dieser klassischen Ausprägung des IR zeigen sowohl die stagnierenden Systemleistungen in kontrollierten Evaluationskampagnen wie TREC und CLEF als auch Nutzerstudien in realen Systemen der wissenschaftlichen Fachinformation und der digitalen Bibliotheken. Trotz großer Datenmengen in hochspezialisierten Fachdatenbanken wie Solis oder FIS Bildung, datenbankübergreifenden digitalen Bibliotheken wie Sowiport oder der ACM Digital Library und Websuchmaschinen wie Google Scholar oder CiteseerX (s. Abschnitt 2.2), bleiben viele Nutzerbedürfnisse und -anforderungen unbefriedigt. Zentrales Anliegen ist es, sowohl relevante als auch qualitative Dokumente zu finden – nach Möglichkeit direkt auf der ersten Ergebnisseite. Die Frage, was Relevanz jeweils ausmacht, ist dabei vielschichtig (s. Abschnitt 2.5).

Als Untersuchungsumgebung wird bewusst das Umfeld der digitalen Bibliotheken mit den beiden Beispielen Sowiport und arXiv ausgewählt. Beide Systeme stellen zwei unterschiedliche Ausprägungen der digitalen Bibliotheken dar und folgen gegensätzlichen Strategien bei der Akquise und Pflege der Datenbestände, der Informationsdarstellung und der internen Struktur der Daten. Trotzdem bewegen sich diese beiden Systeme zwischen stark strukturierten und monolithischen Fachdatenbanken wie SOLIS oder FIS Bildung und den weitgehend unstrukturierten Websuchmaschinen (Fox u. a., 2012). Sowohl Sowiport als auch arXiv unterscheiden sich in ihrem Aufbau und ihrem Grad der Strukturiertheit und vereinen somit unterschiedliche Dokumentbestände und Qualitätsstandards (s. Abschnitt 2.1). Die Daten der beiden Systeme sind zudem Teil zweier IR-Testkollektionen, die für eine spätere Evaluierung genutzt werden können (s. Abschnitt 8.1.1). Die Auswahl dieser zwei Systeme stellt somit einen guten Kompromiss zwischen pragmatischen und organisatorischen Gesichtspunkten dar, da so die zu evaluierenden Systeme in unterschiedlichen Kontexten und Datenqualitätsstufen analysiert werden.

## 1 Einleitung

Ausgangspunkt für die Überlegungen in dieser Arbeit ist die Erkenntnis, dass die Retrievalleistung und die Nutzungszufriedenheit in klassischen digitalen Bibliothekssystemen seit Jahren stagniert, wie Metastudien auf Grundlage der Ergebnisse großer IR-Evaluationskampagnen (s. Kap. 5) in den letzten Jahren gezeigt haben (Armstrong u. a., 2009). Es wurde festgestellt, dass das klassische Ad-hoc-Retrieval auf einem hohen Leistungsniveau angekommen sei und sich die erzielten Retrievalleistungen nicht mehr signifikant verbessern. Die evaluierten Retrievalsysteme konnten im Vergleich zu ihren Vorgängerversionen keine substantiellen Verbesserungen in den Retrievalleistungen erbringen (Trotman u. Keeler, 2011). Ein Grund hierfür mag im IR-Modell und der in der IR-Community verbreiteten Evaluationmethodik des Cranfield-Paradigmas selbst liegen (s. Abschnitt 5.1). Gegen die These, dass es sich um ein Problem der Evaluationmethodik handelt, sprechen parallele Untersuchungsergebnisse aus Nutzerstudien im Bereich der DL, die zu ähnlichen Ergebnissen gelangen (te Boekhorst u. a., 2003; Xie, 2008).

Im klassischen IR-Modell wird von einem Abgleich zwischen dem Informationsbedürfnis eines Informationssuchenden und den in einem Informationssystem gespeicherten Dokumenten ausgegangen (s. Abb. 1.1). Üblicherweise wird das Informationsbedürfnis durch Suchterme formuliert und der Datenbestand liegt in Form von indexierten Dokumentsurrogaten vor. Beide werden danach mittels einer Ähnlichkeitsfunktion verglichen. Innerhalb dieses sehr grundlegenden Modells beschreibt Mandl (2006, S. 53) die verschiedenen Optimierungsansätze, die in der IR-Forschung bekannt sind. Er nennt eine Vielzahl an bekannten Ansätzen wie z. B. unterschiedliche Formen des Relevance-Feedbacks, die Optimierung von Gewichtungsfunktionen oder die Extraktion von Dokument- und Korpora-Eigenschaften, die dem Ziel dienen sollen, eine möglichst optimal sortierte Ergebnisliste zu erzeugen. Analog hierzu wurde in Studien von Buckley (2009) beobachtet, dass typische IR-Systeme seit Jahren mit vergleichbaren Problemen kämpfen, dem mit traditionellen, sprachzentrierten Mitteln des IR nicht Einhalt geboten werden konnte (s. Abschnitt 2.6).

Um die Beschränkungen des sprach- bzw. term-basierten Retrievals und dessen Schwächen im IR-Modell zu umgehen, wurden unterschiedliche Lösungsansätze präsentiert, die explizit auf nicht-textuelle Attribute oder Verfahren setzen. Eines der bekanntesten Verfahren, das in diese Kategorie fällt, ist das PageRank-Verfahren, das in der Websuchmaschine Google eingesetzt wird und auf einer Analyse der Verlinkungen von Webseiten basiert. Einen ähnlichen Ansatz verfolgt z. B. auch Mandl (2006, S. 253), indem er eine Erweiterung des IR-Modells um den Qualitätsbegriff vorschlägt. Dies soll zu einem „Qualitätsretrieval“ im Web-Retrieval führen. Einen ähnlichen Ansatz verfolgen Autoren wie z. B. Ingwersen (2012c) oder White (2007a,b), die auf die Gemeinsamkeiten von Informatik als Disziplin zur Messung von (meist wissenschaftlicher) Kommunikation und Informationseinheiten auf der einen und dem IR auf der anderen Seite hinweisen. Ein vertieftes Wissen über die Art und Weise wie Autoren, Verlage und sonstige Akteure im Prozess der Informationsproduktion miteinander verbunden sind (s. Abb. 1.1 links oben), könne im späteren Verlauf des Retrievalprozesses von großem Nutzen sein. Eine Übersicht und Gegenüberstellung von textuellen und nicht-textuellen State-of-the-Art-Verfahren im IR ist in den Kapiteln 3 und 4 zusammengetragen sowie eine Übersicht von Verfahren der Informatik in Kapitel 6.



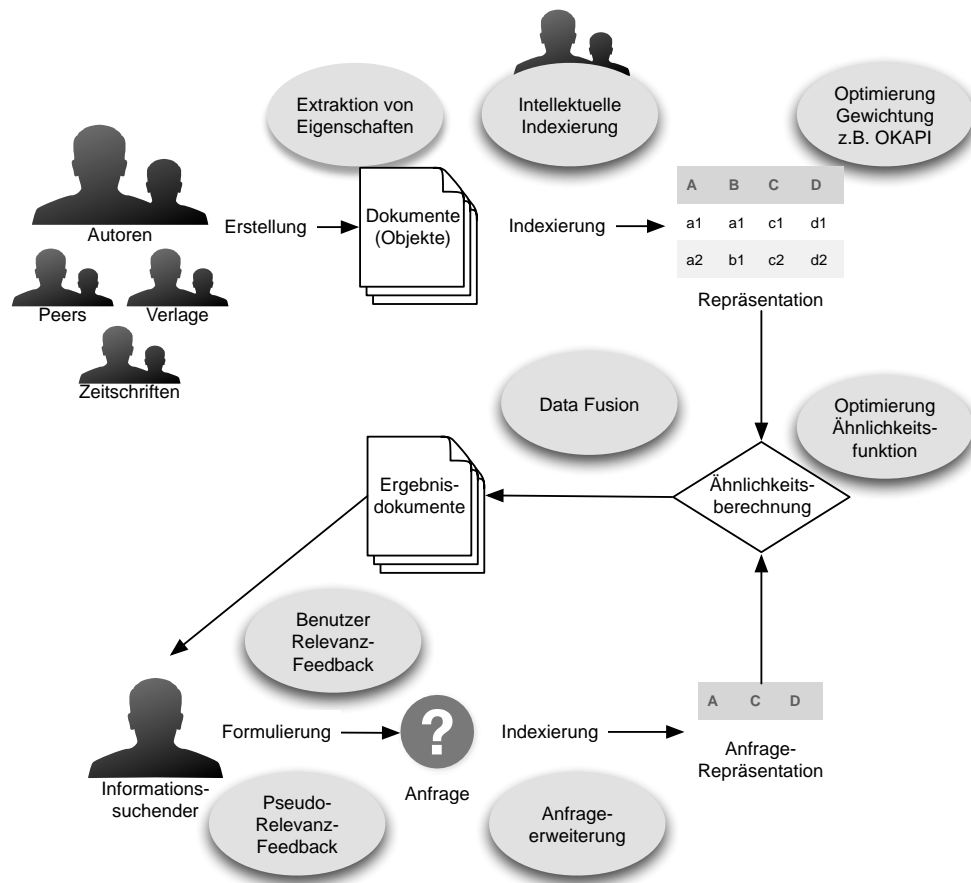


Abbildung 1.1: Ansätze zur Optimierung im klassischen IR-Modell, das auf den Abgleich von Anfrage- und Dokumentenrepräsentationen aufbaut (Abbildung basiert auf Mandl, 2006, S. 53).

## 1.1 Zielsetzung

Ausgangspunkt für diese Arbeit ist die Beobachtung, dass im klassischen, dokumentenzentrierten IR offene Probleme existieren, denen nicht mit traditionellen Lösungsansätzen beizukommen ist. Gleichzeitig ist aber im Umfeld der digitalen Bibliotheken eine Vielzahl an hochwertigen (Meta-)Daten vorhanden, die für das Retrieval genutzt werden können. Im Gegensatz zur Suche mit Websuchmaschinen sind hier die Inhalte kontrolliert und von einer höheren Datenqualität (s. Abschnitt 2.4). Im Bereich der Informatik existiert eine Reihe an Verfahren zur Analyse solcher Metadaten, allerdings bleibt eine konkrete Nutzung dieser hohen Informationsdichte für das Dokumentenretrieval meist aus. Im Fokus dieser Arbeit liegt daher die Fragestellung, wie die beiden Bereiche Informatik und IR angenähert und informatische Verfahren zum praktischen Einsatz im IR gebracht werden können. Hierzu sollen speziell nicht-textuelle Verfahren und Dokumentattribute verwendet werden.

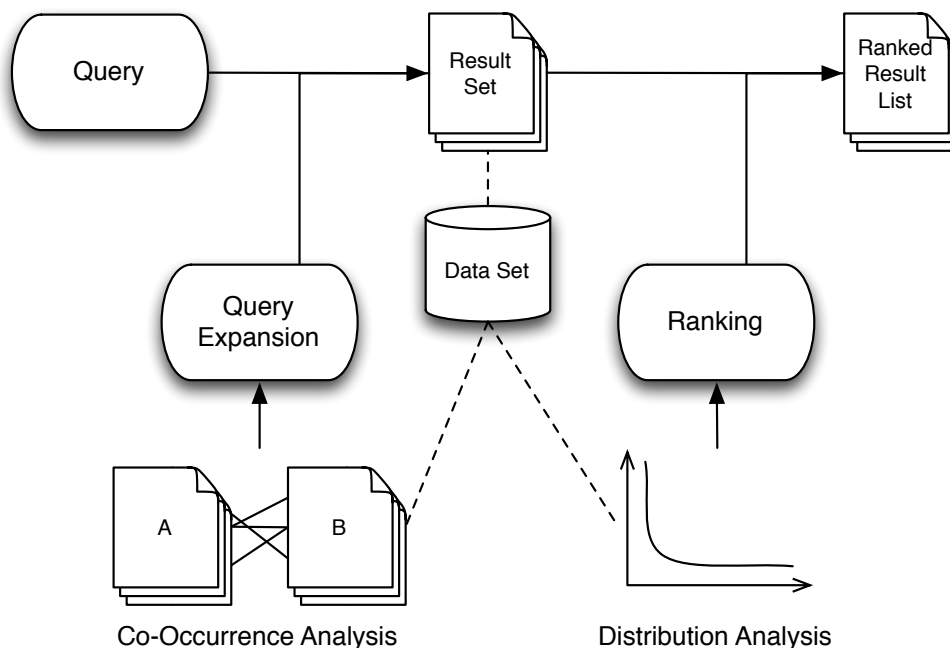


Abbildung 1.2: Systemtische Darstellung der beiden in dieser Arbeit verwendeten Verfahren im Retrievalprozess.

Bei der Analyse des klassischen IR-Modells zeigen sich zwei Stellen im Retrievalprozess, die sich für eine solche Herangehensweise anbieten: (1) der Moment der Anfrageformulierung und (2) die Sortierung der Ergebnismenge. Abbildung 1.2 zeigt die beiden Verfahren in einer systematischen Darstellung und deren Bedeutung für den Suchprozess. Ausgehend von einer Anfrage (*Query*) wird eine Dokumentmenge (*Result Set*) extrahiert, die anschließend von einem Rankingverfahren geordnet wird. Die beiden Verfahren Anfrageerweiterung (*Query Expansion*) und (Re-)Ranking greifen in diesen Prozess ein, um die endgültige Ergebnismenge (*Ranked Result List*) positiv zu beeinflussen.

Zwei konkrete informetrische Verfahren sollen in dieser Arbeit für das Dokumentenretrieval evaluiert werden: (1) die Analyse von gleichzeitig auftretenden Informationseinheiten in Dokumentbeständen (Kookkurrenzanalyse, *Co-occurrence Analysis*) und (2) die Analyse der Produktivität unterschiedlicher Informationsquellen wie z. B. Autoren oder Zeitschriften. Diese beiden Verfahren erlauben es, mit den in einem Retrievalsystem vorhandenen Daten zur Laufzeit in den Retrievalprozess einzugreifen. Mittels der Kookkurrenzanalyse können semantisch nahestehende (d. h. häufig gemeinsam auftretende) Entitäten in einem Datenbestand gefunden werden. So ist es bspw. möglich, herauszufinden, welche Autoren häufig zu bestimmten Themen publizieren und diese Information für das Retrieval zu nutzen. Bei der Messung der Produktivität von einzelnen Informationserzeugern wird bei der Analyse der absoluten Häufigkeiten klar, dass diese einer charakteristischen Verteilung folgen. So erzeugen bspw. einige wenige Autoren zu bestimmten Themen eine große Anzahl an Papieren, während zum gleichen Thema eine Vielzahl anderer Autoren nur sehr

wenig publiziert haben. Diese schiefe Verteilung wird allgemein als Power-Law-Verteilung bezeichnet und lässt sich auch bei vielen anderen Häufigkeitsverteilungen finden.

Durch die Arbeiten an den Schnittpunkten der IR-Forschung und der Informatik soll erforscht werden, ob und in welcher Form sich informatische Verfahren für den praktischen Einsatz in einem Retrievalprozess eignen. Der Fokus liegt hierbei auf einem pragmatischen Ansatzpunkt, der sich für einen Einsatz im organisatorischen Umfeld der digitalen Bibliotheken eignet.

Auf den Forschungsgebieten des IR, der digitalen Bibliotheken und der Informatik soll diese Dissertation durch die folgenden Arbeiten beitragen:

1. durch eine Analyse der offenen Probleme beim Dokumentenretrieval in DL-Systemen, besonders in den Phasen der Anfrageformulierung und des Rankings,
2. durch einen Lösungsentwurf wie mit Hilfe informatischer Analyseverfahren diesen offenen Problemen begegnet werden kann und
3. durch eine Evaluation der vorgeschlagenen Lösungen mit Hilfe einer IR-Laborevaluation nach dem Cranfield-Paradigma.

Eine generelle Machbarkeit der grundlegenden Ideen wurde bereits im DFG-geförderten Projekt „Retrieval-Mehrwertdienste zur Weiterentwicklung wissenschaftlicher Fachportale“ (IRM, s. Mayr u. a., 2011b; Schaer u. a., 2011) gezeigt. Im IRM-Projekt wurden klassische Verfahren des Dokumentenretrievals um informatische Verfahren wie einem Reranking nach Autorenzentralität oder dem sogenannten Bradfordizing erweitert. Die vorliegende Arbeit setzt auf diese Vorarbeiten auf und entwickelte diese substantiell weiter, indem bspw. verschiedene Dokumentattribute miteinander verknüpft und unterschiedliche Einsatzszenarien untersucht werden. Für einen wissenschaftlichen Nachweis der Effektivität der Verfahren wird auf den stark formalisierten Rahmen des Cranfield-Paradigmas gesetzt (s. Kap. 5). Das Ziel ist es, die im folgenden Abschnitt formulierten Forschungsfragen zu beantworten.

## 1.2 Forschungsfragen

Ausgehend von den zuvor beschriebenen Problemen der wissenschaftlichen Literatursuche und den fehlenden Brücken zwischen den Disziplinen des IR und der anwendungsbezogenen Informatik, lautet die übergeordnete Forschungsfrage in dieser Arbeit wie folgt:

*Können informatische Analyseverfahren, insbesondere die Messung der Produktivität und die Kookkurrenzanalyse, genutzt werden, um das Retrievalergebnis in digitalen Bibliotheken gezielt zu verbessern?*

Die Forschungsfragen teilen sich dabei nach den beiden verwendeten Verfahren auf: der Kookkurrenzanalyse und der Messung der Produktivität von Informationserzeugern und der dort charakteristischen Power-Law-Verteilungen.

Der erste Forschungsfragenkomplex betrachtet den Einsatz der Kookkurrenzanalyse für den Einsatz als Werkzeug zur Anfrageerweiterung.

## 1 Einleitung

- 1. Forschungsfrage:** *Für welche Entitäten bzw. Kombinationen von Entitäten einer digitalen Bibliothek eignet sich die Kookkurrenzanalyse als Werkzeug für eine Anfragerweiterung?*
- 2. Forschungsfrage:** *Lassen sich positive Veränderungen in der Retrievalleistung gegenüber einer State-of-the-Art-Baseline im Rahmen einer Cranfield-Evaluation messen?*

Neben der Kookkurrenzanalyse ist die Analyse von Häufigkeitsverteilungen bzw. der Produktivität von Informationserzeugern ein weiteres Werkzeug der Informatik. Die charakteristische Power-Law-Verteilung, die in vielen Ausprägungen in digitalen Datenbeständen zu finden ist, ist bei textuellen Informationen in Form von Zipfs Gesetz in IR-Verfahren eingeflossen, z. B. bei der Erstellung von Stoppwortlisten. Bei nicht-textuellen Dokumentattributen wie der Produktivität von Autoren, Zeitschriften oder Verlagen sind vergleichbare Phänomene zu beobachten, doch werden diese Informationen bisher nicht gezielt für das Retrieval genutzt. Daher lauten die nächsten Forschungsfragen:

- 3. Forschungsfrage:** *Kann durch gezielten Einsatz von informatrischen Kennzahlen wie der Bestimmung einer Power-Law-Verteilung das Retrievalergebnis beeinflusst werden? Was für Auswirkungen hat der Einsatz auf die Ergebnismenge? Kann der Einfluss durch signifikante Veränderungen in Form einer Cranfield-Evaluation gemessen werden?*
- 4. Forschungsfrage:** *Gibt es messbare Unterschiede zwischen den unterschiedlichen Implementierungen? Welche Implementierung liefert die besten Ergebnisse?*
- 5. Forschungsfrage:** *Sind die Art der untersuchten Entitäten und die jeweils zugrundeliegenden wissenschaftlichen Modelle für die Güte des Rankings von Bedeutung? Gibt es eine Verbindung zu bekannten informatrischen Verteilungen wie den Gesetzmäßigkeiten nach Lotka oder Bradford?*

Dieser, wie auch der erste Fragenkomplex soll konkret mittels einer IR-Laborevaluation nach dem Cranfield-Paradigma bearbeitet werden. Die Güte des hier vorgestellten Retrievalverfahrens, unter besonderer Berücksichtigung von informatrischen Verteilungsinformationen, wird anhand der Retrievalleistung des darauf basierenden Systems untersucht. Die Forschungsfragen beziehen sich auf die generelle Machbarkeit des Rerankings mittels der ermittelten Produktivitätszahlen. Unterschiedliche Implementierungen und Versuchsaufbauten sind denkbar, daher werden die Kandidaten vorgestellt und mittels einer Cranfield-Evaluation untersucht. Dieser Fragenkomplex soll betrachten, ob die generelle Annahme, dass durch den gezielten Einsatz von Power-Law-Verteilungen das Ranking positiv beeinflusst und bisher ungenutzte Metadaten im Retrieval genutzt werden können, stimmig ist.

Für die IR-Evaluation wird der aus den internationalen Evaluationskampagnen TREC und CLEF bekannte Datensatz GIRT sowie der iSearch-Korpus verwendet. GIRT enthält primär Dokumente (Metadaten) aus den Sozialwissenschaften, iSearch enthält Dokumente (Metadaten und Volltexte) aus der Physik. Neben der reinen Retrievalleistung in den

Fragen 3 bis 5 sollen zusätzlich informatrische Analysen auf den vorhandenen Korpora durchgeführt werden, um die Frage 6 zu beantworten.

- 6. Forschungsfrage:** *Sind Power-Law-Verteilungen direkt für gutes Ranking verantwortlich? Ist ein „sauberes“ Power-Law verantwortlich für eine bessere Retrievalleistung? Kann man durch Ermittlung der Exponenten oder einer anderen Kennzahl mit Bestimmtheit sagen, ob eine Power-Law-Verteilung geeignet ist, um damit zu ranken? Kann man im Vorfeld berechnen, ob sich ein Reranking lohnt?*

In der letzten Forschungsfrage sollen die speziellen Eigenarten von Power-Law-Verteilungen untersucht werden. Hierbei sind ggf. vorhandene Korrelationen zwischen der Effektivität des Rankings und den Charakteristika der Verteilungen von besonderem Interesse. Sollte die Art der Verteilung eine nachweisbare Auswirkung auf die Retrievalleistung haben, wäre dies ein starker Anhaltspunkt für die Gültigkeit der zugrundeliegenden Modelle und eine Bestätigung für den gezielten Einsatz im IR.

## 1.3 Aufbau der Arbeit

Die Arbeit teilt sich in vier Teile auf: (1) einen Grundlagenteil, (2) einen Teil für die verwendeten Verfahren und Materialien, (3) einen Teil zur Beschreibung und Diskussion der Evaluationsergebnisse und (4) einen Anhang.

Der Grundlagenteil I besteht aus den Kapiteln 2 bis 6. In Kapitel 2 wird das Einsatzszenario der digitalen Bibliotheken vorgestellt und die besonderen Nutzerbedürfnisse analysiert. Die folgenden beiden Kapitel beschäftigen sich mit der Thematik der Anfrageerweiterung und des Rankings im Dokumentenretrieval. Beides sind zentrale Schwachstelle im klassischen IR-Modell. Kapitel 5 fasst den aktuellen Stand der Forschung im Bereich der IR-Laborevaluation, das sogenannte Cranfield-Paradigma, zusammen. Bekannte Analyseverfahren der Informatris, mit Fokus auf die Konkurrenz- und die Produktivitätsanalyse, werden in Kapitel 6 besprochen. So wird in diesem Grundlagenteil der Bogen gespannt von der Analyse der Schwachstellen bestehender Systeme, über die Beschreibung des aktuellen Stands der Forschung und den Ausblick auf mögliche Lösungsstrategien. Ein besonderer Fokus liegt dabei auf der Unterscheidung zwischen bekannten Technologien, die auf textuelle Analysen setzen und solchen, die mit nicht-textuellen Attributen arbeiten.

In Teil II werden die verwendeten Verfahren und Materialien vorgestellt. In Kapitel 7 werden die beiden nicht-textuellen Mehrwertdienste für die Anfrageerweiterung und das Reranking beschrieben. Beide Verfahren versuchen, mit den Ergebnissen von informatrischen Analysen und den so gewonnenen nicht-textuellen Attributen Einfluss auf das Verhalten von IR-Systemen zu nehmen. Die zur Evaluation der Systeme verwendeten Testkollektionen werden in Kapitel 8 ausführlich beschrieben.

Der inhaltlich abschließende Teil III fasst die Ergebnisse der Evaluation zusammen und analysiert und diskutiert die Ergebnisse. In den beiden Kapiteln 9 und 10 werden jeweils die Anfrageerweiterung und das Reranking analysiert und die Ergebnisse in Bezug auf die eingangs gestellten Forschungsfragen diskutiert. Die Arbeit schließt mit einer

## 1 Einleitung

ausführlichen Diskussion der Ergebnisse in Kapitel 11 und einer Zusammenfassung der Forschungsbeiträge dieser Arbeit sowie mit einem Ausblick auf zukünftige Arbeiten in Kapitel 12.

Im Anhang der Arbeit finden sich darüber hinaus noch eine detaillierte Aufstellung aller Einzelergebnisse der informatrischen Analysen aus Kapitel 10, eine Übersicht über die im Rahmen dieser Dissertation erstellten Publikationen und Software sowie ein Lebenslauf.

### Textkonventionen

Folgende Konventionen wurden bei der Erstellung des Textes eingehalten, um den Lesefluss zu fördern:

- Solange es einen adäquaten deutschen, bzw. eingedeutschten, Begriff gibt, wird dieser verwendet. So wird in dieser Arbeit der Begriff Vektorraummodell, anstelle des *vector space model*, benutzen. Trotzdem sprechen wir von Retrieval und Ranking, da beide gebräuchliche Fachbegriffe sind. Bei der ersten Verwendung des Begriffs wird der jeweilige englische Originalbegriff in Klammern und kursiv angegeben, um Unklarheiten zu vermeiden.
- Sollte es keine gebräuchliche Alternative geben (z. B. bei Begriffen wie *Exact-Match*, *Best-Match* oder *Bag-of-Words*), wird der englische Begriff bei der ersten Verwendung kursiv markiert und später im normalen Fließtext als Substantiv verwendet, analog zu den Empfehlungen des Dudenverlags<sup>1</sup>.
- Programmcode, Pseudocode oder Auszüge aus der Anfragesprache werden in **Maschinenschrift** gesetzt.
- Aus Gründen der Lesbarkeit wird grundsätzlich nur die männliche Form angegeben. Begriffe wie Benutzer, Autor oder Herausgeber sind in dieser Arbeit explizit geschlechtsunabhängig zu verstehen.

---

<sup>1</sup><http://www.duden.de/sprachwissen/rechtschreibregeln/fremdwoerter>

## **Teil I**

# **Grundlegende Modelle, Technologien und Konzepte**





## 2 Digitale Bibliotheken

Im Laufe der letzten Dekade hat sich sowohl beim wissenschaftliche Nachwuchs als auch bei etablierten Wissenschaftlern eine große Veränderung im Informationsverhalten vollzogen. So ist bspw. bei der Suche nach wissenschaftlicher Literatur eine sogenannte „Now or Never“-Mentalität zu beobachten. Dies mag mit elektronischen Zugängen zu wissenschaftlichen Dokumenten in Form von Websuchmaschinen oder digitalen Bibliotheken, zusätzlich zu den etablierten physischen Bibliotheken an Universitäten oder anderen Forschungseinrichtungen, zu tun haben (mehr dazu in Abschnitt 2.5). Im folgenden Kapitel werden sowohl der typische Aufbau und die Charakteristika digitaler Bibliotheken als auch der aktuelle Stand der Forschung zum Nutzungsverhalten dieser Systeme beschrieben. Aus der Analyse der Stärken und Schwächen der Systeme als auch den Bedürfnissen der Nutzer lassen sich im weiteren Verlauf der Arbeit die konkreten Anforderungen und Forschungsfragen ableiten.

### 2.1 Aufbau, Charakteristika und Begrifflichkeit

Eine klassische Definition digitaler Bibliotheken geht auf die Digital Library Federation (1998) zurück:

„Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities.“ (Digital Library Federation, 1998)

Diese klassische Definition von 1998 ist breit aufgestellt und spiegelt das Definitionsproblem der frühen DL-Community wieder, die im späteren Verlauf als zu beliebig angeprangert wurde. In einer späteren Definition von Arms (2000) ist die Definition bereits knapper und präziser formuliert:

„[...] a digital library is a managed collection of information, with associated services, where the information is stored in digital formats and accessible over a network.“ (Arms, 2000, S. 2)

Weiterhin unterstreicht der Autor seine eigene Aussage: „A key part of this definition is that the information is managed“. Das ausschlaggebende Attribut ist für Arms, dass eine digitale Kollektion (intellektuell) verwaltet und gepflegt wird. Dies würde einer reinen Internetsuchmaschine die Bezeichnung als digitale Bibliothek absprechen – ihr

## 2 Digitale Bibliotheken

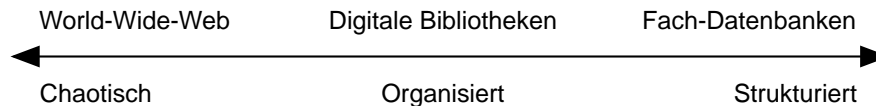


Abbildung 2.1: Unterschiedliche Strukturgrade von Schlüsseltechnologien der Informationsversorgung (angelehnt an Fox u. a., 2012, S. 6).

Index ist nicht im klassischen Sinne strukturiert bzw. wird nicht verwaltet. Dieses Unterscheidungsmerkmal zu Internetsuchmaschinen ist auch bei Lagoze u. a. (2005) ein entscheidendes Attribut:

„Digital libraries need to distinguish themselves from web search engines in the manner that they add value to web resources. This added value consists of establishing context around those resources, enriching them with new information and relationships that express the usage patterns and knowledge of the library community.“ (Lagoze u. a., 2005)

Saracevic (2009) nennt daher auch andere virtuelle Bibliotheksangebote als Beispiele, wie die rein digital existierende WWW Virtual Library<sup>1</sup>, die von Tim Berners-Lee ins Leben gerufen wurde, sowie die seit 1985 bestehende themenorientierte Perseus Digital Library<sup>2</sup>, eine Kollektion von historischen Texten und Bildern, die die Zeitspanne zwischen Antike und dem 19. Jahrhundert abdeckt. Ein anderes Beispiel aus der Informatik und angrenzenden Disziplinen ist die ACM Digital Library (s. Abschnitt 2.2).

Fox u. a. (2012) grenzen digitale Bibliotheken klar von Fachdatenbanken und dem Web ab, da beides Pole eines Kontinuums seien (s. Abb. 2.1). Auch hier wird die Struktur als Unterscheidungsmerkmal angesehen:

„One dichotomy often posed about DLs is Managed vs. Comprehensive. Thus, a library is managed while the WWW is unmanaged (but closer to being comprehensive). [...] we generally use the term *structure*. We argue that DLs must be organized, thus having a moderate degree of structure.“ (Fox u. a., 2012, S. 6)

Die Verwendung des Begriffs digitale Bibliothek geht wahrscheinlich auf ein Arbeitspapier von Kahn u. Cerf (1988) zurück, in dem eine nationale Infrastruktur für ein sogenanntes Digital Library System skizziert wurde. Andere Begriffe, wie elektronische oder virtuelle Bibliothek wurden lange Zeit synonym verwendet, später dann aber in ihren Bedeutungen voneinander abgesetzt. Auch heute sind unterschiedliche Begriffe in Gebrauch, die gleichbedeutend eingesetzt werden, so z. B. digitales Archiv, Repositorium oder digitales Fachportal. Allgemein kann man die folgenden Ausprägungen in Referenzdatenbanken (*reference databases*) und Quelldatenbanken (*source databases*) unterscheiden (Chowdhury, 2010, S. 17), wobei die Referenzdatenbanken in drei Unterkategorien aufzuteilen sind:

<sup>1</sup><http://vlib.org/>

<sup>2</sup><http://www.perseus.tufts.edu>

**Bibliografische Datenbanken**, die Literaturnachweise, teilweise mit Zitationen und erweiterten bibliografischen Metadaten wie einem Abstract, enthalten.

**Bibliothekskataloge**, die physische Katalogsysteme von Bibliotheken nachbilden oder mehrere Bibliothekskataloge zusammenfassen. Diese Systeme enthalten meist nur sehr wenige zusätzliche Metadaten.

**Verweisdatenbanken**, die Verweise auf Referenzen zu Informationen wie Name, Adresse oder Forschungsbereiche einer Person, Institution oder einem Informationssystem enthalten.

In Referenzdatenbanken werden nur Verweise auf die eigentlichen Informationsquellen verwaltet. Dies können Monografien, Zeitschriftenaufsätze, Web-Ressourcen, Personenverzeichnisse und vieles mehr sein. Dem entgegengesetzt sind die Quelldatenbanken, die den Nutzer nicht zu einer weiteren Informationsquelle weiterleiten, wie die Referenzdatenbank, da sie selbst die benötigten Informationen enthalten. Chowdhury unterteilt diese abhängig von ihrem Inhalt in:

**Numerische Datenbanken**, die numerische Datensätze unterschiedlichster Form beinhalten, so z. B. Statistiken oder Umfragedatensätze.

**Volltext-Datenbanken**, die Volltexte von Dokumenten, gleich welcher Form, enthalten, auch Volltext-Repositoryum genannt.

**Multimedia-Datenbanken**, in denen Texte, Bilder, Audio- und Video-Informationen gespeichert sind.

Darüber hinaus sind weitere fachspezifische Datenbanktypen wie Fachdatenbanken, virtuelle Fachbibliotheken oder Fachportale bekannt, die meist eine Mischform der o. g. Referenz- und Quelldatenbanken sind. Diese entstanden aus den Strukturen der Informationsversorgung durch Fachinformationszentren, kommerziellen Fachinformationsanbietern und Bibliotheken.

Allen genannten Bezeichnungen ist gemein, dass die Bezeichnungen in unterschiedlichen Fachkontexten und Disziplinen sehr heterogen verwendet werden. Dies führt in einem interdisziplinären Umfeld zu einer weitgehenden Rückbesinnung auf den Begriff der digitalen Bibliothek. Diese ist nicht immer korrekt, da sich ein Repositoryum natürlich von einem Fachportal in Aufbau und Inhalt unterscheidet. Allerdings wird die digitale Bibliothek als begriffliche Klammer bzw. Oberbegriff für die vielen unterschiedlichen und teilweise spezifischeren Bezeichnungen verwendet. Es ist davon auszugehen, dass die gegenwärtig fließende Bedeutung einem weiteren Wandel unterliegt und eine „Ausweitung des Begriffs auf digitalen Content gleich welcher Art“ (Seadle, 2009, S. 216 f.) zu erwarten ist.

## 2.2 Beispielsysteme

Im folgenden Abschnitt werden einige aktuelle DL-Systeme für wissenschaftliche Literatur beispielhaft vorgestellt. Die Beispiele sind grob in vier Untergruppen zusammengefasst: (1)

kommerzielle Systeme und Kataloge, (2) öffentliche Systeme und Metakataloge, (3) Open-Access-Repositoryn und (4) Suchmaschinen- und Web-Crawl-Systeme. Eine tabellarische Übersicht über die jeweiligen Systeme, ihre Betreiberorganisationen, die Anzahl der verfügbaren Nachweise und Volltexte sowie des inhaltlichen Fokus ist der Tabelle 2.1 zu entnehmen. Die Auswahl enthält sowohl nationale wie internationale Beispielsysteme, kann aber keinen Anspruch auf Vollständigkeit erheben. Die typischen System- und Anbietergruppen sind jedoch enthalten.

### 2.2.1 Kommerzielle Systeme und Verlagskataloge

Die Systeme ScienceDirect, SpringerLink, ACM Digital Library und das Web of Science sind typische Vertreter kommerzieller Systeme. Hinter allen Systemen stehen Unternehmen bzw. Verlage, die gegen ein entsprechendes Entgelt Nutzern Zugriff auf ihre Systeme gewähren. Bei den ersten drei genannten Systemen sind jeweils die Zugänge zu den erfassten Literaturnachweisen und einem Großteil der Metadaten kostenfrei möglich. Für zahlende Kunden wird darüber hinaus ein erweiterter Zugriff u. a. auf Volltexte, Zitationsdaten und weitere Mehrwertdienste gewährt. Diese Systeme sind also sowohl Nachweis- als auch Volltextdatenbanken, die sich zu einem großen Teil aus den Kataloginformationen der jeweiligen Betreiber speisen. Darüber hinaus werden durch Kooperationsverträge mit anderen Anbietern die Nachweissysteme mit externen Informationen erweitert. Im Gegensatz zu den drei vorgenannten Systemen ist das Web of Science eine reine Nachweisdatenbank, die allerdings einen starken Fokus auf eine breite fachliche Abdeckung und eine Erfassung von Zitationen und Referenzen legt.

Die Qualität der angebotenen Systeme ist für gewöhnlich sehr hoch, da die hinterlegten Nachweisinformationen entweder direkt aus den Katalogen der Verlage stammen, eingekauft oder mit großem Ressourcenaufwand selbst erstellt werden. Der kommerzielle Charakter der Systeme erlaubt eine wissenschaftliche Auswertung meist nur über Kooperationsverträge bzw. gegen ein entsprechendes Entgelt.

### 2.2.2 Öffentliche Systeme und Metakataloge

Eine Alternative zu den kommerziellen Systemen stellen die mit öffentlichen Mitteln geförderten Systeme Sowiport, Europeana Libraries, RePEc, DBLP oder FIS Bildung dar. Auch hier liegt der Fokus auf der Erfassung von Literaturnachweisen, obwohl auch vermehrt Volltexte bzw. Verlinkungen zu Volltexten in die Systeme aufgenommen werden. Die vier Systeme Sowiport, RePEc, DBLP und FIS Bildung besitzen im Gegensatz zu Europeana Libraries einen engen fachlichen Fokus. Für Sowiport sind dies die Sozialwissenschaften, für RePEc die Wirtschaftswissenschaften, für DBLP die Informatik und für FIS Bildung die Pädagogik. Europeana Libraries dient als Sammelstellen für wissenschaftliche Literatur für die Europeana und kooperiert mit wissenschaftlichen Bibliotheken in ganz Europa, allerdings ohne einen vergleichbaren inhaltlichen Fokus wie die zuvor genannten Systeme.

Ähnlich zu den kommerziellen Anbietern stehen die betreibenden Organisationen für eine hohe Qualität der angebotenen Daten ein. Über Kooperationsverträge oder Lizenzen

Name	Betreiber	Art	NW	VT	Domäne	Link
BASE	Universität Bielefeld	S	40918	40918	Interdisziplinär	<a href="http://www.base-search.net">http://www.base-search.net</a>
ScienceDirect	Elsevier	K	11467	11467	Interdisziplinär	<a href="http://www.sciencedirect.com">http://www.sciencedirect.com</a>
Google Scholar	Google	S	9470	2730	Interdisziplinär	<a href="http://scholar.google.com">http://scholar.google.com</a>
Sowiport	GESIS	P	7372	494	Sozialwissenschaft	<a href="http://www.gesis.org/sowiport">http://www.gesis.org/sowiport</a>
SpringerLink	Springer	K	6642	6642	Interdisziplinär	<a href="http://www.springerlink.com">http://www.springerlink.com</a>
Europeana Libraries	Europeana Foundation	P	5000	5000	Interdisziplinär	<a href="http://www.europeana-libraries.eu">http://www.europeana-libraries.eu</a>
Web of Science	Thomson Reuters	K	4968	-	Interdisziplinär	<a href="http://www.webofknowledge.com">http://www.webofknowledge.com</a>
DBLP	Uni Trier	P	2167	-	Informatik	<a href="http://dblp.uni-trier.de">http://dblp.uni-trier.de</a>
ACM Digital Library	ACM	K	2074	366	Informatik	<a href="http://dl.acm.org">http://dl.acm.org</a>
CiteseerX	Pennsylvania SU	S	1472	1472	Informatik	<a href="http://citeseerx.ist.psu.edu">http://citeseerx.ist.psu.edu</a>
RePEc	dezentral organisiert	P	1337	1200	Wirtschaft Physik,	<a href="http://www.repec.org">http://www.repec.org</a>
arXiv	Cornell University	O	812	812	Informatik, Mathematik	<a href="http://arxiv.org">http://arxiv.org</a>
FIS Bildung	DIPF	P	798	183	Pädagogik	<a href="http://www.fachportal-paedagogik.de/fis_bildung">http://www.fachportal-paedagogik.de/fis_bildung</a>
EconStor	ZBW	O	47	47	Wirtschaft	<a href="http://www.econstor.eu">http://www.econstor.eu</a>
SSOAR	GESIS	O	26	26	Sozialwissenschaft	<a href="http://www.ssoar.info">http://www.ssoar.info</a>
pedocs	DIPF	O	4	4	Pädagogik	<a href="http://www.pedocs.de">http://www.pedocs.de</a>

Tabelle 2.1: Beispiele für digitale Bibliotheken. Es wird unterschieden zwischen kommerziellen Systemen und Katalogen (K), öffentlichen Systemen und Metakatalogen (P), Open-Access-Repositoryn (O) sowie Suchmaschinen- und Web-Crawl-Systemen (S). Die Größe der Systeme ist in Tausend angegeben jeweils für die Anzahl der Nachweise (NW) und der Volltexte (VT). Die Daten beziehen sich auf den Stand Januar 2013. Systeme, deren Größe nur geschätzt werden konnte, sind durch ° markiert.

werden auch teilweise Inhalte kommerzieller Anbieter (z. B. im Rahmen der Nationallizenzen der Deutschen Forschungsgemeinschaft, DFG) in die Systeme eingebunden und der Öffentlichkeit zur Verfügung gestellt. Dies ermöglicht auch vielfach den kostenfreien Zugriff auf die vorgehaltenen Daten zum Zwecke der wissenschaftlichen Untersuchung.

### 2.2.3 Open-Access-Repositoryen

Einen klaren Fokus auf die Akquise von Volltexten legen sogenannte Open-Access-Repositoryen. Diese Dokumentenserver werden meist von Hochschulen oder Forschungseinrichtungen betrieben und lassen sich in institutionelle und disziplinäre Angebote unterteilen. Institutionelle Angebote sammeln vorwiegend die Publikationen der Angestellten der eigenen Institution und stellen diese frei zugänglich über das Repository zur Verfügung. Disziplinäre Repositoryen sammeln nicht nur mit dem Fokus auf eine Institution, sondern für einen bestimmten Fachbereich. Die vier in Tabelle 2.1 vorgestellten Systeme arXiv, EconStor, SSOAR und pedocs sind ausnahmslos disziplinäre Repositoryen, jeweils für die Fachgebiete Physik und Informatik, Wirtschaftswissenschaften, Sozialwissenschaften und Pädagogik.

Die Volltexte werden zu einem großen Teil auf den Systemen selbst gespeichert und üblicherweise nur rudimentär erfasst. Dies liegt vor allem an der häufig durchgeführten Selbsterschließung der Nachweise durch die Autoren selbst. Im Open-Access-Umfeld hat sich die Erfassung mit Dublin Core und der Klassifizierung nach DDC als kleinster gemeinsamer Nenner etabliert. Da alle in den Repositoryen gesammelten Inhalte kostenfrei (meist auch unter einer sehr offenen Lizenz wie z. B. den Creative Commons) angeboten werden, ist eine wissenschaftliche Nachnutzung der Nachweise und der Volltexte problemlos möglich. Mit OAI-PMH (Open Archive Initiative Protocol for Metadata Harvesting) existiert sogar ein eigenes Austauschprotokoll, um die Inhalte von Open-Access-Repositoryen zu sammeln. Dieser Vorgang wird auch als *Harvesting* bezeichnet.

### 2.2.4 Suchmaschinen- und Web-Crawl-Systeme

Einen übergreifenden Zugang zu den zuvor genannten Systemen möchten Suchmaschinen wie z. B. BASE, Google Scholar und CiteseerX bereitstellen. Im Gegensatz zu den bisher vorgestellten Systemen halten diese Systeme die Nachweis- bzw. Volltextdaten nicht selbst vor, sondern pflegen nur einen Index, in dem nach den Informationen gesucht werden kann. Die eigentlichen Nachweise und Volltexte sind über Harvestingvorgänge (BASE) oder Web-Crawling (Google Scholar und CiteseerX) gesammelt worden. Systembedingt haben Suchmaschinen daher oft ein Dublettenproblem.

Während die beiden Systeme BASE und CiteseerX öffentlich finanziert und somit auch die Inhalte frei verfügbar sind, ist Google als kommerzieller Anbieter nur schwer zu untersuchen. So musste bspw. die Größe von Google Scholar nach einem Verfahren von Jacsó (2008) geschätzt werden, da es als einziges System keine genaue Größenangaben enthält. Die Datenqualität ist stark abhängig von den jeweiligen Datenquellen und ist, da das primäre Interesse nur auf dem Füllen eines Suchindex liegt, nicht mit denen von kommerziellen oder öffentlichen Nachweisverzeichnissen zu vergleichen.

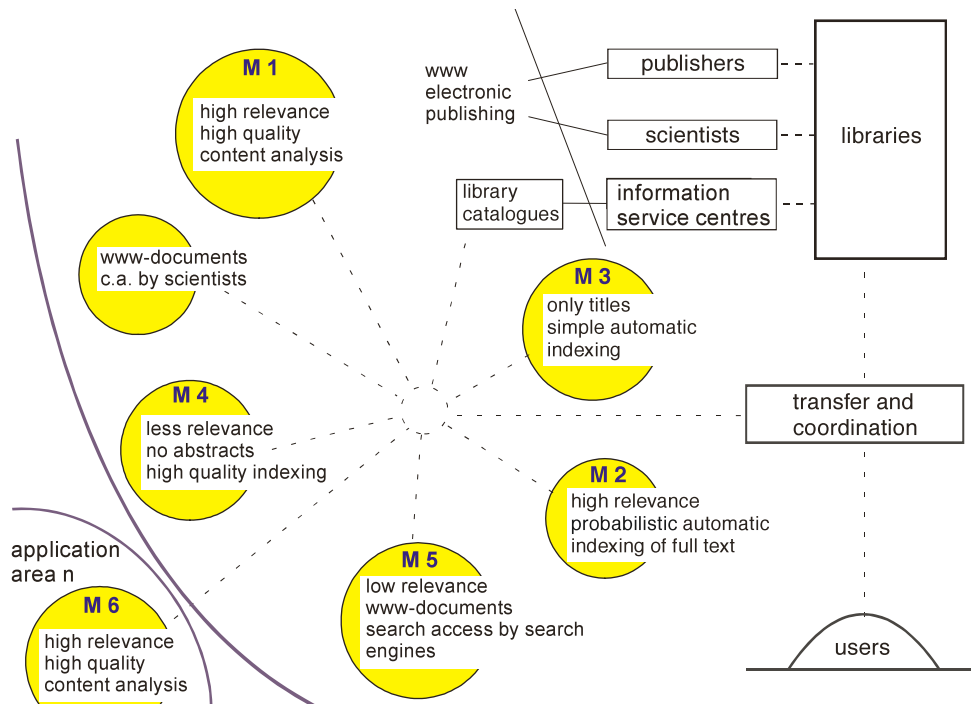


Abbildung 2.2: Polyzentrische Informationslandschaft (entnommen aus Krause, 2006, S. 98).

## 2.3 Polyzentrische Informationslandschaft

Die in Abschnitt 2.2 vorgestellten Einzelsysteme sind Bestandteil der sogenannten polyzentrischen Informationslandschaft (s. Abb. 2.2), wie sie u. a. von Krause (2006) beschrieben wird. Weder gibt es ein System, das alle anderen Einzelsysteme einbindet, noch gibt es eine umfassende Kooperation oder Verbindung zwischen ihnen, sowohl auf technischer als auch inhaltlicher Ebene. Die Inhalte unterscheiden sich in der kostenfreien Verfügbarkeit, Erschließungsqualität und fachlicher Abdeckung. Ein Wissenschaftler, der in dieser Informationslandschaft einen Überblick über die für ihn relevante Literatur erlangen möchte, muss zwangsläufig auf mehrere Informationsanbieter und -dienstleister zurückgreifen.

Um sowohl die quantitativen wie auch qualitativen Anforderungen an die wissenschaftliche Informationsversorgung befriedigen zu können, gibt es eigens eine spezielle Förderlinie der DFG, wie auch nationale bzw. europäische Bestrebungen einer umfassenden Literatur- und Kulturgüterversorgung. Hintergrund der Bestrebungen aller Beteiligten ist dabei die polyzentrische Informationslandschaft, die eine konzeptuelle Trennung zwischen selbständig agierenden Fachportalen, virtuellen Fachbibliotheken und Informationsdienstleistern auf der einen und übergeordneten Portalen und sogenannten Heterogenitätskomponenten zur Übersetzung zwischen diesen Systemen auf der anderen Seite vorsieht. Obwohl das fachübergreifende Wissenschaftsportal Vascoda 2011 als Einstiegsportal für die heterogene Wissenschaftslandschaft Deutschlands eingestellt wurde, erfreuen sich Nachfolgeprojekte

## 2 Digitale Bibliotheken

wie z. B. die Europeana, die einen europaweiten einheitlichen Zugang zu digitalisierten Literatur- und Kulturgütern liefern soll, steigender Nutzungszahlen und einer ungebrochenen umfangreichen Förderung durch staatliche Mittelgeber. Zentraler Punkt aller Systeme ist dabei wiederum der Qualitätsanspruch, der „primär fachbezogen organisiert“ ist und darauf basiert „ausschließlich Dokumente [...], deren wissenschaftlicher Wert verifiziert ist“ (Krause u. a., 2008, S. 3) anzubieten.

Ein konkretes Beispiel für ein solches System ist das zuvor beschriebene sozialwissenschaftliche Informationsportal Sowiport, das bei GESIS entwickelt und betrieben wird. Sowiport ist sowohl Aggregator für unterschiedliche Informationsquellen und Informationstypen, gleichzeitig aber auch ein gemeinsamer Sucheinstieg mit einer Reihe an Mehrwertdiensten wie z. B. die zentrale Heterogenitätskomponente zur Vereinheitlichung von Suchanfragen oder die sogenannten Search Term Recommender (s. Abschnitt 3.3) zur Unterstützung der Nutzer bei der Formulierung der Suchanfrage. Gerade die Heterogenitätskomponente ist eine explizite Lösung für Probleme, die sich aus der polyzentrischen Erschließung ergeben. Die Heterogenität von Datenerschließung und -qualität ist systemimmanent.

Neben den organisatorischen Ansätzen der Forschungsgemeinschaften wie Vascoda oder Sowiport gibt es die Vision des Semantic Web (Berners-Lee u. a., 2001) als Methode zur Beherrschung der polyzentrischen Informationslandschaft. Dabei sollen vorhandene Informationen, die ggf. bereits im Web verfügbar sind, mit semantischen Annotationen versehen werden, um Mehrwerte daraus zu generieren:

„The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. The first steps in weaving the Semantic Web into the structure of the existing Web are already under way. In the near future, these developments will usher in significant new functionality as machines become much better able to process and "understand" the data that they merely display at present.“ (Berners-Lee u. a., 2001, S. 35)

Zur Umsetzung dieser Vision wurden in der Vergangenheit eine Reihe von Standards und Technologien vorgeschlagen, so z. B. XML, RDF(S) und OWL (Hitzler u. a., 2008, S. 10–11). Im Bereich der digitalen Bibliotheken wurden diese technischen und konzeptionellen Möglichkeiten erst langsam und zögerlich angenommen, führten allerdings zu ersten Systemen wie JeromeDL oder BRICKS (Kruk u. McDaniel, 2009, S. 139 ff.).

Tatsächlich war in den frühen 2000er Jahren eine paradoxe Entwicklung zu beobachten: Digitale Bibliothekssysteme mit ihren jahrzehntelang gepflegten Datenbeständen und präzisen Anfragemöglichkeiten eiferten im Schatten erfolgreicher Websuchmaschinen der vermeintlich einfachen Suche über ein einziges Suchfeld nach, während die Semantic Web Community versuchte die Unstrukturiertheit des Webs zu kontrollieren:

„Wie kann es möglich sein, dass die erweiterten Oberflächen für Digitale Bibliotheken in der Entwicklung zurückfallen, wenn moderne Web-Informationssysteme (wie z. B. aus der Semantic Web Initiative resultierend) sich gerade bemühen, mehr Struktur und semantische Klarheit zu schaffen?“ (Mayr, 2009b, S. 18)



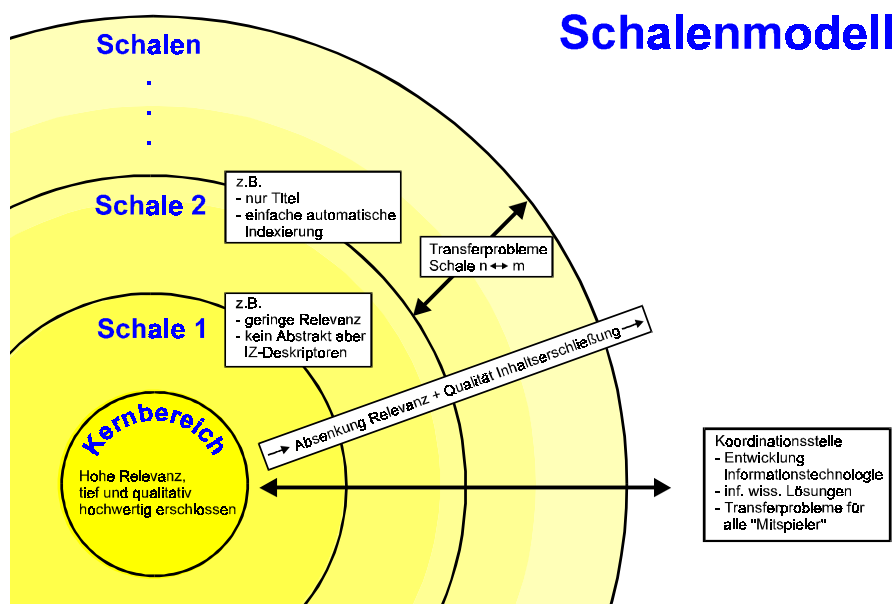


Abbildung 2.3: Schalenmodell (entnommen aus Krause, 1996, S. 18).

Der eingangs beschriebene Mangel an übergreifenden Systemen und einheitlichen Sucheinstiegen ist durch beide Herangehensweisen nicht abschließend aufzulösen. Autoren wie Krause (2012, S. 187–188) plädieren für eine pragmatische Lösung wie das Schalenmodell (s. Abb. 2.3) und die Ausnutzung vorhandener Erschließungsverfahren wie Thesauri und deren Verknüpfung (Krause, 2008). Das World Wide Web Consortium (W3C) hat mit der Veröffentlichung des SKOS-Standards (Simple Knowledge Organization System<sup>3</sup>) eine Möglichkeit geschaffen, Thesauri als „Alternative zur den stärkeren (aber auch aufwändigeren) Ontologiesprachen“ (ebenda) zu integrieren.

Für die digitalen Bibliotheken bedeutet dies, dass sie trotz der beschriebenen Entwicklungen des Semantic Web oder der Öffnung von Fachinformationssystemen immer noch einen Ankerpunkt zwischen den beiden Extremen der potentiellen Unstrukturiertheit des Webs und der Strukturiertheit der Fachdatenbanken darstellen. Sie versuchen auf pragmatische Weise die Stärken und Schwächen der beiden Extreme bestmöglich miteinander zu verbinden (s. Abb. 2.1 und Fox u. a., 2012).

## 2.4 Metadaten in digitalen Bibliotheken

In allen bisher beschriebenen Varianten von Informationssystemen wird den sogenannten Metadaten eine besondere Bedeutung beigemessen. Metadaten, die primär zunächst „Daten über Daten“ sind, werden eingesetzt, um Daten oder Objekte strukturiert zu beschreiben. In frühen digitalen Bibliotheken und den zuvor genannten Referenz- und Quelldatenbanken war die Suche über diese Metadaten die einzige Möglichkeit, die

<sup>3</sup><http://www.w3.org/TR/skos-reference/>

The screenshot shows the Europeana search results page for the query 'golem'. At the top left is the Europeana logo with the tagline 'think culture'. The main heading is 'Explore Europe's cultural collections'. Below this is a search bar containing the text 'golem' and a 'Search' button. A 'Help' link is also visible. Below the search bar, there are links for 'Return to search results' and 'Next'. The main content area features a large image of a film still from 'Der Golem, wie er in die Welt kam'. To the right of the image is a metadata section with the following text: **Paul Wegener als Golem, Albert Steinrück als Rabbi Loew und Ernst Deutsch als Famulus im Stummfilm "Der Golem, wie er in die Welt kam" von Paul Wegener und Carl Boese. Ufa, 1920**. Below the title, it lists 'Creator: Ross-Verlag (Fotograf) | > Unbekannter Fotograf (Fotograf) | >', 'Contributor: Wegener, Paul Steinrück, Albert Deutsch, Ernst Boese, Carl', and 'Date: 1935'. The 'Type' is 'Image'. The 'Subject' is 'Schauspieler | > Der Golem, wie er in die Welt kam | > Fotos | >'. The 'Description' reads: 'Paul Wegener als Golem, Albert Steinrück als Rabbi Loew und Ernst Deutsch als Famulus im Stummfilm "Der Golem, wie er in die Welt kam" von Paul Wegener und Carl Boese. Ufa, 1920, Vom Werden deutscher Filmkunst. Der stumme Film (Bilder aus dem Sammelwerk Nr. 10)'. The 'Provider' is 'Deutsche Fotothek | > Germany | >'. To the right of the image and metadata are social media sharing options (Like, +1, Tweet) and a 'Translate details' section with a language selector and 'Powered by Microsoft?? Translator'. There is also an 'Embed' section with 'Auto-generated tags'.

Abbildung 2.4: Screenshot aus der Europeana. Dargestellt ist eine Ergebnisseite zu einer Suchanfrage nach „Golem“. Enthalten sind, neben dem eigentlichen Bild, Metadaten wie *Creator*, *Contributor* etc. (Screenshot vom 11. Juli 2012).

hinterlegten Daten und Dokumente zu finden. Die Erschließungsqualität der Metadaten hatte direkte Auswirkung auf die Suchmöglichkeiten innerhalb der Referenzobjekte, die durch die Metadaten beschrieben wurden (Voß, 2012).

Dennoch war und ist eine direkte Suche innerhalb der Datenbestände in vielen Systemen nicht möglich. In großen digitalen Bibliotheken wie z. B. der Europeana (s. Abb. 2.4), die neben textuellen Dokumenten auch Bild-, Audio- und Videomaterial enthält, ist die Suche über die Metadaten die einzige Möglichkeit auf die Datenbestände zuzugreifen. Obwohl die direkte Suche in solchen nicht-textuellen (in diesem Kontext multimedialen) Datenbeständen große Fortschritte macht, ist diese Art der Suche bisher noch Spezialsystemen und der Forschung vorbehalten – im Moment dominiert die text-orientierte Suche in Metadaten (Neal, 2012).

Auch in Systemen, die Volltexte durchsuchbar machen, sind Metadaten von Bedeutung. Eine Volltextsuche ist bspw. nicht in der Lage, zwischen dem Autor eines Textes und der Erwähnung eines Namens in den Zitationen zu unterscheiden. Eine entsprechende Auszeichnung der Metadaten nach klar definierten Regeln erlaubt eine gezielte Suche u. a. nach Personen, Institutionen oder Publikationsjahren. In bibliografischen Informationssystemen gibt es daher mehrere Standards zur Beschreibung von Publikationen und Dokumenten unterschiedlichster Art. Diese sorgen für eine Normierung der Erschließungsvorgänge und erlauben eine Grundannahme über die Qualität der Erschließung.

Beispiele hierfür sind deutsche Normen wie DIN 1502 oder DIN 1505-2, die Regeln zur Vereinheitlichung von Zeitschriftennamen oder Zitierungen beinhalten. Für die Entwicklung und Anwendung von Thesauri gibt es die DIN 1463-1 und das internationale Äquivalent ISO 2788, die unterschiedliche Relationsarten innerhalb der Thesauri oder Abkürzungen beschreiben. Seitens der Deutschen Nationalbibliothek wurde 2012 die Vielzahl an verschiedenen Normdateien zur Gemeinsamen Normdatei zusammengefasst (Behrens-Neumann, 2012).

Im folgenden Abschnitt werden die Erschließungsregeln für die in dieser Arbeit betrachteten Dokumentarten Monografien, Sammelwerke und Zeitschriftenaufsätze kurz beschrieben, da auf die jeweiligen Eigenschaften später bei der Entwicklung der Retrievalmodelle zurückgegriffen wird. Neben den jeweils genannten Quellen bieten auch Umlauf u. Gradmann (2009) einen guten Überblick über weitere relevante Erschließungsregeln.

### 2.4.1 Zeitschriften

Zur eindeutigen Identifizierung von Zeitschriften wurde die Internationale Standardnummer für fortlaufende Sammelwerke (*International Standard Serial Number*, ISSN) entwickelt, die von der ISO unter der Nummer 3297 zertifiziert wurde. Die Vergabe einer international verbindlichen und weltweit einmaligen Nummer zur Identifikation von Zeitschriften ist notwendig, da diese über die oft lange Lebenszeit einer Zeitschrift unter verschiedenen Namen und Schreibweisen bekannt sein können. Die Zeitschrift „Schmollers Jahrbuch – Journal of Applied Social Science Studies“ (ISSN 1439-121X) hat seit ihrer Gründung 1871 als „Jahrbuch für Gesetzgebung, Verwaltung und Rechtspflege des Deutschen Reiches“ unterschiedliche Namen getragen (zur Entwicklungsgeschichte s. Abb. 2.5). Mit einer ISSN können nicht nur Zeitschriften, sondern allgemein Periodika wie „Jahrbücher, Jahresberichte, Adressbücher, Schriftenreihen, Loseblattwerke, CD-ROMs, DVDs oder Datenbanken“ versehen werden (Emrich, 2011a).

Das Problem der unterschiedlichen Namensgebung über die Lebensdauer wird durch die Vergabe von ISSN nur indirekt gelöst, da entsprechend der Richtlinien des ISSN International Centre die ISSN bei jeder Namensänderung und bei Änderung der Publikationsform von Print- auf Onlinepublikation ebenfalls geändert werden muss. Dies ermöglicht es aber zumindest, an einer zentralen Stelle die verzeichneten Änderungen nachzuvollziehen. Das ISSN Portal<sup>4</sup> ermöglicht nach vorheriger kostenpflichtiger Registrierung einen zentralen Zugriff auf die ISSN-Bestände über die Webseite des ISSN International Centre. Somit kann die ISSN trotz dieser Einschränkung als wichtiges Werkzeug zur klaren Unterscheidung von Zeitschriften gesehen werden, da durch ihre Verwendung die Probleme der Ansetzungsformen von Zeitschriftennamen, Sonderausgaben, mehrsprachigen Titeln oder simpler Rechtschreibfehler vermieden werden.

Die achtstellige ISSN wird aus den Ziffern 0–9 gebildet und in zwei vierstellige Blöcke aufgeteilt. Die letzte Stelle ist dabei eine Prüfziffer, die sich durch die Differenz von 11 zum Modulo 11 der Quersumme berechnet. Die Quersumme wird zuvor von der ersten bis zur letzten Stelle von 8 bis 2 absteigend gewichtet. Anstelle des Wertes 10

---

<sup>4</sup><http://portal.issn.org>

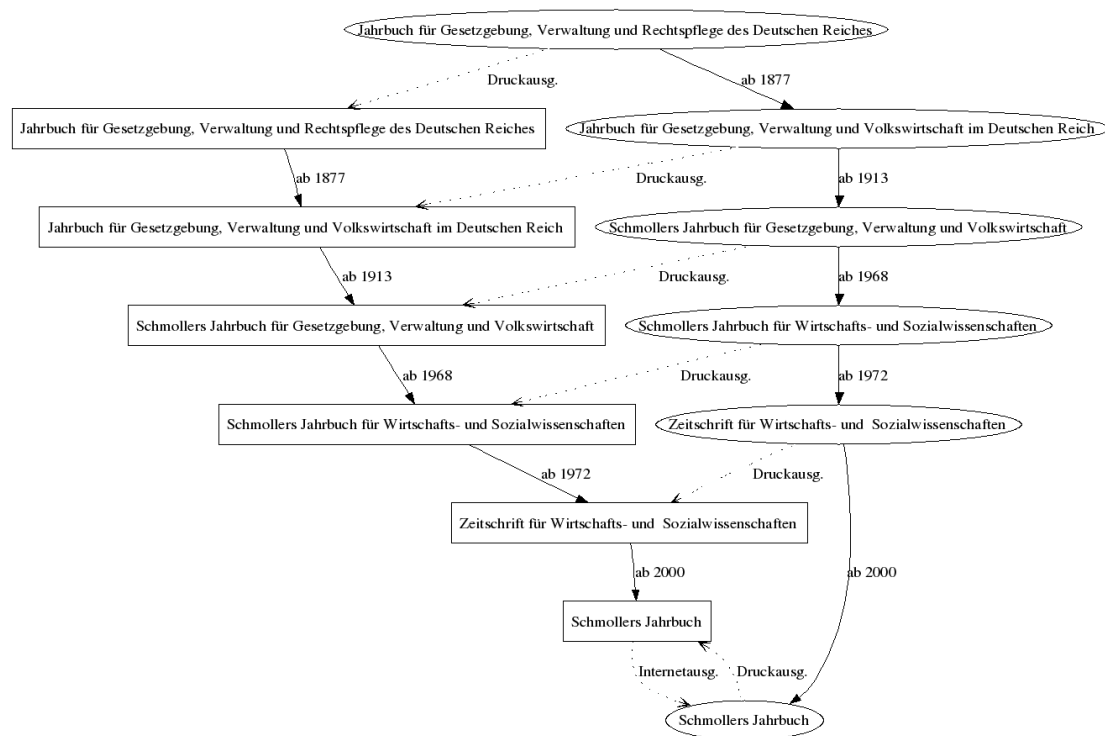


Abbildung 2.5: Die aktuell als „Schmollers Jahrbuch – Journal of Applied Social Science Studies“ bekannte Publikation hat eine bis ins Jahr 1871 zurückreichende Geschichte. Die Zeitschrift ist eine Fachzeitschrift für Wirtschafts- und Sozialwissenschaften und wurde in den letzten 140 Jahren mehrfach umbenannt und ist mittlerweile in digitaler Form vorhanden (entnommen aus Voß, 2007).

wird das Zeichen X angegeben, was auch den Wert für die o.g. Zeitschrift erklärt:  $11 - (8 \cdot 1 + 7 \cdot 4 + 6 \cdot 3 + 5 \cdot 9 + 4 \cdot 1 + 3 \cdot 2 + 2 \cdot 1) = 111$  und weil  $111 \bmod 11 = 1$  gilt  $11 - 1 = 10 \Rightarrow X$ .

Neben der Prüfziffer ist aber kein weiteres strukturierendes Element enthalten. Weder gibt es einen Code, der Verlag oder Herkunftsland beschreibt, wie dies bei der Internationalen Standardbuchnummer (s. Abschnitt 2.4.2) der Fall ist, noch ist eine Information über die Periodizität, d. h. die Erscheinungshäufigkeit, enthalten. Die ISSN wird international dezentral vergeben. In Deutschland ist die Deutsche Nationalbibliothek für die Zuteilung der ISSN zuständig (Schütz, 2011). Insgesamt sind bis zum Jahr 2010 mehr als 1,55 Millionen Nummern vergeben worden.

## 2.4.2 Monografien

Ähnlich zu den ISSN für Periodika gibt es die Internationale Standardbuchnummer (*International Standard Book Number*, ISBN), die strukturelle Informationen beinhaltet.

Sie ist als DIN 1462 bzw. als ISO 2108 genormt. Die ISBN besteht aus dreizehn Stellen und ist in fünf Teile aufgeteilt, die Emrich (2011b) im Einzelnen wie folgt beschreibt:

- 1. Teil: Präfix** (dreistellig, wird international vergeben),
- 2. Teil: Gruppennummer** für nationale, geographische Sprach- oder ähnliche Gruppen (ein- bis fünfstellig, Vergabe durch nationale ISBN-Agentur; die Ziffer 3 steht für deutschsprachige Länder),
- 3. Teil: Verlagsnummer** für den einzelnen Verlag innerhalb einer Gruppe (zwei- bis siebenstellig, Vergabe durch nationale ISBN-Agentur),
- 4. Teil: Titelnummer** für das einzelne Buch des in Teil 3 bezeichneten Verlages (ein- bis sechsstellig, Vergabe durch Verlage aus einem festen Nummernkontingent),
- 5. Teil: Prüfziffer** (einstellig).

Durch die entsprechende Aufteilung ist folglich in der ISBN sowohl das Herkunftsland und damit indirekt die Sprache als auch der herausgebende Verlag kodiert. Ein großer Verlag, dem eine dreistellige Verlagsnummer zugeordnet wurde, kann bis zu 100 000 Titelnummern vergeben. Bis Ende 2006 waren ISBN nur zehnstellig und der dreistellige Präfix fehlte. Eine zehnstellige ISBN kann aber im Falle von Büchern mit dem konstanten Präfix 978 versehen werden und ist damit eine korrekte dreizehnstellige ISBN. Wie auch bei der ISSN muss bei einer Änderung der Monographie, z. B. einer Neuauflage, auch die ISBN angepasst werden.

Durch die vorgegebene Struktur der ISBN kann diese analysiert werden<sup>5</sup>. Die dreizehnstellige ISBN des Buchs „Introduction to modern information retrieval“ mit der ISBN 978-1-85604-694-7 erlaubt z. B. folgende Rückschlüsse auf das Dokument:

- Die ISBN-10 lautet 1-85604-694-X,
- beide ISBN-Varianten sind in Gebrauch,
- sie gehören der Gruppe „Englischsprachige Gebiete“ an
- und der Verlag „Facet Publishing“ stammt aus dem Vereinigten Königreich oder Irland.

In Deutschland ist der Marketing- und Verlagsservice des Buchhandels für die Verwaltung der Verlagsnummern zuständig. Die internationale Koordination erfolgt durch die internationale ISBN-Agentur mit Sitz in London.

---

<sup>5</sup>Online über den ISBN-Tester der Wikimedia Foundation möglich: <https://toolserver.org/isbn/IsbnCheckAndFormat>

### 2.4.3 Autorennamen

Die Namen von Autoren oder Herausgebern (im Englischen auch allgemeiner als *creator* bezeichnet) sind ein weiterer elementarer Bestandteil von in digitalen Bibliotheken erfassten Metadaten. Vielfach werden Autorennamen in einfacher Textrepräsentation erfasst. Dies führt aber, wie auch schon bei Zeitschriften oder Monografien beschrieben, zu gewissen Ungenauigkeiten. Am Beispiel des Autors selbst ist dies anschaulich zu zeigen, da er selbst in verschiedenen Ansetzungsformen in unterschiedlichen Datenbanken verzeichnet ist: Sei es, dass der Nachname mit einem Umlaut oder ohne verwendet wird (Schaer/Schar/Schär), der Vorname unterschiedlich abgekürzt wird (Philipp, Ph., P. oder einfach P) oder Vor- und Nachname nicht richtig auseinandergehalten werden. Natürlich sind beliebige Mischungen dieser Fehler möglich. Neben den Ansetzungsformen sind mehrdeutige Namen ein Problem, wie van Noorden (2011) anschaulich aufzeigt: „Most-published researchers in 2011? Wang Y, Zhang Y, Liu Y, Wang J, Li Y, Wang L, Li J, Zhang J, Zhang L, Kim JH.“.

Scharnhorst u. Garfield (2010) erkennen ebenfalls das Identifikationsproblem und beschrieben drei unterschiedliche Lösungsansätze mit denen dem Problem entgegenge- arbeitet werden soll: (1) mittels eindeutiger Personenidentifikatoren, (2) einheitlicher Erschließungsregeln und (3) automatischer Extraktion und Zusammenführung von hete- rogenen Datenbeständen.

Sie geben eine Übersicht über die unterschiedlichen Systeme zur Vergabe eines eindeuti- gen digitalen Identifikators für Autoren. Sie unterscheiden dabei zwischen kommerziellen und öffentlich finanzierten Systemen, die nebeneinander, miteinander und auch gegenein- ander arbeiten. Kommerzielle Verlage wie z. B. Thomson Reuters oder Elsevier haben eigene Systeme zur Autorenidentifikation eingeführt wie z. B. die ResearcherID<sup>6</sup>. Dem gegenüber stehen öffentliche Systeme, wie der Digital Author Identifier (DAI) der niederländischen SURF Foundation oder der Personennamendatei (PND) der deutschen Nationalbibliothek (DNB). In diesen Systemen werden sowohl normierende Regeln für die Erschließung als auch eindeutige Identifikatoren vergeben. Im Falle der DNB ist dies die sogenannte PND-Nummer. Neben den Namen werden auch Pseudonyme, Beruf und weitere ergänzende Daten erfasst.

Neben den kommerziellen und öffentlich-finanzierten Anbietern gibt es auch Bestre- bungen, die aus der wissenschaftlichen Gemeinschaft selbst stammen. Beispiele hierfür sind das Identifikatorenprogramm des arXiv<sup>7</sup> oder das Authorclaim-System<sup>8</sup> des RePEc- Initiators Thomas Krichel. Das System ORCID<sup>9</sup> ist eine neue Entwicklung, das zum Ziel hat, vorhandene Lösungen wie die ResearcherID von Thomson Reuters, die Autorenprofile von RePEc, Scholar Universe und Scopus in einer einheitlichen technologischen Basis zusammenzubringen.

---

<sup>6</sup><http://www.researcherid.com>

<sup>7</sup>[http://arxiv.org/help/author\\_identifiers](http://arxiv.org/help/author_identifiers)

<sup>8</sup><http://authorclaim.org>

<sup>9</sup><http://orcid.org/>

#### 2.4.4 Kontrollierte Schlagwörter und Thesauri

Nach Bertram u. Umlauf (2012) ist ein Thesaurus

„[...] eine Indexierungssprache, die sich auf die natürliche Sprache gründet. Ein Thesaurus ist eine geordnete Zusammenstellung eindeutig aufeinander bezogener Begriffe und Benennungen, die zur Indexierung, zum Speichern und zum Wiederauffinden dokumentarischer Bezugseinheiten (DBE) dient.“ (Bertram u. Umlauf, 2012)

Ein Thesaurus liefert ein sogenanntes kontrolliertes Vokabular, das ermöglicht Dokumente zu beschreiben. Diese terminologische bzw. begriffliche Kontrolle ist eine der größten Stärken von Thesauri, da sie durch die Kombination der kontrollierten Begriffe (auch Deskriptoren genannt) eine große Ausdrucksmächtigkeit und -klarheit schaffen. Dies kann insbesondere für den Retrievalprozess genutzt werden, da hierbei das eingeschränkte Vokabular eine gezieltere Suche ermöglicht – vorausgesetzt, der Suchende ist in der Lage, das Vokabular des Thesaurus korrekt zu verwenden.

Ein konkretes Beispiel für einen Thesaurus ist der Thesaurus Sozialwissenschaften, der von GESIS entwickelt und gepflegt wird. Er wird hauptsächlich für die sozialwissenschaftliche Literaturdatenbank SOLIS und die Forschungsdatenbank SOFIS verwendet und liegt in insgesamt drei Sprachen vor (deutsch, englisch und russisch). Die deutsche Ausgabe enthält ca. 8 000 Deskriptoren und 4 000 Nicht-Deskriptoren, die fachlich die Disziplinen der Sozialwissenschaften abdecken. Der Thesaurus ist online verfügbar<sup>10</sup> und kann auch als SKOS-Version abgefragt werden (Zapilko u. Sure, 2009).

## 2.5 Nutzerverhalten und -bedürfnisse

In Bezug auf digitale Bibliotheken stehen aktuelle Systeme vor der Aufgabe, Nutzern qualitativ hochwertige (Fach-)Information auf schnellem Wege und mit einfacher Bedienbarkeit bereitzustellen. Aktuelle Studien (z. B. die JISC-Studie Wong u. a., 2010) zeigen, dass durch die Angebote moderner Websuchmaschinen die Erwartungshaltung der Nutzer schneller wächst, als ihre Fähigkeit, mit den Systemen umzugehen:

„For example, at a simplistic level, many of the participants do not understand how to assess the quality of materials they find. Google or Google Scholar have lower thresholds of information technology literacy, and are considered their '... friends' because of the apparent higher yield or success rate.“ (Wong u. a., 2010, S. 5)

Der folgende Abschnitt wird das gewandelte Rollenbild von digitalen Bibliotheken beleuchten und die Ergebnisse aktueller Nutzerstudien aus dem Umfeld der Informationsrecherche zusammenfassen.

<sup>10</sup><http://www.gesis.org/sowipport/nc/suche/thesaurus.html>

### 2.5.1 Die Rolle digitaler Bibliotheken

Wie Anderson (2011) beschreibt, befinden sich die wissenschaftlichen Bibliotheken aktuell in einer „Crisis in Research Librarianship“, womit der Zustand, dass in reinen Zahlen immer weniger Nutzer die Fachdienstleistung einer wissenschaftlichen Präsenzbibliothek in Anspruch nehmen, gemeint ist. So habe sich in den letzten 15 Jahren, nach den offiziellen Zahlen der Association of Research Library (ARL), die Zahl der Geschäftsgänge (*reference transactions*), die sich innerhalb einer der erfassten Bibliotheken ereignet haben, mehr als halbiert. Angesichts der steigenden Zahl an Studenten pro Jahrgang, sei der Rückgang noch drastischer: Waren 1995 noch 10,1 Geschäftsgänge pro eingeschriebenem Student zu verzeichnen, waren dies 2009 nur noch 3,6. Für Anderson sind diese Zahlen ein Indiz für die These, dass Nutzer wissenschaftlicher Bibliotheken (in diesem Falle Studenten), ihre Informationsbedürfnisse zunehmend außerhalb der traditionellen Institutionen befriedigen – zum einen, weil es durch die Vielzahl an elektronischen Zugriffsmöglichkeiten umfangreiche Alternativen gibt und zum anderen, weil sich die Nutzer zunehmend als fähig betrachten, Informationen selbst zu recherchieren und keine Unterstützung durch Bibliothekare mehr in Anspruch nehmen. Auch Hätscher u. a. (2007) stellen fest, dass Präsenzbibliotheken bei der Informationssuche keine wichtige Rolle mehr spielen, wohl aber bei der Beschaffung oder der Lizenzierung von Datenbanken.

Diese Thesen sind vor den steigenden Zugriffszahlen in digitalen Bibliotheken, Fachportalen und anderen Online-Angeboten schlüssig. Allerdings zeigen verschiedene Studien, dass zwar quantitativ eine Bewegung weg von den Präsenzbibliotheken und dortiger Beratungen zu sehen ist, diese Entwicklung aus Sicht der Nutzer jedoch sehr ambivalent ausfällt. So werden entfallene Fußwege und eine größere Datenbasis als positiv bewertet, doch viele Nutzer der elektronischen Angebote sind bei der eigenständigen Suche schlichtweg überfordert, trotz spezieller Hilfestellungen wie Schritt-für-Schritt Anleitungen oder Fibeln, die speziell für die wissenschaftliche Online-Recherche ausgelegt sind (Day, 2001). Zu gleichen Ergebnissen kam auch eine Studie von Xie (2008): Nutzer sind bei der Suche häufig überfordert und werden durch DL-Systeme nicht genügend unterstützt.

In Studien von te Boekhorst u. a. (2003) wird der „Information Overload“ als eines der Hauptprobleme bei der Suche von wissenschaftlicher Literatur herausgearbeitet, wobei zwischen einer nachfrageinduzierten und einer angebotsinduzierten Ausprägung des Problems unterschieden wird:

„Im Fall des nachfrageinduzierten Overload-Problems liegt die Ursache vor allem in der mangelnden Informationskompetenz (information literacy) der Benutzenden [...]. Im Fall des angebotsinduzierten Problems übersteigt das Angebot, trotz ausreichender Kenntnisse und Fähigkeiten bei der Recherche, das Maß der Informationsmenge, die für die jeweiligen Wissenschaftler konsumierbar ist.“ (te Boekhorst u. a., 2003, S. 4 f.)

Die Autoren der Studie weisen auf zwei Strategien der Informationssuchenden hin, um das Overload-Problem zu beherrschen. Zum einen vertrauen Wissenschaftler primär auf ihre persönlichen Informationsnetzwerke (*invisible colleges*), die sie mit aktueller und relevanter Literatur versorgen, zum anderen wird vermehrt nur die Art von elektronischer



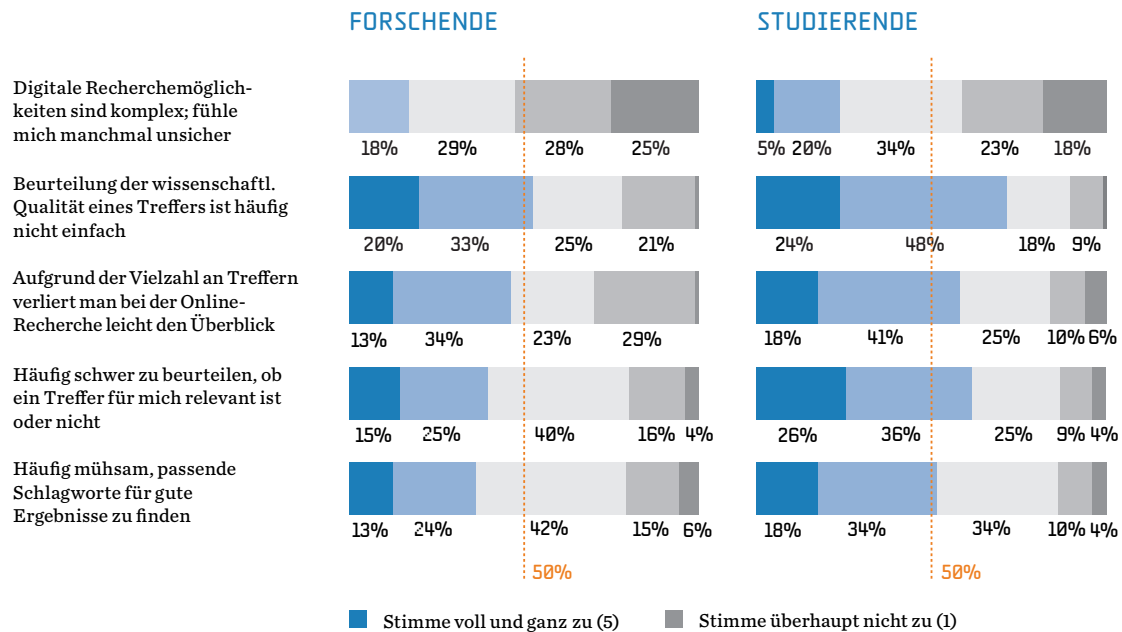


Abbildung 2.6: Typische Problemfelder bei der digitalen Recherche aus einer Studie mit N=80 forschenden und N=80 studierenden Teilnehmern der Wirtschaftswissenschaften (entnommen aus Siegfried u. Flieger, 2011, S. 6).

Information wahrgenommen, die aktuell verfügbar ist („Verfügbarkeit hat Priorität“, ebenda S. 8 ff.). Man spricht hierbei auch von einer „Now or Never-Mentalität“ (Sühl-Strohmer, 2008, S. 58).

### 2.5.2 Relevanzbeurteilung durch Nutzer

Für die Zentralbibliothek der Wirtschaftswissenschaften (ZBW) haben Siegfried u. Flieger (2011) in einer Nutzerstudie mit 160 Studenten und Forschern (wissenschaftliche Mitarbeiter, Doktoranden, Juniorprofessoren, Professoren oder Forschende) die zuvor aufgestellten Thesen zum Nutzungsverhalten in DL-Systemen untersucht. Tatsächlich berichteten die Nutzer über teilweise erhebliche Probleme bei der Recherche. So werden u. a. Probleme bei der Beurteilung der wissenschaftlichen Qualität eines Suchtreffers, Einordnung der Relevanz für das eigene Forschungsvorhaben und bei der Suche nach einem passenden Schlagwort angeführt (s. Abb. 2.6). Ähnliche Ergebnisse präsentiert auch eine OCLC-Studie (Calhoun u. a., 2009, S. 11), die einen eindringlichen Appell der Nutzer zusammenfasst: „Search results must be relevant and the relevance must be obvious.“

Gleichzeitig zeigt die ZBW-Studie sehr gut, wie und nach welchen Kriterien die Teilnehmer der Studie die Relevanz von Suchergebnissen sortieren und deren Einschlägigkeit beurteilen. Die drei wichtigsten Kriterien sind zunächst die Aktualität der Suchergebnisse (gemeint ist das Publikationsdatum), gefolgt vom Renommee der Zeitschrift und des

Suchentitäten	Anzahl der Anfrage	Anteil an Gesamtanfragen
Freitextsuche (alle Felder)	58 754	45,5 %
Person(en)	40 979	31,7 %
Schlagwort	26 959	20,9 %
Titel	1 108	0,9 %
Quelle	581	0,4 %
Andere	354	0,3 %

Tabelle 2.3: Ergebnis einer Logfile-Analyse des sozialwissenschaftlichen Portals Sowiport. Analysiert wurden die 1 000 populärsten Suchanfragen (insgesamt N=129 251). Aufgelistet werden die Anzahl der Fragen nach Suchentität und deren prozentualen Anteil an der Gesamtmenge der Anfragen. Die letzte Zeile fasst die Entitäten Publikationsjahr, Institution und Ort zusammen (Tabelle angelehnt an Schaer u. a., 2012c).

Autors. Weiterhin werden die Häufigkeit der Zitation und die Anzahl der Downloads als Qualitäts- und damit Relevanzindikator herangezogen (Siegfried u. Flieger, 2011, S. 6 f.). Wie aber schon zuvor beschrieben, sind diese Relevanzkriterien für 53 % der Forschenden nur schwer nachzuvollziehen und einzuordnen.

Weiterhin greifen Forschende gerne auf „nahestehende“ Experten zurück, sowohl bei der Suche als auch bei der Einordnung nach Relevanz. Die meist dem sogenannten Web 2.0 zugeordnete Funktionalität der Bewertung von Dokumenten durch Benutzer (*social feedback*) wird von Wissenschaftlern hingegen meist als störend wahrgenommen – ausgenommen davon sind professionelle Reviews und Expertenmeinungen (Calhoun u. a., 2009, S. 18).

Neben Studien, die auf Nutzerbefragungen und Interviews setzen, werden zunehmend auch Analysewerkzeuge wie z. B. Google Analytics eingesetzt, um Studien zum Nutzerverhalten (*User Behaviour Studies*) durchzuführen (Hess, 2012). So wurde im Rahmen einer Logfile-Analyse des sozialwissenschaftlichen Portals Sowiport von Schaer u. a. (2012c) ermittelt, dass von der Vielzahl an erfassten Metadaten nur wenige in der direkten Suche eingesetzt werden. Bei der Analyse der 1 000 populärsten Suchanfragen (insgesamt N=129 251) zeigte sich, dass in knapp 1/3 der Fälle explizit nach einer Personen bzw. deren Namen gesucht wurde; nur 1/5 der Benutzer verwendete explizit kontrollierte Schlagwörter für die Suchanfrage. Die weiteren vorhandenen Suchfelder für Publikationsjahr, Quelle, Institutions- oder Ortsnamen wurde in weniger als 1 % der untersuchten Fälle verwendet. Nahezu die Hälfte der Suchanfragen waren einfache Freitextsuchen, die alle vorhandenen Metadaten abdeckten, sodass hier keine bestimmten Entitäten bestimmt werden konnten (s. Tab. 2.3).

Das Suchverhalten der Benutzer im Portal Sowiport kann dabei als relativ typisch für die Nutzer in digitalen Bibliotheken allgemein angesehen werden: In vielen Fällen wird (ob aus Unkenntnis oder Unzufriedenheit angesichts der Alternativen) auf erweiterte Suchunterstützungsmöglichkeiten in Form spezialisierter Dienste oder Suchmasken nicht

zurückgegriffen. Dies mag daran liegen, dass entgegen des tatsächlichen Nutzerinteresses an bspw. Personen, Dienste angeboten werden, die speziell auf kontrollierte Vokabulare abzielen, z. B. in Form sogenannter Search Term Recommender (Hienert u. a., 2011a; Petras, 2006).

### 2.5.3 Disziplinarität digitaler Angebote

Wichtig sind die disziplinären Eigenheiten, die unterschiedliche wissenschaftliche Fächer bzw. die Fachkulturen voneinander trennen. So sind z. B. in den Geistes- und Sozialwissenschaften (*Humanities and Social Sciences*) vor allem Monografien von Bedeutung, die sich in eher technisch-naturwissenschaftlichen Fachkulturen (*Science, Technology and Medicine*) so nicht wiederfinden lassen (Adema u. Rutten, 2010). Aber nicht nur in der Publikationskultur unterscheiden sich die Fachkulturen, auch das Informationsverhalten selbst unterliegt einer hohen Varianz („one size doesn't fit all“, Connaway u. Dickey, 2010, S. 50).

„Den größten Wunsch, den Sozialwissenschaftler an Fachdatenbankanbieter haben ist es, die Wissensstruktur der Sozialwissenschaften abzubilden und gleichzeitig den Fachzugriff auf Informationen zu bieten.“ (Zens, 2010)

Die allgemeinen Ansätze, die Suchmaschinenbetreiber wie Google Scholar in ihren Angeboten bieten, können den speziellen Anforderungen der jeweiligen Fächer nur bedingt gerecht werden. Dies zeigt sich auch in Studien zum Informationsverhalten von jungen College-Studenten von Head u. Eisenberg (2009): Suchmaschinen werden häufig für sehr gezielte Suchen eingesetzt (z. B. für Volltexte oder bestimmte Literaturangaben), wobei die teilweise differenzierten Möglichkeiten der Suchunterstützung in fachspezifischen Systemen nicht bekannt sind oder generell ungenutzt bleiben. Laut Wong u. a. (2010) sehen sich Informationsanbieter und Betreiber von digitalen Angeboten einem klassischen Zielkonflikt in den Bedürfnissen der Benutzer gegenüber, die einerseits nach einfachen und übersichtlichen Suchmasken, gleichzeitig aber nach weitergehenden Filter- und Personalisierungskomponenten in der Suche verlangen, die allerdings die Komplexität der Suchsysteme erhöhen. Gleichzeitig zeigt sich aber auch in Nutzerstudien zu Websuchmaschinen (Zhang, 2008), dass die Komplexität der Anfragesysteme nicht verstanden bzw. genutzt wird (nicht einmal 1/10 der Benutzer nutzt/kennt boolesche Operatoren; knapp die Hälfte kennt die Funktionalität der Phrasensuche). Auch die interne Funktionsweise des Rankings und dessen jeweiligen Kriterien sind den untersuchten Personen größtenteils unbekannt.

## 2.6 Dokumentenretrieval

Ein essentielles Problem digitaler Bibliotheken ist die nutzerseitige Suche und das damit verbundene Ranking der Ergebnismenge. Bedingt durch die Charakteristika der Systeme, welche meist aus intellektuell verwalteten Metadaten bestehen (vgl. Abschnitt 2.4), sind viele anspruchsvolle und komplexe Retrieval- und Rankingfunktionen nicht anwendbar. Dies ist größtenteils auf das Fehlen passender Daten zur Berechnung dieser Verfahren

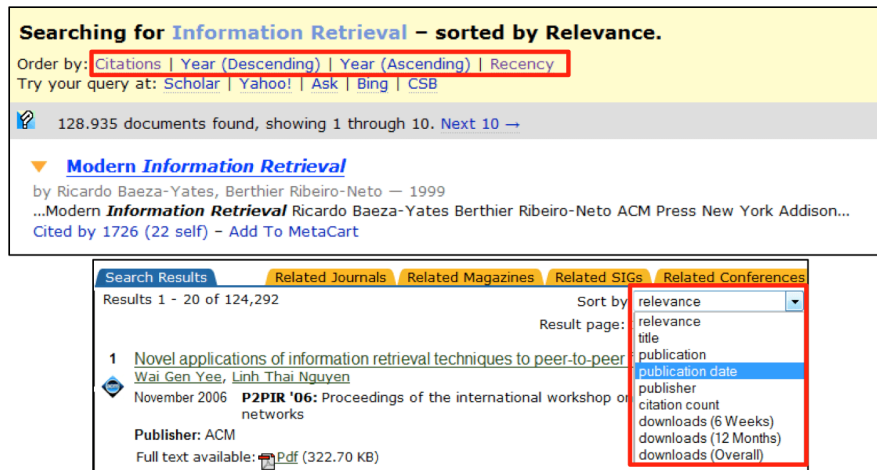


Abbildung 2.7: Beispiele für Ranking- bzw. Sortierverfahren aus den beiden digitalen Bibliotheken Citeseer (oben) und ACM Digital Library (unten). Markiert sind die Sortiermöglichkeiten, die von den Systemen angeboten werden (Screenshots vom 2. Dezember 2011).

zurückzuführen. So scheitert z. B. eine Linkanalyse (vgl. PageRank oder HITS, Abschnitt 4.4.1) am Fehlen solcher Verknüpfungen der Metadaten untereinander, bzw. an einer ausreichenden Abdeckung der Metadaten mit den dazugehörigen Volltexten, die z. B. eine Zitationsanalyse zuließe. Weiterhin sind nicht in allen Systemen Mechanismen zur Verwaltung von User-Profilen eingebaut, die eine Auswertung von Nutzungsverhalten und Relevance-Feedback erlauben würden.

In einer Studie von Buckley (2009) zeigte sich, dass einige seiner Testsysteme einen Großteil der vorhandenen Metadaten schlichtweg ignorierten und so die Retrievalsysteme nicht darauf zurückgreifen konnten. Selbst wenn die Metadaten indiziert wurden, wurden sie meist nicht für das Retrieval verwendet. Buckley stellt die Frage, ob Potential verschenkt wurde und ob zusätzlich zu sprachzentrierten Verbesserungen wie Stemming oder einer NLP-getriebenen Anfrageerweiterung andere Eigenschaften der Dokumente hätten genutzt werden können, um das Suchergebnis zu verbessern. Buckley fasst in der Debatte die folgenden essentiellen Probleme aktueller Retrievalsysteme zusammen: (1) technische Fehler (Stemming, Tokenization etc.), (2) Überbewertung oder Fehlbewertung einzelner Begriffe oder Aspekte in der automatischen Anfragegenerierung und (3) weitere Probleme, die aus Bereichen des Natural Language Understanding (NLU) und Natural Language Processing (NLP) stammen (dazu auch Spärck Jones, 1999). Neben den sprachzentrierten Problemen wurde unter anderem festgestellt, dass die in den Systemen vorhandenen Metadaten nur unzureichend für das Retrieval genutzt wurden. In der mit mehr als tausend Personenstunden und insgesamt sieben bewerteten Retrievalsystemen groß angelegten Studie kam man zu dem Schluss, dass nicht-term-basierte Sucheigenschaften das Suchergebnis erheblich verbessern können. Im Falle der Buckley-Studie war dies die Angabe der Zeitschrift Financial Times, in der die Artikel erschienen.

Wirft man einen Blick auf typische digitale Bibliotheken wie die ACM Digital Library, das CiteseerX-System, das Web of Science oder ScienceDirect von Elsevier, werden dem Benutzer nur eine recht begrenzte Anzahl an Sortierungsmöglichkeiten geboten (s. Abb. 2.7), doch ein wirkliches Ranking ist meist nur in Form von Relevance-Ranking (s. Kap. 4) implementiert. Neben klassischem text-basierten Relevanzranking wird die Umsortierung nach nominellen Kriterien (Erscheinungsjahr) als auch alphabetischen Kriterien (Titel oder Herausgebername) angeboten.

Systeme wie z. B. arXiv verzichten vollständig auf ein Ranking der Dokumente und bieten nur ein boolesches Retrieval mit einfacher Sortierung nach Eingangsdatum oder alphabetischen Kriterien. Im Falle von ScienceDirect kommen hingegen neben einfachen Sortierkriterien wie Erscheinungsjahr oder Downloadzahlen, das bekannte TF\*IDF-Ranking (mehr dazu in Abschnitt 4.3.2) hinzu:

„Lists results according to the relative frequency that your search terms appear in each publication. If a term appears frequently in a publication, that publication is listed before one in which the term appears less frequently.“  
(Elsevier, 2012)

Die Popularität des sogenannten Vektorraummodells für digitale Bibliotheken zeigt sich auch in der Nutzung der freien Suchmaschinentechologie Solr, die bei Citeseer oder der OPAC-Software Blacklight<sup>11</sup> eingesetzt wird. Blacklight ist als freier „next-generation catalog“ für den Einsatz mit heterogenen Dokumentensammlungen konzipiert worden.

Lewandowski (2009) stellt eine Übersicht über Rankingfaktoren zusammen, die in digitalen Bibliothekssystemen Verwendung finden. Er schlägt fünf unterschiedliche Kategorien dieser Faktoren vor, die genutzt werden können, um Felder in der Datenbank entsprechend der Herkunft ihrer – textuellen – Information zu gewichten. Seiner Beschreibung nach sei es „best practice“ Titeltermine höher in der Rankingformel zu gewichten, als Terme, die im Abstract eines Dokuments enthalten sind. Durch den rein textuellen Abgleich von Anfrage- und Dokumentterminen wird das Retrievalsystem allerdings dazu gezwungen die zuvor sorgsam codierte Semantik zu ignorieren – eine Eigenschaft, die allen Bag-of-Words-Ansätzen gemein ist. Neben Termgewichtungen werden noch Faktoren wie Popularität, Neuigkeit oder Lokalität beschrieben. Hinzu kommen sehr domänenspezifische Eigenschaften, die nur im konkreten Anwendungsfall eines OPAC einer Präsenzbibliothek sinnvoll sind. Beispiele hierfür sind der Aufenthaltsort des Benutzers und die physischen Entsprechungen der digitalen Objekte in Form von Printmedien in einem Regal.

Andere von ihm aufgeführte Faktoren sind allerdings auch auf andere Anwendungsfelder übertragbar, wie die Dokumentengröße, Dokumententyp oder Popularitätswerte wie die Anzahl der Dokumentenansichten, Benutzerbewertungen, Zitationen etc. Speziell in modernen, Web-basierten digitalen Bibliotheken erlangt diese Art der Rankings zusehends an Zuspruch. Eine weitgehende Auswertung und Evaluation ist diesen Faktoren aber bislang noch nicht gewidmet worden, von einigen Papieren wie z. B. von Schlögl u. Gorraiz (2012) abgesehen.

<sup>11</sup><http://projectblacklight.org/>

## 2.7 Zusammenfassung

Digitale Bibliotheken sind als Ankerpunkt zwischen der Unstruktiertheit des Webs und der Strukturiertheit der Fachdatenbanken zu begreifen, die versuchen einen pragmatischen und auf den Nutzer zugeschnittenen Zugang zu Informationen zu ermöglichen. Trotzdem haben aktuelle Nutzerstudien (s. Abschnitt 2.5) gezeigt, dass zum derzeitigen Zeitpunkt wissenschaftliche digitale Bibliotheken noch keine ausreichende Unterstützung des Nutzers gewährleisten, obwohl ihre Nutzung in den letzten Jahren rapide angestiegen ist. Datenbanken und Suchmaschinen sind die wichtigsten Recherchemittel, wobei die Nutzung der Systeme unterschiedlich ist. So wird die freie explorative Suche eher in digitalen Fachangeboten genutzt (aufgrund der besseren Erschlossenheit der Dokumente), wohingegen Suchmaschinen eher für gezielte Suchen genutzt werden.

Die Bedürfnisse der Nutzer von digitalen Bibliotheken zeigen allerdings den vorherrschenden Zielkonflikt auf: Einerseits recherchieren immer mehr Forschende selbstständig in und mit Fachdatenbanken, digitalen Bibliotheken und Suchmaschinen, andererseits sind sie mit der konkreten Nutzung überfordert. So geben in der ZBW-Studie von Siegfried u. Flieger 53 % der Forschenden an, Probleme mit der Relevanzbewertung von Rechercheergebnissen zu haben. Gleichzeitig werden als die drei wichtigsten Kriterien zur Relevanzbewertung die Aktualität der Dokumente sowie die Reputation der Zeitschriften und der Autoren angeführt. Heutige digitale Bibliotheken unterstützen diese teilweise sehr abstrakten Kriterien jedoch nur unzureichend. Die vorliegenden Studien können demnach als Indikator für den Bedarf an alternativen Retrievalmöglichkeiten gewertet werden.

### 3 Anfrageerweiterung und Recommendersysteme

Bei der Benutzung von hochspezialisierten Informationssystemen wird ein besonderes Problem schnell ersichtlich: Bei der Formulierung eines Informationsbedürfnisses, das in eine Suchanfrage überführt werden soll, gibt es theoretisch eine unbegrenzte Menge an Ausdrucksmöglichkeiten. Dies trifft insbesondere auch für wissenschaftliche digitale Bibliotheken zu. Jede wissenschaftliche Disziplin hat ihre eigene fachspezifische Sprache mit einem speziellen Vokabular, die nicht zwangsläufig von anderen Disziplinen geteilt wird. Indexierer und die Betreiber von digitalen Bibliotheken verwenden seit jeher diese fachspezifischen Sprachen in Form von dokumentarischen Werkzeugen wie Thesauri oder Klassifikationssystemen. Mit Hilfe dieser Werkzeuge sollte die Vielfalt der wissenschaftlichen Ausdrucksmöglichkeiten in ein einheitliches System überführt werden.

Wird Information Retrieval als „fundamentally a linguistic process“ (Blair, 2003, S. 3) betrachtet, dann müssen Nutzer in der Lage sein, diese hochspezialisierten Werkzeuge zu verwenden. Dies ist häufig aber nicht der Fall und so verhalten sich Nutzer, wie dies ein interviewter Experte in einer Studie von Aula u. a. (2005) ausdrückte:

„I choose search terms based not specifically on the information I want, but rather on how I could imagine someone wording [...] that information.“ (Aula u. a., 2005, S. 589)

Ein klassisches Werkzeug, um die o. g. Probleme bei der Anfrageformulierung zu umgehen und den Nutzer zu unterstützen, sind Mechanismen zur Anfrageerweiterung (*Query Expansion*, QE). Hierbei werden zur ursprünglichen Anfrage des Nutzers beliebig viele weitere Suchterme hinzugefügt, mit dem Ziel, eine effizientere Suche zu ermöglichen (Efthimiadis, 1996).

Neben den zuvor beschriebenen Vagheitsproblemen der sprachlichen Formulierung eines Informationsbedürfnisses ergibt sich in einer wissenschaftlichen digitalen Bibliothek ein weiteres Formulierungsproblem. Wie in Kapitel 2 gezeigt, sind für gewöhnlich Informationsobjekte mit einer Fülle an ergänzenden nicht-textuellen Metadaten wie Zeitschriften- oder Verlagsinformationen versehen, die prinzipiell für die Suchanfrage genutzt werden können. Hier kommt das Prinzip der Polyrepräsentation (Ingwersen, 1996; Ingwersen u. Järvelin, 2005) zum Tragen, das gezielt auf solche zusätzlichen Metadaten zurückgreift, um eine effektivere Suche zu ermöglichen.

Obwohl diese Metadaten in aktuellen Systemen z. B. in Form einer „erweiterten Suche“ oder einer reinen Volltextsuche nutzbar sind, wird im Kontext der Anfrageerweiterung selten auf solche Zusatzinformationen eingegangen. Um die Fülle eines Informationsbedürfnisses adäquat in eine Suchanfrage umzusetzen, sollten aber auch zusätzliche

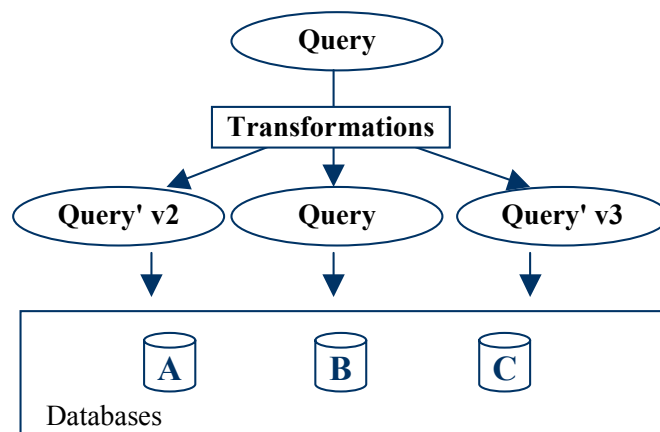


Abbildung 3.1: Vagheitsbehandlung im Information Retrieval (entnommen aus Hellweg u. a., 2001).

Aspekte in Systeme, die die Formulierung unterstützen können, einbezogen werden. Im vorliegenden Kapitel werden daher zwei verschiedene Herangehensweisen an die Thematik der Anfrageerweiterung vorgestellt: Sprachlich-textuelle Anfrageerweiterung und nicht-textuelle Erweiterungen.

### 3.1 Sprachproblem des IR

Eine Suche in textuell orientierten Informationssystemen kann nur funktionieren, wenn der Nutzer sein Informationsbedürfnis möglichst nah an der Art und Weise, wie das System sie speichert, formuliert. Dies sind im besten Fall exakt die Begriffe, die auch im Informationssystem verwendet werden. Das Problem selbst ist als das klassische *Sprachproblem des IR* (Blair, 2002, 2003) oder als das *Vagheitsproblem* (s. Abb. 3.1) bekannt: „Jedem Bibliothekar und jedem, der sich mit Information Retrieval befasst, war schon immer klar, dass zwischen den semantischen Termen, die in der Datenbank ein Dokument charakterisieren, und dem Term, den der Benutzer anwendet, nicht immer eine 1:1-Relation besteht.“ (Krause, 2003, S. 10).

Während Benutzer eine Anfrage an ein Informationssystem formulieren, befinden sie sich laut Belkin (1980) in einem „Anomalous State of Knowledge“ in dem sie versuchen die Worte und Konzepte, die für sie selbst ihr Informationsbedürfnis beschreiben, auf die Begrifflichkeit des Systems abzubilden. Dabei werden sie mit den Unzulänglichkeiten der Sprache konfrontiert: Mehrdeutigkeiten und auch Vagheit. Dieses Problem tritt vermehrt in hochspezialisierten Bereichen wie der wissenschaftlichen Literatursuche auf, die meist nur auf bibliografische Metadaten als Suchbasis zurückgreifen kann.

Wie Petras (2006) zeigt, entwickeln sich in wissenschaftlichen Fachbereichen sehr spezielle Terminologien, die den wissenschaftlichen Diskurs widerspiegeln und sich mit ihm weiterentwickeln. Diese „different language spaces or dialects (specialties) within an



information system“ (Petras, 2006, S. 27), also Dialekte des wissenschaftlichen Diskurses, sind nicht zwangsläufig deckungsgleich mit dem dokumentarischen Vokabular, das ein menschlicher Indexierer oder Information-Professional verwenden würde (z. B. in Form von Thesauri). In ihrer Arbeit verwendeten sie zwei Dokumentenkorpora: Inspec, der die Fachgebiete Physik und Ingenieurwesen abdeckt und Ohsumed, eine Untermenge von Medline, die medizinische Dokumente beinhaltet. Sie konnte zeigen, dass maschinelle Verfahren zur Anfrageerweiterung davon profitieren, wenn sie die Besonderheiten fachlicher Terminologie berücksichtigen. In ihren Versuchen mit diesen beiden Datenbeständen, konnten die spezialisierten Methoden eine Steigerung der Anzahl der perfekten Ergebnisse von 0,43 % auf 0,60 % für Inspec und von 6,36 % auf 10,26 % erreichen. Das Verfahren zur automatischen Abbildung (Mapping) von Terminologien und Vokabularen nannte sie den *Search Term Recommender* (STR).

Neben den maschinellen Verfahren gibt es auch intellektuelle Ansätze, die versuchen Vokabulare aufeinander abzubilden. Mittels sogenannter Crosskonkordanzen soll, z. B. im Szenario fachübergreifender Datenbanken, eine Abbildung der unterschiedlichen Fachvokabulare aufeinander ermöglicht werden. Die Arbeit von Keil (2012) zeigt einen aktuellen Stand der Grundlagen und aktuelle Standardisierungsbestrebungen auf ISO-Ebene in diesem Bereich auf. Ein Beispiel Verknüpfungen auf technischer Ebenen zu ermöglichen wird in Arbeiten von Zapilko u. Sure (2009) dargestellt, die den vorhandenen Thesaurus Sozialwissenschaften in das SKOS-Format übertragen. Dies ermöglicht es, einen Thesaurus mittels definierter Relationen wie `skos:closematch`, `skos:exactmatch`, `skos:broadmatch`, `skos:narrowmatch` oder `skos:relatedmatch` mit anderen Ressourcen zu verknüpfen.

Unabhängig von der eingesetzten Technologie können intellektuelle oder maschinelle Verfahren genutzt werden, um das Problem der Sprachvagheit zu reduzieren und eine Nutzeranfrage automatisch oder interaktiv auf das entsprechende Vokabular der Informationssysteme abzubilden (Mayr u. Petras, 2008). Es wird ein Instrument benötigt, das die unterschiedlichen Vokabulare ineinander übersetzt. Ein solches Werkzeug zur Abbildung ist besonders im bereits beschriebenen Szenario der wissenschaftlichen digitalen Bibliotheken notwendig. Nutzerstudien in digitalen Bibliotheken haben gezeigt, dass viele Nutzer nichts von speziellen kontrollierten Vokabularen wissen, die zur Indexierung der Dokumente verwendet werden (Shiri u. Revie, 2006) und die somit auch nicht zur Unterstützung und Präzisierung des Suchvorgangs eingesetzt werden können.

Basierend auf den Arbeiten von Petras setzen Schaer u. a. (2011) zur Verringerung des Sprachproblems im IR einen STR ein, der auf zwei Komponenten basiert: (1) einem kontrollierten Vokabular, das genutzt wurde, um ein Dokument zu beschreiben und (2) Begriffen der natürlichen, freien Sprache, die von einem Nutzer zur Anfrage an ein Informationssystem verwendet werden. Durch die Abbildung der beiden Vokabulare aufeinander, können auch bei einer Suchanfrage ohne entsprechende Fachbegriffe passende Thesaurusterme zur Anfrageerweiterung verwendet werden. Ein STR ist ein System, das mit Verfahren der Kookkurrenzanalyse (s. Abschnitt 6.2) arbeitet und hierbei mit freien Begriffen aus Titel und Abstract sowie kontrollierten Begriffen aus einem Thesaurus trainiert wird. Aus den Wortassoziationen wird ein Wörterbuch zwischen zwei Vokabularen hergestellt: (1) der Sprache des wissenschaftlichen Diskurses, die Autoren in den Titeln der

Arbeiten bzw. den Abstracts verwendet haben und (2) der Sprache der Dokumentation, die von Indexierern vergeben wurden.

## 3.2 Polyrepräsentation

Die Ausnutzung der in Datenbeständen vorhandenen Heterogenität steht beim Prinzip der sogenannten Polyrepräsentation nach Ingwersen (1996) im Vordergrund, wonach Dokumente in Retrieval-Systemen häufig polyrepräsentiert bzw. sehr heterogen in ihrer Erschließung und Repräsentation sind. Eine der Kernideen der Polyrepräsentation ist, dass entgegen des traditionellen Bag-of-Words-Ansatzes, bei dem nur eine term-basierte Suche durchgeführt wird, auch die Struktur der Dokumente oder ganz generell zusätzliche Daten über die Dokumente (in Form von Metadaten) genutzt werden soll. Diese zusätzlichen Daten, die neben den eigentlichen Dokumenteninformationen wie Titel, Abstract oder Volltext erhoben werden, sind bspw. intellektuell oder maschinell vergebene Deskriptoren, Strukturinformationen wie Seitenangaben, Dokumentenart, Umfang oder gänzlich andere Werkzeuge der bibliografischen Erschließung. Die Polyrepräsentation beschreibt die Ausnutzung dieser mehrfachen und teilweise redundanten Repräsentation von Dokumenten für das Information Retrieval.

Die Polyrepräsentation (*polyrepresentation*) ist ein konzeptionelles Modell, das auf Ingwersen (1996) zurückgeht. Larsen (2004) fasst die Polyrepräsentation wie folgt zusammen:

„In short, the theory of polyrepresentation . . . 'hypothesises that overlaps between different cognitive representations of both users' information needs as well as documents can be exploited for reducing the uncertainties inherent in IR [information retrieval], and thereby improve the performance of IR systems'“ (Larsen, 2004, S. 24)

Anders beschreibt Bachelier (2001) das Konzept:

„Im Gegensatz zur Redundanz soll unter Polyrepräsentation die mehrfache und jeweils unterschiedliche Repräsentation eines Informationsobjektes verstanden werden. Es wird dadurch eine Effektivitätssteigerung angestrebt, da bei jeder der Einzelrepräsentationen unterschiedliche Perspektiven fokussiert werden können, die durch ein einzelnes Modell nicht bzw. nicht in diesem Umfang repräsentierbar sind. Weiterhin soll damit im Kontext eines IRS Unsicherheiten bezüglich der Repräsentationen und die Diversität von Agenten modellierbar und somit handhabbar werden.“ (Bachelier, 2001, S. 43)

Nach Larsen u. a. (2006) wurde das Konzept im Laufe der 1990er Jahre entwickelt und umfassend im Journal of Documentation 1996 vorgestellt. Wie Hjørland (2006) feststellt, ist der Hintergrund des Konzeptes aber wesentlich älter.

Im Konzept der Polyrepräsentation wird davon ausgegangen, dass es *kognitive* und *funktionale* Merkmale von Dokumenten gibt, die in unterschiedlichen Repräsentationen vorliegen. So beschreibt bspw. ein Autor sein Dokument, indem er ihm einen seiner

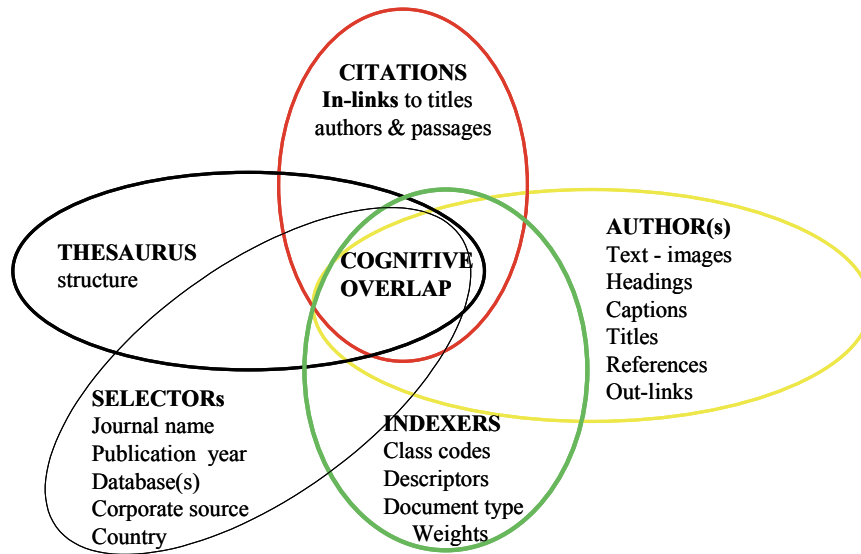


Abbildung 3.2: Das Prinzip der Polyrepräsentation in wissenschaftlichen Dokumenten, dargestellt durch die Überlappung kognitiven und funktionalen Ausprägungen von Informationseinheiten (entnommen aus Larsen u. a., 2006).

Meinung nach aussagekräftigen Titel gibt und hierzu sein persönliches Fachvokabular verwendet. Ein Indexierer einer Fachdatenbank vergibt zusätzlich kontrollierte Begriffe, um das Dokument zu beschreiben. Die beiden Akteure, Autor und Indexierer, beschreiben zwar dasselbe Informationsobjekt (das vorliegende Dokument), verwenden jedoch unterschiedliche Begriffe aufgrund ihrer kognitiv unterschiedlichen Einschätzung. Die Anwendung der Polyrepräsentation von sprachlichen Eigenschaften liegt folglich nahe zum bereits beschriebenen Vagheitsproblem im IR (s. Abschnitt 3.1). Neben der kognitiven Ebene gibt es im Konzept auch die funktionale Ebene. So können z. B. Autorennamen im Fließtext eine andere (funktionale) Bedeutung haben, als die gleichen Namen im Literaturverzeichnis. Ziel des Konzeptes ist es nun, die vorhandene kognitive und funktionale Polyrepräsentation gezielt auszunutzen, um das Retrieval zu verbessern. Hierzu sind besonders die Bereiche von Interesse, in denen sich die verschiedenen Repräsentationen überlappen (s. Abb. 3.2):

„...the more interpretations of different cognitive and functional nature, based on an IS&R situation, that point to a set of objects in so-called cognitive overlaps, and the more intensely they do so, the higher the probability that such objects are relevant (pertinent, useful) to a perceived work task/interest to be solved, the information (need) situation at hand, the topic required, or/and the influencing context of that situation.“ (Ingwersen u. Järvelin, 2005, S. 208)

Ausgehend vom Prinzip der Polyrepräsentation, das laut Larsen (2005, S. 22) und Ingwersen (1996, S. 18) auch die kognitive Wahrnehmung und die Erwartungen eines

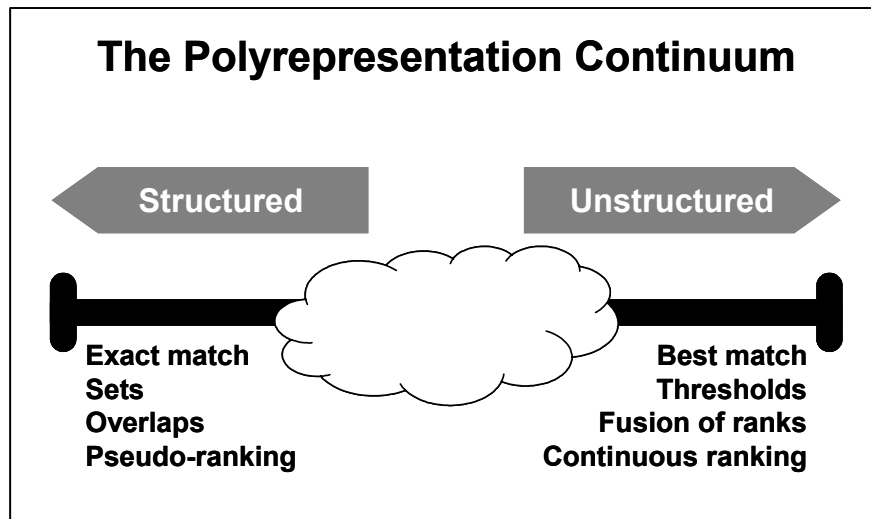


Abbildung 3.3: Das Polyrepräsentationskontinuum (entnommen aus Larsen u. a., 2006, S. 92).

Suchenden umfasst, können verschiedene Aspekte bei der Suche in Betracht gezogen werden: „polyrepresentation of the seeker’s cognitive space is to be achieved by extracting a number of different representations from the seeker“. Die Autoren listen dabei explizit drei verschiedene Repräsentationen bzw. Intentionen auf, die den Kontext einer Informationssuche beschreiben:

- Die Frage nach dem Was („what“), d. h. eine Frage, die das bisher Bekannte über das Unbekannte enthält (das Informationsbedürfnis);
- das Warum („why“), das eigentliche Problem, bzw. die Motivation, die das Informationsbedürfnis ausgelöst hat;
- eine Beschreibung der Arbeitsaufgabe und des Umfeldes („work task and domain description“).

Vergleicht man dies mit den Erkenntnissen, die aus Benutzerstudien und Log-Datei-Analysen entstanden sind (s. Abschnitt 2.5), so sieht man, dass diese Annahmen um weitere, spezifischere Fragen erweitert werden können. So ist die Suche nach einem „Wer“ für fast 1/3 der Nutzer von Bedeutung. Dies kann als das Bedürfnis nach einer zusätzlichen Informationshilfe bei der Anfrageformulierung verstanden werden. Welche Autoren oder Journale kann der Benutzer seiner Suchanfrage hinzufügen, um das Ergebnis zu verbessern? Es scheint intuitiv sinnvoll unterschiedliche Ausprägungen eines Dokuments gezielt auszunutzen und so die Nachteile eines reinen Bag-of-Words-Ansatzes auszugleichen.

Das Konzept konnte bisher u. a. erfolgreich im Bereich des *Fusion of Ranks/Data Fusion* (s. Abb. 3.3) von Larsen u. a. (2009) sowie innerhalb einer *Best-Match*-Retrievalumgebung (Skov u. a., 2008) umgesetzt werden. Beide Studien nutzten traditionelle

IR-Testkollektionen und eine TREC-ähnliche Evaluationsumgebung. Die Verbesserungen für das Retrieval wurden mit Evaluationsmaßen wie *Precision*, *Recall*, *Cut-off-Precision* und *Cumulative Gain* (s. Erläuterungen in Abschnitt 5.2) gemessen. Während bei Skov nur eine verhältnismäßig kleine Dokumentmenge von 1 239 Dokumenten mit 29 dazugehörigen Topics untersucht wurde, betrachtete Larsen den größeren Ad-hoc-Korpus aus TREC 5 mit 30 Topics. Beide Studien berichten von einer signifikanten Verbesserung (gemessen mittels eines Friedman Tests mit  $\alpha = 0,05$ ) durch die Kombination von bis zu vier verschiedenen IR-Modellen.

White (2006) nutzt das Prinzip der Polyrepräsentation aus, um Entscheidungen hinsichtlich des Interaktionsdesigns zu fällen. Er argumentiert, dass inhaltsreiche Suchoberflächen und die Präsentation zu vieler Meta-Informationen innerhalb der Ergebnisliste zu einer kognitiven Informationsüberladung des Benutzers führen können. Mittels verschiedener Repräsentationen der gefundenen Dokumente, die gleichzeitig angezeigt wurden, konnte er ein Software-Tool entwickeln, das bei der Simulation des Benutzerverhaltens eingesetzt wurde. Das Prinzip der Polyrepräsentation war hierbei die konzeptionelle Basis und gab „potential to [...] designers by giving them information about the potential impact of their design decisions [...]“.

Erste Arbeiten im Bereich der Anfrageerweiterung haben gezeigt, dass die Polyrepräsentation sich auch auf die Domäne der Anfrageerweiterung anwenden lässt (Schaer u. a., 2012c).

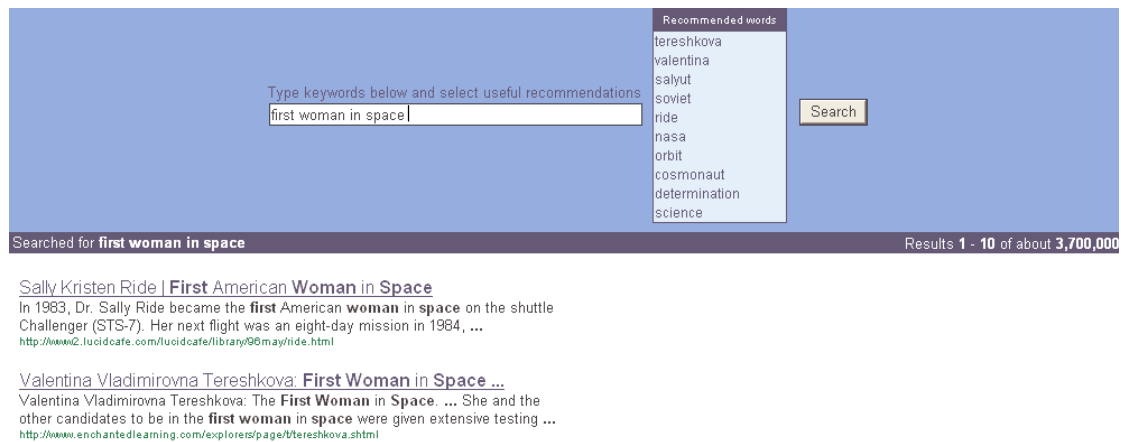
### 3.3 Textuelle Anfrageerweiterung

Anfrageerweiterung beschreibt den Prozess, eine Anfrage zu verändern bzw. anzureichern, um die Retrievalleistung eines Suchprozesses zu verbessern (Efthimiadis, 1996). Wenn die Erweiterung mit textuellen Attributen wie z. B. Thesaurustermen geschehen soll, bietet sich der Einsatz eines sogenannten STR an. Er kann verwendet werden, um eine ursprüngliche Suchanfrage um spezialisierte kontrollierte Terme zu erweitern. Grundsätzlich kann bei allen Arten der Anfrageerweiterung zwischen zwei Ausprägungen unterschieden werden: Der manuellen bzw. interaktiven und der automatischen Anfrageerweiterung.

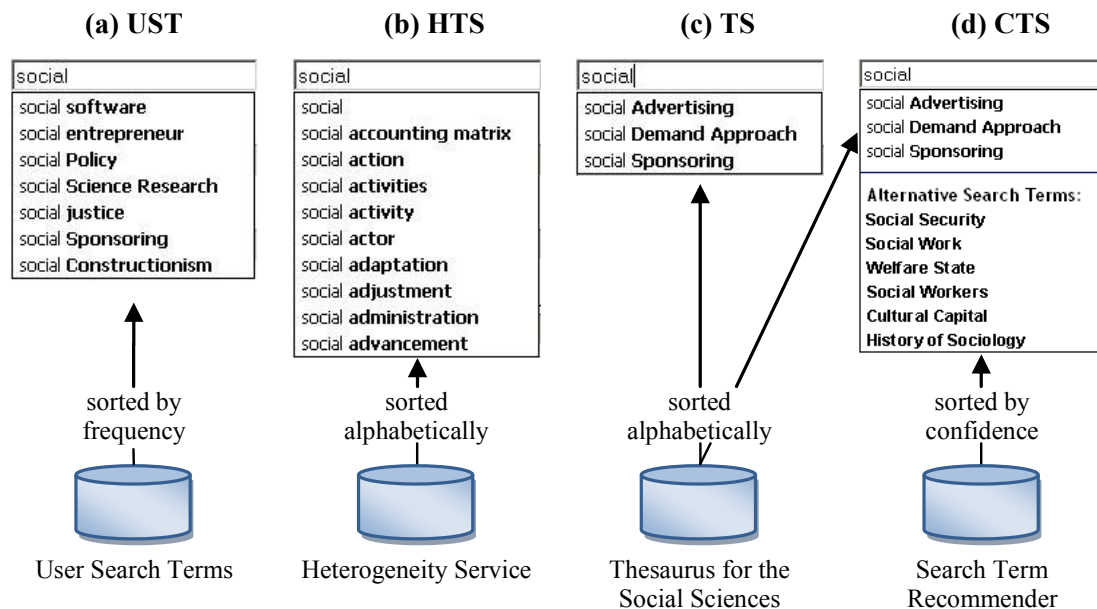
Die interaktive Variante der Anfrageerweiterung (*interactive query expansion* oder *interactive term suggestion*) wird generell als probates Mittel angesehen, um den Suchprozess von Benutzern zu unterstützen. So wird durch die (pro-)aktiven Vorschläge des Suchsystems die kognitive Last des Benutzers bei der Formulierung der Suchanfrage gesenkt, da ihm Formulierungsvorschläge als Entscheidungsgrundlage dienen können. Anstatt sich die Suchbegriffe selbst auszudenken oder zu abstrahieren, kann im besten Fall ein passender Begriff aus einer kurzen Liste von Suchtermvorschlägen ausgewählt werden. Ein solches System wurde u. a. von White u. Marchionini (2007) vorgestellt und evaluiert. Es handelte sich dabei um eine Form der Echtzeit-Anfrageerweiterung (*real-time query expansion*, RTQE), bei der schon während des Tippens der einzelnen Suchbegriffe dynamisch weitere Suchterme vorgeschlagen wurden (s. Abb. 3.4a).

Zusätzlich zu den positiven Effekten in der reinen Retrievalleistung durch den Einsatz von fachspezifischen Vokabularen und der verringerten kognitiven Last, können interaktive

### 3 Anfrageerweiterung und Recommendersysteme



- (a) Interaktive Anfrageerweiterung mit Termvorschlägen, die in Echtzeit vom System generiert werden (entnommen aus White u. Marchionini, 2007).



- (b) Vier im Portal Sowiport implementierte Termvorschlagsdienste: (a) Termvorschläge aus einer Log-Dateianalyse (UST), (b) intellektuell auseinander abgebildete Begriffe aus dem Heterogenitätsservice von GESIS (HTS), (c) Begriffe aus dem Thesaurus Sozialwissenschaften (TS) und (d) eine Kombination von Thesaurusbegriffen und per Kookkurrenzanalyse ermittelter Begriffe (CTS). Abbildung entnommen aus Hienert u. a. (2011a).

Abbildung 3.4: Zwei unterschiedliche Implementierungen eines interaktiven Search Term Recommenders.

Anfrageerweiterungen weitere Einflüsse auf den Suchprozess haben. So kann durch die Termvorschläge ein neues Konzept in die Suche eingebracht werden, das so zunächst vom Benutzer nicht bedacht wurde. In diesem Szenario könnte ein Benutzer durch einen Termvorschlag veranlasst werden, seine Suche anzupassen und somit den „anchoring bias“ (Blair, 2002) abzuschwächen. Dieses Prinzip beschreibt die menschliche Tendenz, sich bei der Suche nach Informationen oder ganz allgemein bei Entscheidungsprozessen zu sehr auf einen einzigen Aspekt oder ein Konzept zu fixieren. Diesem kognitiven Effekt und der menschlichen Eigenart „nicht links, noch rechts“ zu schauen, kann somit entgegengewirkt werden, indem durch die Termvorschläge zusätzliche Suchkonzepte oder -varianten vorgeschlagen werden.

Hienert u. a. (2011a) stellen eine interaktive Termerweiterung auf Grundlage der STR-Implementierung von Schaer u. a. (2011) vor. In dem beschriebenen System wurden insgesamt vier unterschiedliche Verfahren miteinander verglichen und mit Hilfe von jeweils 1 000 Einzelpersonen pro Verfahren evaluiert. In der Untersuchung konnte gezeigt werden, dass nur eine kleine Anzahl an Benutzern direkt von vorgeschlagenen Thesaurustermen profitierte und eine Kombination von lexikografisch- und statistisch-generierten Termvorschlägen den größten Zuspruch fand (s. Abb. 3.4b).

Auch bei der automatischen Anfrageerweiterung (*automatic query expansion*) ist die Erweiterung mit textuellen Dokumentelementen wie Deskriptoren oder Synonymen, die zusätzlich ggf. noch linguistische Techniken der Wortdisambiguierung verwenden, der Stand der Technik. Der gleiche STR, der für die interaktive Termerweiterung genutzt wurde, konnte in einem Szenario der automatischen Anfrageerweiterung die Retrievalleistung eines Systems von einer Average Precision von 0,57 auf 0,68 steigern, was einer Verbesserung von 21,4% entspricht (Schaer u. a., 2010b).

Diese Art der Erweiterungen bewegen sich im Rahmen dessen, was als gängige Praxis in der Anfrageerweiterung angenommen wird. Efthimiadis (1996) definiert als besondere Merkmale und Ausprägungen der automatischen Anfrageerweiterung drei Mechanismen:

- Solche, die auf Wissen aus der Dokumentkollektion basieren;
- solche, die auf unabhängigen Wissensstrukturen basieren;
- und solche, die auf der Dokumentmenge einer vorherigen Suche basieren.

Für ihn liegen die Unterschiede in den verschiedenen Implementierungen, d. h. primär in der Quelle der Daten, die für die Anfrageerweiterung genutzt werden. Die Terme, die für die Erweiterung genutzt werden, können entweder von externen Wissensstrukturen wie z. B. Thesauri oder aus den Dokumenten selbst stammen. Diese verschiedenen Quellen haben Vechtomova u. a. (2003) und Vechtomova u. Karamuftuoglu (2004) gegeneinander evaluiert und dabei das gleiche Verfahren, das auf Kookkurrenzanalyse basierte, verwendet. Zum Einsatz kamen sowohl ein globaler Ansatz, der die gesamte Dokumentenkollektion analysierte als auch ein lokaler Ansatz, der nur auf der Dokumentmenge einer initialen Suche basierte. Das lokale Verfahren konnte den globalen Ansatz eindeutig übertreffen, da beim globalen Ansatz zu viele allgemeine und unspezifische Terme für die Erweiterung hinzugezogen wurden. Ein ähnliches Verhalten konnte in Versuchen von Lücke u. a. (2012)

reproduziert werden. Sie verwendeten hierzu die sozialwissenschaftliche Datenbank SOLIS. Auch hier waren für jeweilige Unterklassen der Datenbank trainierte Recommender im Szenario der automatischen Anfrageerweiterungen effektiver, als die allgemeinen Verfahren. Im Vergleich von 16 spezialisierten und einem generellen Recommender konnte die Mean Average Precision von 0,155 auf 0,179 und die Precision@10 von 0,509 auf 0,601 signifikant gesteigert werden, dies allerdings unter der Voraussetzung, dass der jeweils passende Recommender ausgewählt wurde. In einem automatisierten Verfahren bliebe die Schwierigkeit das jeweils passende Verfahren auszuwählen. Ein erster Ansatz zur automatischen Recommenderauswahl konnte war nicht an die theoretisch besten Werte heranreichen, doch war auch hier eine Steigerung der Retrievalqualität zu verzeichnen.

Bei beiden Untersuchungen zeigte sich das Verhalten des sogenannten *Query-Drift*, bei dem zwar semantisch relevante, aber trotz allem fehlleitende Begriffe für die Anfrageerweiterung verwendet wurden. Bei interaktiven Systemen kann zwar der Nutzer einschreiten und die irreführenden Begriffe manuell aussortieren, jedoch kann es zu Komplikationen kommen, wie Studien von Schatz u. a. (1996) zeigen. Sie evaluierten zwei verschiedene Termvorschlagsdienste gegeneinander, die in einem interaktiven Szenario einmal Thesaurusterme und einmal freie Terme aus einer Kookkurrenzanalyse vorschlugen. In beiden Fällen wurden die zu unspezifischen Vorschläge kritisiert. Die Kombination der beiden Verfahren zeigte aber eine Verbesserung, die zu einer erhöhten Retrievalqualität führten. Auch bei Bhogal u. a. (2007) zeigten sich die zuvor beschriebenen positiven Effekte einer Kombination unterschiedlicher Verfahren und Datenquellen.

## 3.4 Nicht-textuelle Anfrageerweiterungen


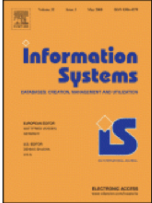
Wie wir gesehen haben, ist die interaktive Empfehlung bzw. die automatische Anfrageerweiterung in heutigen Informationssystemen eine gebräuchliche Funktion. Im vorherigen Abschnitt wurden einige Systeme vorgestellt, die textuelle Vorschläge machen, z. B. auf Grundlage von Thesaurustermen.

Eng verbunden mit dem Thema der Anfrageerweiterung sind sogenannte Recommender-Dienste, wie sie für digitale Bibliotheken z. B. von Geyer-Schulz u. a. (2003) vorgestellt wurden. Bei diesen für die Universitätsbibliothek Karlsruhe entwickelten Systemen wurde mittels einer Log-Datei-Analyse ein damaliges Legacy-OPAC-System um eine Recommender-Funktion erweitert. Ihr Ansatz basierte auf der Repeat-Buying-Theorie und war in der Lage neue Dokumente auf Grundlage von bereits bekannten Dokumenten vorzuschlagen („others also use...“). Anstelle von Dokument-Empfehlungen wurde von Heck u. a. (2011) ein Autoren-Empfehlungsdienst entwickelt, der mit Hilfe von Nutzungsinformationen und Zitationsdaten des sozialen Netzwerks CiteULike für die eigene Arbeit relevante Autoren vorschlagen kann.

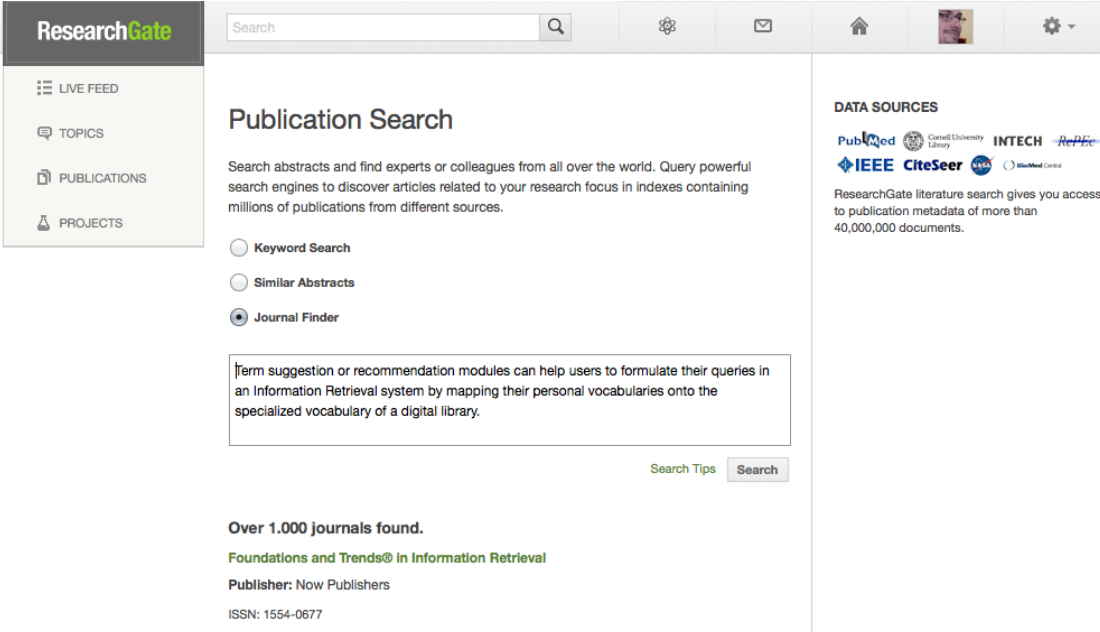
Eine klare Vorstellung davon wie der aktuelle Stand der Technik in tatsächlichen Bibliothekssystemen ist, bekommt man bei der Betrachtung des Journal Recommender der TH Wildau (Keidel u. Müller, 2010), einen Mehrwertdienst der Hochschulbibliothek Wildau. Im Rahmen des Journal Recommender werden spezielle fachliche Zugänge für Studenten der Hochschule angeboten, so z. B. für die Fachgebiete Maschinenbau, Soziologie oder



### 3.4 Nicht-textuelle Anfrageerweiterungen

<b>Informationssysteme</b> <small>Ansprechpartner(innen): Prof. Dr. Gerd Dehnert, Prof. Dr. Michael Hendrix, Prof. Dr. Christian Müller,</small>				<a href="#">(nach oben)</a> <a href="#">(Volltexte)</a> <a href="#">(Hilfe)</a> <a href="#">(Inhalte ändern)</a>
Titel	ISSN	Archiv	Bemerkung	(© der Anbieter gilt für Texte und Cover)
<b>Informatik-Spektrum</b> <small>Anmerkung der Bibliothek: Springer Berlin / Heidelberg</small>	1432-122X	Springer Basic	Computer Science Computer Science, general Computer Hardware Computer Systems Organization and Communication Networks Software Engineering/Programming and Operating Systems Data Structures, Cryptology and Information Theory	
<b>Information Systems</b>	03064379	Elsevier Social	Information systems are the software and hardware systems that support data-intensive applications. Information Systems publishes articles concerning the design and implementation of languages, data models, algorithms, software and hardware for information systems. Subject areas include data management issues as presented in the principal international database conferences (e.g. ACM SIGMOD, ACM PODS, VLDB and EDBT) as well as data-related issues from the fields of data mining, information retrieval, natural language processing, internet data management, visual and audio information systems, scientific computing, and organisational behaviour. All papers should motivate the problems they address with compelling examples from real or potential applications. Systems papers must be serious about experimentation either on real systems or simulations based on traces from real systems. Papers from industrial organisations that discuss the practical	

- (a) Journal Recommender der TH Wildau für den Bereich Informationssysteme, der händisch gepflegt und redaktionell betreut wird.



The screenshot shows the ResearchGate 'Publication Search' interface. It features a search bar at the top, a navigation menu on the left with options like 'LIVE FEED', 'TOPICS', 'PUBLICATIONS', and 'PROJECTS', and a main content area. The main area displays search options: 'Keyword Search', 'Similar Abstracts', and 'Journal Finder' (which is selected). Below these options is a text box containing a snippet of text: 'Term suggestion or recommendation modules can help users to formulate their queries in an Information Retrieval system by mapping their personal vocabularies onto the specialized vocabulary of a digital library.' At the bottom of the main area, it states 'Over 1.000 journals found.' and lists 'Foundations and Trends® in Information Retrieval' as a result, with publisher 'Now Publishers' and ISSN '1554-0677'. On the right side, there is a 'DATA SOURCES' section listing various databases like PubMed, IEEE, and CiteSeer.

- (b) Journal Finder des sozialen Netzwerks ResearchGate, der anhand eines kurzen Abstracts passende Zeitschriften vorschlagen kann. Hierzu werden über 40 Millionen Dokumente ausgewertet.

Abbildung 3.5: Gegenüberstellung zweier wissenschaftlicher Recommendersysteme (Screenshots der Systeme vom 26. Juli 2012).

Wirtschaftsinformatik. Diese sind wiederum in einzelne Lehrgebiete und Spezialisierungen wie Datenbanken, E-Learning oder Informationssysteme unterteilt. Für jedes Untergebiet werden eine Reihe händisch gepflegter Journal-Empfehlungen aus den Beständen der Bibliothek zusammengetragen. Das System besitzt eine ausgearbeitete Administrationsoberfläche und ist erweiterbar, doch basiert die eigentliche Empfehlungsleistung rein auf menschlichen Redakteuren (s. Abb. 3.5a).

Ein ähnliches System ist innerhalb des wissenschaftlichen sozialen Netzwerks ResearchGate<sup>1</sup> implementiert: Der Journal Finder (s. Abb. 3.5b). Die Grundlage des Systems bilden 40 Millionen Dokumente, die innerhalb des Netzwerks vorgehalten oder referenziert werden. Eine Ähnlichkeitssuche erlaubt nach Eingabe einiger Schlagworte oder eines Abstracts den Vorschlag eines thematisch passenden Journals durch das System. Diese ist zum einen für Autoren konzipiert, die für ein verfasstes Papier ein geeignetes Publikationsorgan wie einem Journal suchen, oder für Autoren, die nach alternativen Journalen für ihre Forschungsinteressen suchen.

Diese Arbeiten fallen in einen Bereich, der als Expertensuche (*Expert-Search*) oder allgemein als *Recommender* bekannt ist. Die generelle Idee hinter diesen Verfahren ist es, Wissenschaftler während der Suche zu unterstützen, indem ihnen die Namen oder die Arbeiten von Kollegen (*Peers*) vorgeschlagen werden. Dabei ist es nicht zwingend notwendig, dass die Kollegen dem Suchenden bekannt sind, vielmehr ist es ein oft gewünschtes Verhalten des Systems, wenn neue Arbeiten bzw. Namen vorgeschlagen werden, um den *Serendipity-Effekt* zu unterstützen (Gollapalli u. a., 2011).

Basierend auf den Learning-to-Rank Ansätzen, wie sie in Abschnitt 4.5.2 vorgestellt werden, wird in Arbeiten von Moreira u. a. (2011) ein Verfahren vorgestellt, das die Expertensuche unterstützt. Hierbei wird versucht in einem Dokumentenkörper verschiedene Methoden zur Identifizierung von Experten, d. h. einschlägigen Personen in der Wissenschaftsgemeinde, miteinander zu kombinieren. Diese Arbeiten reihen sich in die recht neue Disziplin der sogenannten entitätszentrierten Suche (Ernde u. a., 2010; Demartini u. a., 2012) ein. Bei dieser steht nicht die dokumentenzentrierten Suche im Mittelpunkt, sondern die Suche nach anderen Entitäten wie Personen oder Lokalitäten.

Nahezu alle der hier vorgestellten Systeme könnten interaktiv genutzt werden und so während der Suche als Mehrwertdienst genutzt werden. Die extrahierten Autoren, Journale etc. werden allerdings weder in wirklichen Systemen für die interaktive noch die automatisierte Anfrageerweiterung eingesetzt (Ausnahmen sind die Systeme aus Abb. 3.5a).

## 3.5 Zusammenfassung

Die Anfrageerweiterung ist eine Standardtechnologie des modernen Information Retrieval und hat sich sowohl für automatischen als auch interaktive Anwendungen in Retrievalsystemen etabliert (s. Übersicht in Abb. 1.1). Trotz der langen Historie an Erfolgen nutzen viele Verfahren der Anfrageerweiterung im Bereich der digitalen Bibliotheken nicht die Fülle an Möglichkeiten aus, die sich aus den so generierbaren Metadaten ergeben würden.

---

<sup>1</sup><http://www.researchgate.net>

Vor dem theoretischen Hintergrund des Sprachproblems im IR und der Polyrepräsentation sollte eine Anfrageerweiterung diese zusätzlichen Möglichkeiten jedoch ausnutzen. Zwar werden in Verbindung mit dem Feld der Recommendersysteme teils sehr aufwändige und ausgefallene technische Systeme geschaffen, die aber nur zu einem Bruchteil ihren Weg in tatsächliche digitale Bibliotheken gefunden haben. Schon relativ simple Systeme wie der Journal Recommender der TH Wildau zeigen, dass Ansätze, die klar durch das Nutzungsinteresse motiviert sind, einen Mehrwert für Nutzer schaffen können.



## 4 Ranking und Relevanz

Wie im vorherigen Kapitel dokumentiert, sind traditionelle digitale Bibliotheken zwar mit einer Fülle an meist gut strukturierten Metadaten gefüllt, doch bestehen klare Defizite in einem nutzergerechten Retrieval innerhalb dieser Datenbestände. Im folgenden Kapitel werden daher zunächst die traditionellen Rankingverfahren aus dem Ad-hoc-Dokumentenretrieval als auch aktuelle Verfahren aus dem Web-Retrieval und verwandten Disziplinen aufgezeigt, um den aktuellen Stand der Technik zu dokumentieren. Die Struktur ist dabei so aufgeteilt, dass zunächst die vier traditionell textuell-orientierten Verfahren des booleschen Modells, des Vektorraummodells, des probabilistischen Modells und der statistischen Sprachmodelle vorgestellt werden. Im nächsten Abschnitt folgen eine Reihe primär nicht-textueller Rankingverfahren, die z. B. wie der PageRank auf Link-Analyse oder wie die Autorenzentralität auf einer Analyse der Ko-Autorenschaften basieren. Das Kapitel schließt mit einer Gegenüberstellung der beiden Ansätze des automatisierten und interaktiven IR unter Berücksichtigung der historischen Debatte zwischen Agentensystemen und direkter Manipulation.

In dieser Arbeit werden sogenannte *textuelle* und *nicht-textuelle* Verfahren und Konzepte unterschieden. Während in der Literatur nicht-textuell häufig mit multimedial gleichgesetzt wird, sind damit im Rahmen dieser Arbeit Eigenschaften gemeint, die keinen textuellen Bezug haben oder zumindest nicht textuell behandelt werden. Metzler (2011, S. 5) führt einige Beispiele für solche nicht-textuellen Eigenschaften auf, u. a. den PageRank, das Zählen von Inlinks einer Netzwerkanalyse, die Lesbarkeit eines Textes, die Wahrscheinlichkeit von Spam usw. Weitere Beispiele sind die Produktivität eines Autors oder die semantische Nähe zwischen einem textuellen Begriff (z. B. einem Suchterm) und einer Zeitschrift. Zwar sind viele Metadaten, die für die Suche genutzt werden, grundsätzlich als Text zu verstehen und können auch so durchsucht werden, doch wird in dieser Arbeit speziell auf ihre nicht-textuellen Eigenschaften, etwas solche, die aus einer informatrischen Analysen stammen (s. Kap. 6), eingegangen.

Eine sehr gute Einführungen in die allgemeine Thematik des IR und speziell des Rankings bieten die Arbeiten von Fuhr (2006) sowie Manning u. a. (2008), auf die hier explizit als zusätzliche Quellen hingewiesen werden soll.

### 4.1 Sortierung vs. Ranking

In diesem Abschnitt wird als kurzer Einschub die begriffliche Unterscheidung zwischen Sortierung und Ranking erläutert. Technisch betrachtet sind Sortierung und Ranking einer Dokumentmenge zunächst einmal das Gleiche: Die unsortierte Menge wird in eine geordnete Liste überführt.

## 4 Ranking und Relevanz

Ganz allgemein bedeutet dies, dass eine beliebige Dokumentmenge  $D$ , die aus einer Menge von Dokumenten  $d \in D$  besteht, in eine linear geordnete Liste  $L$  mittels einer Sortierfunktion  $f$  überführt wird, sodass gilt:

$$f : D \rightarrow L. \quad (4.1)$$

Das Ordnungskriterium ist dabei zunächst nicht von Bedeutung und frei wählbar. Dies kann eine lexikographische, alphabetische Ordnung nach Autorennamen oder eine numerische Ordnung nach Dokumenten-ID sein. Es kann aber auch nach einem abstrakten Konzept der Relevanz sortiert werden. Die Rankingfunktion  $r$  weist dabei jedem Dokument  $d$  einen Wert  $\tau$ , den sogenannten Relevanzwert (*relevance score*), zu. In Anlehnung an Lange u. a. (2010) ist die Rankingfunktion  $r$  wie folgt definiert:

$$r(d) \rightarrow \tau. \quad (4.2)$$

Die Relevanzwerte können geordnet werden, sodass gilt:

$$\tau_1 < \tau_2 \Rightarrow \tau_1 \text{ ist weniger relevant als } \tau_2. \quad (4.3)$$

Das endgültige Ranking ist dann die Ordnung der Menge  $D$  entlang der Relevanzwerte:

$$(\tau_1, \dots, \tau_n) \mid \forall \tau : \tau_{n-1} < \tau_n. \quad (4.4)$$

Es macht aus rein technischer Sicht keinen Unterschied, ob eine numerische Ordnung auf Grundlage von Dokumenten-IDs oder berechneten Relevanzwerten hergestellt wird. Die Begriffe werden auch in vielen Systemen und Lehrbüchern synonym verwendet. Ein Beispiel dafür sind die Screenshots der Systeme in Abbildung 2.7. Dem Benutzer wird angeboten, seine Ergebnisliste nach unterschiedlichen Kriterien zu sortieren, wobei dies sowohl einfache Kriterien wie die alphabetische oder numerische Sortierung, aber auch die Sortierung nach Relevanz sind. Relevanz ist der eigentliche Schlüsselbegriff, der ein simples Sortieren von einem Relevanzranking (*relevance ranking*) unterscheidet.

## 4.2 Der Relevanzbegriff

Saracevic (2007a,b) beschreibt in seiner Aufsatzreihe die Relevanz als den Kernbegriff der Informationswissenschaft und des Information Retrievals. Der Relevanzbegriff ist ein fundamentales Konzept des Information Retrievals, da die Hauptaufgabe eines IR-Systems darin besteht, relevante Informationen zu liefern (Song u. a., 2011). Die Definition der Relevanz ist vielschichtig und wurde von Borlund (2003) in Form einer ausführlichen Literaturstudie zusammengefasst. Sie führt ein allgemeines Modell der Relevanz in die Diskussion ein, das die verschiedenen Aspekte der Relevanzbewertung zusammenbringt, wie unterschiedliche Klassen, Typen, Grade, Kriterien und Ebenen der Relevanz. Borlund bezeichnet dies als die Multidimensionalität der Relevanz („*multidimensionality of relevance*“). Die allgemeinste Unterscheidungsform der Relevanz ist dabei die Aufteilung in objektive, bzw. systembedingte und subjektive, bzw. menschliche Relevanzkriterien.

Borlund (2003, S. 913) betrachtet vor allem die drei zentralen Schlussfolgerungen von Schamber u. a. (1990) über den Relevanzbegriff als solchen:

- Relevanz als ein *multidimensionales, kognitives Konzept*, das von der Wahrnehmung des Benutzers und dessen persönlichem Informationsbedürfnis abhängt;
- Relevanz als ein *dynamisches Konzept*, das von der Bewertung der Beziehung zwischen Information und Informationsbedürfnis zu einem bestimmten Zeitpunkt abhängt;
- Relevanz als ein *komplexes Konzept*, das trotz allem systematisch und messbar erfassbar sei, wenn es dabei aus der jeweiligen Nutzerperspektive betrachtet werde.

Alle drei Aspekte des Relevanzbegriffes eint die Abhängigkeit von einer Nutzerperspektive bzw. einem subjektiven Informationsbedürfnis. Da diese Sichtweise schlecht in Algorithmen und tatsächliche IR-Systeme umsetzbar ist, wird von Fuhr (2006, S. 12 f.) eine Aufteilung vorgeschlagen. Er fasst vier unterschiedliche Arten der Relevanz zusammen.

**Situative Relevanz:** Hiermit ist die tatsächliche Nützlichkeit einer Informationseinheit in einer konkreten Situation gemeint.

**Subjektive Relevanz (bzw. Pertinenz):** Die Nützlichkeit einer Informationseinheit für eine Person mit einem bestimmten Informationsbedürfnis.

**Objektive Relevanz:** Im Unterschied zur subjektiven Relevanz steht hier nicht die subjektive Einschätzung einer Person im Vordergrund, sondern mehrere Einschätzungen unterschiedlicher Personen. Die subjektiven Einschätzungen des Einzelnen sollen durch die Mehrfachbewertungen verschiedener sogenannter Assessoren, z. B. im Rahmen einer IR-Evaluation, abgeschwächt werden.

**Systemrelevanz:** Die algorithmische Berechnung der Relevanz durch ein IR-System. Hierbei handelt es sich nur um einen Schätzwert des Systems, der auch mit dem Begriff RSV (*Retrieval Status Value*) abgekürzt wird.

Die Aufgabe eines IR-Systems ist es nun, eine auf den berechneten RSV-Werten gerankte Liste für den Suchenden zusammenzustellen, gemäß seinen Ansprüchen an die situative und subjektive Relevanz.

White (2007a, S. 538 ff.) betrachtet den Relevanzbegriff aus einer etwas anderen Perspektive, da er sich der Relevanz über die Relevanztheorie nach Sperber und Wilson annähert, die aus der Position der linguistischen Pragmatik<sup>1</sup> heraus die Relevanz beurteilen. Vor diesem Hintergrund ist die Relevanz für White keine eindeutige Entscheidung, sondern das Verhältnis von kognitivem Effekt auf den Suchenden (*cognitive effect*) zum Verarbeitungsaufwand bzw. zur Verarbeitungsleichtigkeit (*ease of processing*). Diese menschliche Sicht auf die Relevanz (subjektive Sicht) wird nun von ihm auf die Systemsicht abgebildet, die mit Mechanismen wie TF\*IDF arbeitet. Er bildet dabei den

#### 4 Ranking und Relevanz

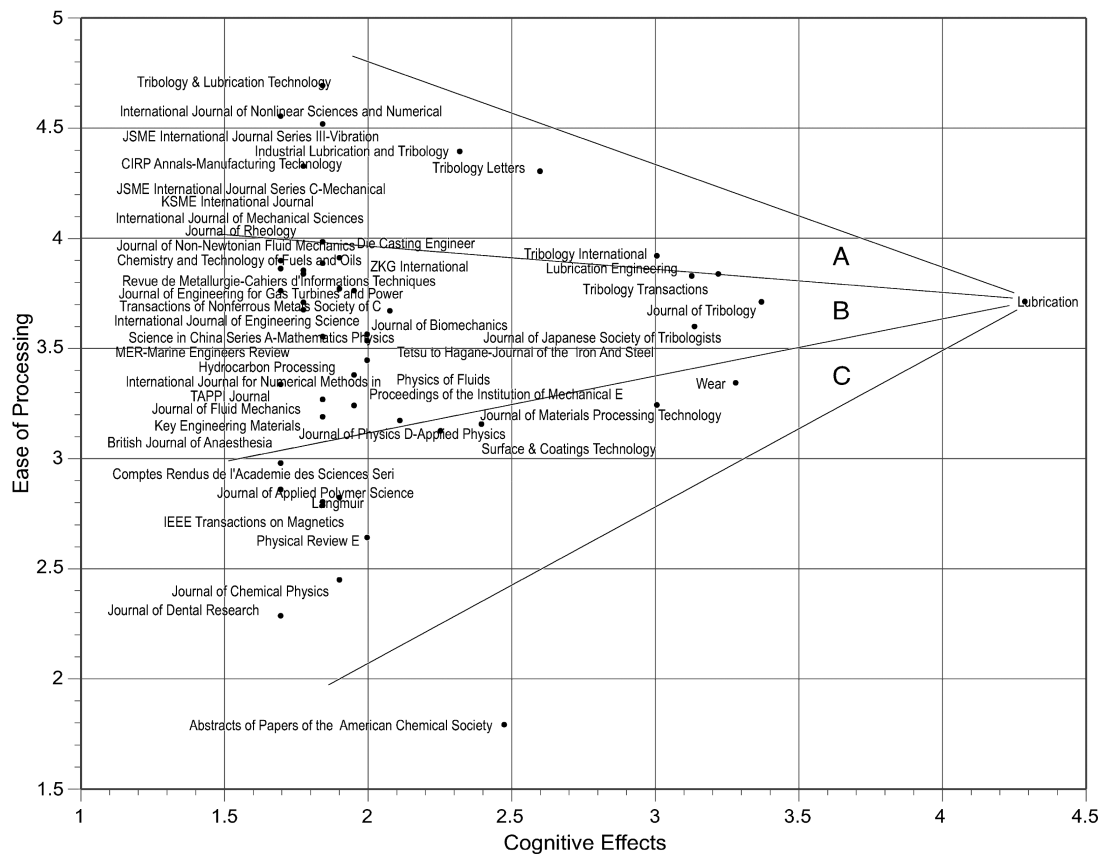


Abbildung 4.1: Ein Pennant-Diagramm zu Zeitschriften, die Artikel veröffentlichten, die das Wort „Lubrication“ im Titel enthielten (entnommen aus White, 2007b, S. 602).

kognitiven Effekt (d. h. Güte und Mehrwert für den Informationssuchenden) auf TF und die Verarbeitungsleichtigkeit auf IDF ab (s. Abb. 4.1).

Neben der Verknüpfung von kognitivem Effekt und Verarbeitungsleichtigkeit als Messgrößen für die Relevanzbeurteilung stellt White eine Visualisierung seiner These vor, die sogenannten *Pennant*-Diagramme. Es handelt sich dabei um doppelt logarithmierte Streudiagramme (*scatter plots*) der beiden Komponenten TF und IDF. Durch diese Art der Darstellung hat man den Vorteil, dass man die unterschiedlichen Komponenten, die zur Bildung des  $TF \cdot IDF$ -Wertes führen, besser überblicken kann. Das Beispiel-Pennant in Abbildung 4.1 zeigt einen Plot der unterschiedlichen Zeitschriften der SciSearch-Datenbank, die Artikel mit dem Wort „Lubrication“ im Titel enthalten. Durch die besondere Darstellung der Pennants lassen sich nun unterschiedliche Zonen A, B und C markieren. Die Zeitschriften in Zone A sind Fachzeitschriften, die als einschlägige Quellen

<sup>1</sup> „Die linguistische Pragmatik befasst sich mit denjenigen Aspekten sprachlicher Bedeutung, die vom (sprachlichen und situativen) Kontext abhängig sind.“ (Hagen, 2005)



	Systemsicht (gemessen)	Menschliche Sicht (nicht gemessen)
TF	Vorhergesagter kognitiver Effekt	Tatsächlicher kognitiver Effekt
IDF	Vorhergesagte Leichtigkeit der Verarbeitung	Tatsächliche Leichtigkeit der Verarbeitung
TF*IDF	Vorhergesagte Relevanz	Tatsächliche Relevanz

Tabelle 4.1: Abbildung von Systemsicht und menschlicher Sicht auf die Faktoren TF und IDF und den Relevanzeffekte nach White (2007a).

für dieses Forschungsgebiet angesehen werden können, wohingegen die Zeitschriften der Zone C sich nur peripher mit dem Thema „Lubrication“ auseinandersetzen. Dies ist analog zu den Beobachtungen von Bradford zu sehen, die in Abschnitt 6.1.4 genauer besprochen werden.

Festzuhalten aus der Pennantdiskussion ist, dass beim Einsatz von TF\*IDF die sogenannte „*topical relevance*“ überwiegt (White, 2007a, S. 547), was in vielen Anwendungsfällen auch genau der intendierte Effekt eines Informationssuchenden ist. Allerdings wird in der Diskussion klar, dass es auch andere Relevanzkriterien gibt, die zwar implizit auch im Pennant zu sehen sind. In einer typischen Ergebnisliste eines IR-Systems sind diese Feinheiten allerdings unter ersten 10, 20 oder 100 Treffern nicht zu erkennen. Durch die Einteilung in Zonen werden die unterschiedlichen Arten der Relevanz eher sichtbar und erlauben eine erweiterte Sicht auf das Themenfeld. White schließt daraus, dass der Einsatz von TF\*IDF die leicht erfassbare Relevanz bevorzugt (z. B. durch Gleichheit von Anfragetermen und Termen im Titel eines Dokuments).

Die Unterscheidung der zwei Begriffe Relevanz und Qualität ist zentrales Anliegen der Arbeiten von Mandl (2006, 93 ff.). Wie er ausführt sei für die Bewertung der Relevanz immer ein konkretes Informationsbedürfnis bzw. ein „Informationsproblem“ notwendig, zu dem eine konkrete Informationseinheit relevant oder nicht relevant sein könne. Bei Qualität handelt es sich hingegen um ein Attribut, das unabhängig von einem konkreten Informationsproblem zugewiesen werden kann. Allerdings ist eine Qualitätsbewertung nicht objektiv, eine Systemsicht wie beim Relevanzbegriff faktisch nicht umsetzbar. Subjektive Einschätzungen seien aber jederzeit auch ohne konkreten Informationsbedarf möglich. Trotzdem überlappen sich die Begriffe häufig in ihrer Bedeutung und Interpretation. Von den zuvor beschriebenen drei Ebenen der Relevanz nach Borlund ist das Qualitätsurteil dabei nicht zwangsläufig abhängig. Gerade der fehlende Bezug zu einem Informationsbedürfnis macht den Unterschied klar: So mag ein Dokument, das bereits bekannt ist, nicht mehr relevant sein, da es das Informationsbedürfnis nicht weiter befriedigen kann. Dennoch behält das Dokument trotz des verlorenen Neuigkeitswerts nach wie vor seine Qualität.

Gleichzeitig scheint der Relevanzbegriff ein grundlegendes Konzept menschlicher Kognition zu sein:

„People understood and understand relevance similarly over time, space, cultures, and domains. 'Nobody has to explain to users of IR systems what

## 4 Ranking und Relevanz

relevance is, even if they struggle (sometimes in vain) to find relevant stuff. People understand relevance intuitively.” (Saracevic, 2007b, S. 1918)

Jeder Mensch ist folglich intuitiv in der Lage ein Urteil über die Relevanz einer Informationseinheit, wie z. B. einem Dokument, zu fällen, wobei dabei immer sein persönliches Informationsbedürfnis im Vordergrund steht. Für ein IR-System besteht nun die Herausforderung darin, diese sehr subjektive Sicht auf ein Informationsproblem aus der System Sicht heraus zu beantworten.

Die vorangegangene Diskussion zeigt, dass die Relevanzbewertung und damit die Schlussfolgerungen für das Relevanzranking mehrschichtig und vage sind. Während der klassische TF\*IDF-Ansatz die themenbezogene Relevanz einfach und deutlich darstellen kann, vernachlässigt er latent relevante oder versteckte Konzepte, die erst durch alternative Verfahren zum Vorschein kommen. Für diese Arbeit ist es daher eine besondere Motivation, nach alternativen Relevanzkonzepten und deren Anwendungen bzw. deren Berechenbarkeit für das Information Retrieval zu suchen. Die nächsten Abschnitte widmen sich daher den unterschiedlichen Systementwürfen zur automatischen Relevanzbewertung durch IR-Systeme und dem dadurch angebotenen Relevanzranking.

### 4.3 Textuelles Ranking

Im folgenden Abschnitt werden die vier einflussreichsten Modelle, die den aktuellen Stand der Technik darstellen, skizziert und eingeführt. Die hier vorgestellten Verfahren sind in modernen Retrievalplattformen wie Apache Lucene/Solr<sup>2</sup>, der Terrier IR Plattform<sup>3</sup> oder dem Lemur/Indri Toolkit<sup>4</sup> verfügbar.

Die Modelle sind einer sogenannten Bag-of-Words-Annahme unterworfen, was bedeutet, dass alle Anfragen wie auch Dokumente nur als unzusammenhängende Menge an Wörtern verstanden werden. In dieser Menge von Wörtern sind die Wortreihenfolgen verloren gegangen, was zwar für die Verarbeitung der Anfragen und der Dokumentbasis enorme Vorteile, im Detail aber semantische Probleme mit sich bringt. Hinzukommen typische Probleme aus dem *Natural Language Processing* (NLP), u. a. Polyseme, Synonyme oder Wortkomposita. Generell ist die Verwendung der natürlichen Sprache mit Vagheit verbunden, sodass dieses Problem auch als das Vagheitsproblem (s. Abschnitt 3.1) bekannt ist.

#### 4.3.1 Boolesches Modell

Das boolesche Modell ist das älteste und meist-unterstützte Retrievalmodell, das auf der Verwendung der booleschen Algebra aufbaut. Eine Anfrage wird in Form eines booleschen Ausdrucks mit Hilfe der Operatoren UND, ODER sowie NICHT ausgedrückt. Durch eine Klammerung der Anfrageterme und der Operatoren lassen sich Anfragen dabei sehr lang und präzise formulieren. So kann mittels der booleschen Algebra eine

---

<sup>2</sup><http://lucene.apache.org/>

<sup>3</sup><http://terrier.org/>

<sup>4</sup><http://www.lemurproject.org/>

sehr große Ausdrucksstärke erreicht werden. Dies ist einer der Hauptgründe dafür, dass die boolesche Suche besonders bei geschulten Nutzern und Experten beliebt ist. Sie erlaubt es, sowohl eine präzise Anfrage zu stellen als auch das Ergebnis transparent und kontrollierbar zu gestalten (Manning u. a., 2008, S. 3 ff.). Charakteristisch für das boolesche Retrieval ist eine häufige Neu- und Weiterformulierung der Ursprungsanfrage, z. B. um Synonyme zur Suche hinzuzufügen. Boolesches Retrieval unterstützt nur exakte Anfragen, die auf dem *Exact-Match*-Verfahren basieren. Studien haben allerdings auch immer wieder gezeigt, dass die Mächtigkeit der booleschen Logik nur von wenigen Nutzern korrekt verstanden und angewendet wird, wie z. B. in einer Studie von Zhang (2008) zum Suchverhalten von Universitätsstudenten gezeigt. Ein Gros der Nutzer ist mit der korrekten Anfrageformulierung überfordert.

Die einfache technische Umsetzung mit Hilfe eines invertierten Index, die Mächtigkeit der Anfrageformulierung sowie die immense Verbreitung in unzähligen, digitalen Bibliotheken sind wohl die größten Vorteile des Verfahrens. Allerdings sind auch die größten Nachteile des Systems offensichtlich: Durch das *Exact-Match*-Verfahren werden viele relevante Dokumente nicht gefunden; die korrekte Anfrageformulierung ist nicht intuitiv und fehleranfällig. Darüber hinaus beinhaltet das klassische boolesche Retrieval kein Ranking. Lange Ergebnislisten können somit nur nach einfachen Dokumenteneigenschaften, wie in Abschnitt 4.1 beschrieben, sortiert werden. Da dies bei großen Dokumentmengen zu Problemen führt, wurde das boolesche Retrieval erweitert.

### 4.3.2 Erweitertes boolesches Ranking

Eine Erweiterung des Konzeptes ist als *Extended-Boolean-Retrieval* bekannt. Hierbei werden unterschiedliche Gewichte in das Modell, wie z. B. Termhäufigkeiten, eingebracht. Weiterhin können durch die Erweiterungen unterschiedliche, strukturelle Informationen der Dokumente besser verarbeitet werden. So kann unterschieden werden, ob Terme z. B. nur im Titel eines Dokuments vorkommen oder z. B. auch im Abstract. Gerade in Verbindung mit den in Abschnitt 2.4 beschriebenen reichen Beständen an Metadaten entsteht so eine mächtige Anfragesprache, die bis heute in vielen digitalen Bibliotheken Bestand hat. Eine der einfachsten Varianten, ein erweitertes boolesches Retrieval zu implementieren, ist es, die Anfrageterme nach der bereits erwähnten Termhäufigkeit der Anfrageterme in den Dokumenten zu gewichten. Die Termhäufigkeit wird auch Termfrequenz (*term frequency*) genannt und für einen Term  $t$  in einem Dokument  $d$  gilt:  $tf(t,d)$ . Je häufiger ein Anfrageterm in einem Dokument vorkommt, desto relevanter ist er und verschafft dem Dokument eine höhere Position in der Ergebnisliste.

Die Termfrequenz ist zwar ein guter Indikator für die Wichtigkeit von Termen in einem Dokument, doch steigt die Relevanz eines Textes nicht proportional mit der Häufigkeit von Termnennungen an. Um dieses Ungewicht auszugleichen, wird die inverse Dokumenthäufigkeit (*inverse document frequency*) hinzugefügt (Spärck Jones, 1972):

$$\text{idf}(t) = \log \frac{N}{1 + \text{df}(t)}, \quad (4.5)$$

wobei  $N$  die Anzahl der gesamten Dokumentmenge und  $\text{df}(t)$  die Anzahl der Dokumente

## 4 Ranking und Relevanz

ist, die den Term  $t$  enthalten. Um eine Division durch 0 zu verhindern, wird dem Divisor üblicherweise eine 1 hinzuaddiert. Beide Werte können zum sogenannten TF\*IDF-Gewicht (*term frequency inverse document frequency*) zusammengefasst werden:

$$\text{TF*IDF} = \text{tf}(t,d) \cdot \text{idf}(t) = \text{tf}(t,d) \cdot \log \frac{N}{1+\text{df}(t)}. \quad (4.6)$$

TF\*IDF weist einem Term  $t$  in einem Dokument  $d$  einen Wert zu, der abhängt vom unterschiedlichen Auftreten von  $t$  und  $d$ . Dieser Wert ist ein statistisches Maß, das zeigt wie wichtig ein Dokumententerm in einer Dokumentmenge ist. Der Wert ist am höchsten, wenn  $t$  häufig in einer geringen Anzahl an Dokumenten auftritt. Man spricht dabei von der diskriminierenden Kraft der Terme für die Dokumentmenge. Der Wert ist niedrig, wenn  $t$  entweder selten in einem Dokument vorkommt oder wenn  $t$  in einer großen Anzahl von Dokumenten enthalten ist. Je häufiger ein Term in vielen Dokumenten enthalten ist, desto weniger ausdrucksstark ist er. TF\*IDF ist daher am niedrigsten, wenn der Term  $t$  in allen Dokumenten der Dokumentmenge enthalten ist. In diesem Fall hat er keine besondere diskriminierende Kraft mehr. Dies tritt vor allem bei Wörtern wie „und“, „ist“ oder Artikeln auf. Diese Wörter können z. B. über eine Stoppwortliste aus dem Verfahren herausgefiltert werden.

Werden nun die TF\*IDF-Werte für alle Terme  $t$  in einer Anfrage  $q$  aufsummiert, kann für jedes Dokument  $d$  der Dokumentmenge ein eigener Relevanzwert berechnet werden. Eine Sortierung der Dokumentmenge anhand dieses Wertes (auch Score genannt) ist eine direkte Anwendung eines Relevanzrankings:

$$\text{score}_{\text{TF*IDF}}(q,d) = \sum_{t \in q} \text{tf}(t,d) \cdot \text{idf}(t). \quad (4.7)$$

Viele Varianten dieses ursprünglichen Modells sind bekannt, so z. B. Abwandlungen, die eine Normalisierung der Dokumentlänge oder eine Skalierung der tf-Werte berücksichtigen. Der eigentliche Ansatz, dass abhängig von der Anfrage jedem Dokument ein Score zugewiesen wird, wird nicht verändert. Ein Beispiel hierfür wäre die Möglichkeit, mittels logarithmischer Skalierung des tf-Wertes Einfluss auf die Gewichtung zu nehmen. Die grundsätzliche Annahme, dass ein  $n$ -faches Auftreten des Terms  $t$  einer  $n$ -fachen Steigerung der Systemrelevanz gleichkommt, wird durch eine Logarithmierung der Termfrequenz  $\text{tf}$  abgeschwächt.

Das erweiterte boolesche Retrieval repräsentiert in diesem Fall ein Dokument als einen Vektor und arbeitet mit Termgewichtungen. Dieser Ansatz bildet den Schnittpunkt zum Vektorraummodell, welches im folgenden Abschnitt vorgestellt wird.

### 4.3.3 Vektorraummodell

Eines der Standardverfahren zur Implementierung eines Relevanzrankings in IR-Systemen ist das Vektorraummodell, das gleich mehrere Schwächen des booleschen Modells auszugleichen versucht: zum einen wird ein Ranking eingeführt, zum anderen kann durch die geometrische Interpretation der Dokumente die Strenge der booleschen Logik entschärft werden; auch Dokumente, in denen nicht alle Anfragerterme vorkommen, können in der

gerankten Ergebnisliste vorkommen und eine hohe Platzierung erreichen, da sie trotzdem relevant sein können (Robertson, 1977b).

Im Gegensatz zum einfachen booleschen Modell werden dabei sowohl die Dokumente selbst, als auch die Anfragen als Vektoren verstanden. Der Vektor  $\vec{V}(q)$  ist der Anfragevektor und der Vektor  $\vec{V}(d)$  ist der Dokumentvektor. Innerhalb dieser Vektoren entspricht jeder Term einer Dimension. Die unterschiedlichen Terme eines Dokuments oder einer Anfrage spannen so einen mehrdimensionalen Raum auf. Da es sich beim Vektorraummodell um einen klassischen Bag-of-Words-Ansatz handelt, wird die Reihenfolge der Terme beim Aufspannen der Vektoren ignoriert. Gesucht werden nun die Dokumentenvektoren, die dem Anfragevektor ähnlich bzw. nahestehend sind. Dies kann z. B. über den Kosinus-Abstand bestimmt werden, der sich aus dem Skalarprodukt geteilt durch das Produkt der Länge der beiden Vektore ergibt:

$$\text{score}_{\text{cosine}}(q,d) = \frac{\vec{V}(q) \cdot \vec{V}(d)}{|\vec{V}(q)| |\vec{V}(d)|}. \quad (4.8)$$

In praktischen Umsetzungen werden die beiden Vektoren mit gewichteten Werten gefüllt, wobei  $\text{score} \in \mathbb{R}^{|V|}$  gilt. Das TF\*IDF-Gewichtungsmaß wird hierbei am häufigsten angewendet (Salton u. McGill, 1983; Manning u. a., 2008).

#### 4.3.4 Probabilistisches Modell

Neben der TF\*IDF-Rankingfunktion ist Okapi BM25 oder die aktuellere Implementation BM25F (für strukturierte Dokumente) bei textuell-orientierten Retrievalkampagnen wie TREC sehr erfolgreich. Das System ist nach dem Okapi-System der City University London benannt und stellt die 25. Iteration einer Reihe von Best-Match-Systemen dar (Robertson u. a., 1995). Obwohl BM25 auf dem sogenannten probabilistischen Modell Robertson (1977a); Jones u. a. (2000) basiert und so folglich eine andere theoretische Grundlage als das Vektorraummodell besitzt, kann es doch auch wie TF\*IDF zu den klassischen Bag-of-Words-Ansätzen gezählt werden. Auch die Berechnung des BM25-Scores ist abhängig von den Anfragetermen und rankt die Ergebnisse einer Treffermenge entsprechend der ermittelten Relevanz zwischen Anfragetermen und der Termverteilung in der Dokumentenbasis. Für eine Anfrage  $q$ , welche die Terme  $t$  enthält, wird der BM25-Score wie folgt berechnet:

$$\text{score}_{\text{BM25}}(q,d) = \sum_{t \in q} \text{idf}(t) \cdot \frac{(k+1) \cdot \text{tf}(t,d)}{\text{tf}(t,d) + k \left( (1-b) + b \frac{dl}{\text{avdl}} \right)}, \quad (4.9)$$

wobei  $\text{idf}$  die inverse Dokumentfrequenz (Gleichung 4.5),  $\text{tf}$  die Termfrequenz sowie  $dl$  und  $\text{avdl}$  die aktuelle Dokumentlänge bzw. die durchschnittliche Dokumentlänge in der gesamten Dokumentmenge sind. Die beiden Parameter  $0 \leq k$  und  $0 \leq b \leq 1$  werden zur Normalisierung genutzt und meist auf feste Werte gesetzt (Robertson u. Zaragoza, 2010, S. 360).

Das probabilistische Modell beruht auf bedingten Wahrscheinlichkeiten und macht in der zuvor skizzierten Umsetzung Annahmen über die Unabhängigkeit der verwendeten Terme

sowie deren Einfluss auf die eigentliche Relevanz eines Dokuments. Die grundsätzliche Idee hinter BM25 und dem probabilistischen Modell ist dabei aber auch die Vagheit der Sprache, die bei der Anfrageformulierung besteht und in das Modell eingearbeitet wurde. So werden im probabilistischen Modell nur die Wahrscheinlichkeit für die spätere Relevanz berechnet (Robertson, 1977a). Obwohl *Relevance-Feedback* (die Einbeziehung von Relevanzbewertungen durch den Benutzer) sowohl für das Vektorraummodell als auch das probabilistische Modell als Konzept existiert, fügt es sich doch besonders im letzten besonders nahtlos in die dahinterstehende Theorie ein. Dies ist auch einer der Vorteile, die von Fuhr (2004, S. 211) skizziert werden: Das Modell kann sowohl über die Retrievalmaße als auch entscheidungstheoretisch über die Kosten begründet werden. Trotzdem ist das Verfahren jenseits von Evaluationskampagnen kaum in kommerziellen Systemen oder Websuchmaschinen vorzufinden, wie Lewandowski (2005, S. 86–87) feststellt.

Eine Sonderform der probabilistischen Modelle sind die sogenannten statistischen Sprachmodelle (*Language-Models*). Sie stellen formale Beschreibungen einer Sprache dar, die bei der Erzeugung von Texten den Regeln dieser Sprache entsprechen. Die Idee, Sprachmodelle für das IR einzusetzen, basiert darauf, dass sich zu einem existierenden Text über die darin enthaltenen Wörter ein passendes Modell ableiten lässt. Für den Einsatz in IR-Systemen kann man somit zu jedem Dokument ein entsprechendes Modell generieren und zu einer Anfrage prüfen, welches Dokumentmodell die Anfrage am ehesten erzeugen kann. Eine weitere Beschreibung der Herleitung und der Stärken und Schwächen der Sprachmodelle findet sich bei Zhai (2007).

#### 4.3.5 Vergleich der Modelle

Mit dem Vektorraummodell, dem probabilistischen Modell bzw. den statistischen Sprachmodellen existieren leistungsfähige Alternativen zum einfachen bzw. erweiterten booleschen Retrieval. Diese Modelle haben sich in den Evaluationskampagnen wie TREC oder CLEF beweisen können und stellen in der Forschung den aktuellen Stand der Technik dar.

Betrachtet man allerdings den derzeitigen Stand der Technik bzgl. der Retrieval- und Rankingmodelle in digitalen Bibliotheken, so herrscht nach wie vor das erweiterte boolesche Retrieval vor. So arbeiten auch große Systeme wie Scopus, Web of Science oder auch RePeC mit einem einfachen, term-basierten Relevanceranking oder vollständig ohne ein Ranking, wie z. B. arXiv (s. Abschnitt 2.6). Dies liegt zum einen daran, dass die alternativen Modelle ihre jeweiligen Stärken nicht voll ausspielen können. Die Retrievalleistungen der unterschiedlichen Ansätze sind je nach Anwendungsgebiet vergleichbar, wenn eine jeweilige domänenspezifische Anpassung und ein Tuning der Parameter sowie entsprechende Vor- und Nachverarbeitungsschritte eingeplant wurden. Hier werden teilweise mit viel händischer Arbeit die jeweiligen Modelle um weitere Schritte und Faktoren erweitert, dem sogenannten *Feature Engineering* (Manning u. a., 2008, S. 311), wobei der Begriff im Bereich der Dokumentklassifizierung geprägt wurde und nicht so häufig im IR angewendet wird. Die eigentliche Retrievalleistung der zugrundeliegenden Rankingmodelle tritt dabei zusehends in den Hintergrund. Es kann folglich keines der Modelle per se als das Beste angesehen werden (Metzler, 2011, S. 22).

Dies zeigt sich auch in Studien von Zhu u. a. (2009, S. 707), in denen die besten MAP-Werte eines TF\*IDF-, BM25- bzw. LM-basierten Systems nur minimale Unterschiede zeigten (TF\*IDF = 0,2233, BM25 = 0,2220 und LM = 0,2244). In dieser Arbeit wurden auf Grundlage der *TREC2006 Expert Search Test Collection* die Rankingmodelle mit internen Dokumentstrukturen, PageRank und einer Anfrageerweiterung angereichert, jedoch zeigten sich sowohl bei der Anwendung der unveränderten Rankingmodelle als auch den besten Kombinationen keine signifikanten Unterschiede (TF\*IDF+QE+PR = 0,4081, BM25+QE+PR = 0,4067 und LM+QE+PR = 0,4087).

Ähnliche Ergebnisse zeigten sich in einer Metastudie von Kürsten u. Eibl (2011), in der fünf unterschiedliche Stemmer, dreizehn IR-Modelle und drei Pseudo-Relevance-Feedback-Mechanismen in allen möglichen Kombinationen miteinander verglichen wurden. Es zeigte sich weiterhin, dass gerade im Zusammenspiel der verschiedenen Systemkomponenten die Wahl des IR-Modells nur ein Faktor unter vielen ist. Konkret konnte eine TF\*IDF-Implementation das zweitbeste Ergebnis hinter einer Kombination mit einem probabilistischen und parameterfreien DLH-Modell (eine Variante des *Divergence from Randomness*-Modells) erlangen. Die arithmetisch gemittelten MAP-Werte unterschieden sich mit 0,3578 an erster (DLH) und 0,3518 an zweiter Stelle (TF\*IDF) nur knapp.

Dies deckt sich mit der Einschätzung von Fang u. a. (2011):

„The state-of-the-art retrieval functions, when optimized, usually have similar MAP values even though their function forms are different and their retrieval results for the same query also tend to differ. This suggests that all the functions may have their own (potentially different) weaknesses and strengths.“  
(Fang u. a., 2011, S. 7:3)

## 4.4 Nicht-textuelles Ranking

Alle bisher vorgestellten Retrieval- und Rankingmodelle basieren auf dem zuvor beschriebenen Bag-of-Words-Ansatz: Anfrage- und Dokumentterme werden dabei verglichen und jeweils der Exact- oder Best-Match dem Benutzer in Form einer gerankten Ergebnisliste angeboten. Die Bewertung der jeweiligen Relevanz erfolgt ausschließlich auf Grundlage von Termhäufigkeiten oder anderen ermittelten sprachlichen Eigenschaften. In dieser Arbeit werden diese als textuelle Dokumentfeatures bezeichnet. Wie in Abschnitt 2.4 beschrieben bestehen die Dokumente in digitalen Bibliotheken allerdings nicht nur aus textuellen Features und es lassen sich, wie in Kapitel 6, gezeigt aus der Fülle an Metadaten weitere Features ermitteln, die ein Dokument und seine Relevanz bzgl. einer Nutzeranfrage beschreiben. Neben den textuellen Eigenschaften, die sich direkt aus den Dokumenten auslesen lassen, können zusätzliche Daten aus und über die Dokumente, ihre Autoren oder sonstige Metadaten verwendet werden um ein Ranking umzusetzen (Mayr, 2009a). Für Ferber (2003, S. 302–306) sind hierbei vor allem die Angaben von Interesse, die von „(möglichst fachkundigen) Menschen gemacht werden, also manuelle Indexierungen, Beschreibungen und Beurteilungen oder Verweise [...]“.

Im folgenden Abschnitt sollen genau diese Eigenschaften vorgestellt werden. Im Einzelnen sind dies u. a. Verfahren aus der Websuche wie der PageRank oder auf Zitationsanalyse

beruhende, wissenschaftliche Rankingverfahren wie der h-Index. Eine Mischform aus beiden Ansätzen stellt der Google-Scholar-Rank dar, der zusammen mit anderen aus der Welt der Netzwerkanalyse entlehnten Ansätzen wie z. B. der Autorenzentralität oder der Popularitätsbestimmung in sozialen Netzwerken im Web 2.0 vorgestellt wird.

### 4.4.1 PageRank und HITS

In der Frühzeit des Web waren traditionelle term-basierte Retrievalmethoden auf Basis der zuvor beschriebenen Verfahren wie dem booleschen oder dem Vektorraummodell verbreitet. Angereichert wurden diese zwischenzeitlich durch Ansätze wie dem *Sponsored Search* (s. Abschnitt 4.4.6), bei dem primär die Höhe der Zahlung des Webseitenbetreibers in das Ranking einging und Suchmaschinenbetreibern wie Overture zu kommerziellem Erfolg verhalf. Mit dem PageRank von Page u. a. (1999) wurde nun erstmals ein Verfahren zur Suche im Web entwickelt, das primär auf ein nicht-textuelles Rankingverfahren setzte.

Ferber (2003, S. 302) beschreibt den PageRank als eine rekursive Funktion:

$$\text{PR}(i) = (1 - d) + d \left( \frac{\text{PR}(1)}{\text{out}(1)} + \dots + \frac{\text{PR}(m)}{\text{out}(m)} \right). \quad (4.10)$$

Der PageRank einer Seite  $\text{PR}(i)$  wird aus den PageRank-Werten der Webseiten berechnet, die auf die aktuelle Webseite zeigen. Mittels des Wertes  $\text{out}$  wird die Zahl der ausgehenden Links einer Seite angegeben. Durch die Division mit diesem Wert erfolgt ein Ausgleich zwischen Webseiten mit vielen und solchen mit wenigen ausgehenden Links (*Outlinks*). Der Parameter  $0 \leq d \leq 1$  kann als Dämpfungsfaktor verwendet werden. Durch die Addition aller PageRanks der Seiten, die auf die aktuelle Seite  $i$  verweisen, zuzüglich dem Dämpfungsfaktor und dem Wert 1 kann der PageRank der aktuellen Seite bestimmt werden.

Die eigentliche Idee des PageRanks basiert auf der sogenannten *Random Surfer*-Annahme (Page u. a., 1999, S. 3), womit das Verfahren inhaltlich motiviert werden kann. Bei der praktischen Umsetzung werden dazu die Anzahl der *Backlinks*, die eingehenden Links auf eine Webseite, gezählt und gleichzeitig auf die unterschiedliche Wertigkeit von Links Bezug genommen. Eine Webseite wird dabei als wichtig oder einflussreich“ bezeichnet, wenn wiederum andere wichtige Seiten auf sie zeigen. Es macht beim PageRank-Verfahren einen großen Unterschied, ob die eingehenden Links von wichtigen oder von weniger wichtigen Webseiten stammen. Es zählt nicht nur die schiere Menge an Verlinkungen, sondern ebenso die Wichtigkeit der verlinkenden Webseite (Page u. a., 1999, S. 4).

Der so ermittelte Wert pro Seite ist dabei vollkommen unabhängig von der jeweiligen Suchanfrage. Es existiert so pro Webseite jeweils ein PageRank-Wert, der global gilt und statisch ist. Zwar kommen durch die Dynamik des Webs ständig neue Webseiten in die Graphenstruktur hinzu, doch wird davon ausgegangen, dass die Links zwischen den Seiten sich nicht innerhalb kurzer Zeit ändern, also stabil sind. Es muss folglich nicht zwangsläufig für jede neue Webseite eine Neuberechnung des PageRanks angestoßen werden. Durch die Unabhängigkeit von der Suchanfrage ist auch die konkrete Umsetzung innerhalb der Websuchmaschine so angelegt, dass zunächst mit Hilfe einer erweiterten



booleschen Anfrage eine Dokumentmenge mit potentiell relevanten Dokumenten ermittelt wird, die anschließend mittels des globalen PageRanks gerankt wird. Page und Brin bezeichnen ihr Verfahren dabei so, dass mit der term-basierten Anfrage für eine hohe Precision und mittels des PageRanks für eine besonders hohe Qualität am oberen Ende der Ergebnisliste gesorgt wird. Die Grundannahme eines globalen Rankings von Webseiten, unabhängig von der jeweiligen Thematik der Suche und dem dadurch unterschiedlichen Informationsbedürfnis eines jeweiligen Nutzers, ist allerdings nicht unumstritten.

Kleinberg (1999) stellt ein alternatives Verfahren vor, das er den HITS-Algorithmus nannte (*Hypertext-Induced Topic Selection*). Wie der Name bereits sagt, soll hier die Schwäche des PageRanks eines undifferenzierten globalen Rankings, über alle Themengebiete hinweg, umgangen werden. Hierzu werden zwei unterschiedliche Arten von Webseiten eingeführt, die sogenannten *Hubs* (Mittelpunkte) und *Authorities* (Autoritäten). Authorities sind Webseiten, die gute, einschlägige Inhalte besitzen. Hubs sind andere Webseiten, die auf diese Inhalte verlinken. Hubs und Authorities stehen in einer wechselseitigen Beziehung, da sie sich gegenseitig in ihrer Wichtigkeit (wie im Sinne des PageRanks) bestärken.

Bei HITS werden jeder Webseite zwei Werte zugewiesen, der Hub- und der Authorities-Wert, jeweils berechnet pro Anfrage. Dies steht im klaren Gegensatz zum PageRank, welcher anfrageunabhängig berechnet wird. Gleichzeitig werden die beiden Werte nicht auf der gesamten Dokumenten/Webseitenmenge berechnet, sondern nur auf einer kleinen Untermenge an zuvor als relevant befundenen Webseiten. Hierzu wird ein einfaches textuell-basiertes Retrievalverfahren verwendet (im Papier von Kleinberg ist dies die Websuchmaschine AltaVista), um die sogenannte Grundmenge (*root set*) zu erzeugen. Mit dieser Grundmenge wird nach einem weiteren Vorverarbeitungsschritt, der die Grundmenge noch einmal erweitert, gerechnet. Durch die verwendeten textuellen Verfahren wird davon ausgegangen, dass die so gebildete Grundmenge größtenteils homogen bzgl. der Thematik der Anfrage ist. Diese Annahme ist vor einer ähnlichen Problematik im Bereich der Query Expansion (s. Abschnitt 3.3) nicht unproblematisch, kann es doch anstelle eines Query-Drift zu einem Topic-Drift kommen. Durch die Erweiterung können nicht-relevante Seiten hohe Authority-Werte erlangen, obwohl sie nicht zur ursprünglichen Suchanfrage passen (Gottron, 2010, S. 157–159).

#### 4.4.2 Google Scholar Citation-Rank

Ein sehr prominentes System, das einen komplexeren Rankingalgorithmus unter der Zuhilfenahme nicht-textueller Features implementiert, ist Google Scholar, welches nach eigener Aussage unterschiedliche Features kombiniert „[...] in a [...] way researchers do, weighing the full text of each document, where it was published, who it was written by, as well as how often and how recently it has been cited in other scholarly literature“ (Google Scholar, 2011). Google Scholar setzt auf ein kombiniertes Ranking, das sowohl Termfrequenzen, den Publikationsort, die Autorschaft und die Zitationen des Textes beinhaltet. Wie üblich bei einem kommerziellen Produkt, ist die genaue Umsetzung ein Geschäftsgeheimnis, doch zeigen Untersuchungen des Retrievalverhaltens von Beel u. Gipp (2009), dass die Zitationsrate und der Abgleich von Anfrage- und Titeltermen

den höchsten Einfluss auf das Ranking haben, wobei andere Kriterien wie bspw. das Alter der Dokumente keinen großen Einfluss haben. Dies führt dazu, dass die höchstgerankten Treffer einer Suche in Google Scholar meist Dokumente mit einer hohen Anzahl an Zitationen sind. Anders als Page und Brin, die in ihrem ursprünglichem Artikel zum PageRank-Verfahren gegen die Auszählung von Zitationszahlen (bzw. den reinen Backlinks) argumentierten, setzt Google Scholar sehr stark auf den Einfluss von Zitationen. Google Scholar erlaubt es seinen Benutzern nicht selbst, das Ranking zu beeinflussen oder alternative Sortiermöglichkeiten auszuwählen.

#### 4.4.3 Wissenschaftliches Ranking: Journal Impact Factor und h-Index

Der *Journal Impact Factor* (JIF) ist ein Maß, das am Institute for Scientific Information (ISI) um Eugene Garfield entwickelt wurde. Der JIF wird jeweils für ein aktuelles Jahr berechnet, indem für die beiden vorgegangenen Jahre die Anzahl der tatsächlich zitierten Artikel aus einer Zeitschrift durch die Anzahl der potentiell zitierbaren Artikel geteilt wird:

$$\text{JIF}(J)_{2012} = \frac{\text{Zitationen}(J)_{2010} + \text{Zitationen}(J)_{2011}}{\text{Zitierbare Artikel}(J)_{2010} + \text{Zitierbare Artikel}(J)_{2011}}. \quad (4.11)$$

Sind für Artikel einer Zeitschrift  $J$  in den Jahren 2010 und 2011 insgesamt 200 Zitationen registriert worden und die Zeitschrift hat in diesem Zeitraum insgesamt 100 Artikel herausgebracht, dann ist der JIF für das Jahr 2012:  $200/100 = 2,0$ .

Die Aussagekraft des JIF ist, wie viele andere Indizes auch, umstritten. Speziell bei dem vom ISI herausgegebenen Journal Citation Report (JCR), wird immer wieder die problematische fachliche Abdeckung als Kritikpunkt angeführt. Es ist offensichtlich, dass für die exakte Bestimmung der Zitationen eine möglichst große Abdeckung der verwendeten Fachdatenbank von Nöten ist. Gleichzeitig besteht häufig Unklarheit über die Definition eines „zitierbaren Artikels“, da je nach Betrachtungsweise Leitartikel, Editorials, Konferenzbände etc. mit in die Zählung einfließen, oder nicht. Ein weiteres Spannungsfeld stellen Selbstzitationen dar, wie z. B. Glänzel u. a. (2008) für die Domäne der Bioinformatik herausstellt. Auch die zeitliche Abdeckung von zwei Jahren wird als problematisch erachtet, da davon ausgegangen werden kann, dass manche Papiere länger benötigen, um von der wissenschaftlichen Community aufgegriffen zu werden. Während der JIF primär für Zeitschriften entwickelt wurde, gibt es zudem Variationen für Webseiten oder Domänennamen im Internet (*Web Impact Factor*).

Ein anderer Indikator, der sich trotz seines noch jungen Alters als sehr einschlägig erwiesen hat, ist der h-Index von Hirsch (2005). Der h-Index wird zur Bewertung einzelner wissenschaftlicher Autoren, Zeitschriften oder Institutionen herangezogen. Er wird dabei sehr einfach als die Anzahl von  $h$  Papieren, die mindestens  $h$ -mal zitiert wurden, berechnet. Ein h-Index von 6 bedeutet, dass ein Autoren mindestens 6 Papiere geschrieben hat, die mindestens 6-mal zitiert wurden. Redner (2010) zeigt, dass der h-Index in enger Verbindung zur Anzahl der Zitationen steht, sodass  $\sqrt{c}$  wie  $h$  skaliert, wobei  $c$  die Anzahl der Zitationen eines Autors darstellt.

Der h-Index hat sich als sehr robuster Faktor herausgestellt, da Ausreißer wie Pu-

blikationen mit einer unverhältnismäßig hohen Zitationsrate, keinen Einfluss auf den Wert des h-Index haben. Auch eine große Anzahl an Publikationen, die jedoch selten zitiert wurden, hat keinerlei Einfluss. Gleichzeitig kann aber festgestellt werden, dass die h-Indizes zweier Forscher nur bedingt miteinander zu vergleichen sind, da z. B. die Dauer der Publikationstätigkeit eines Autors erheblichen Einfluss auf den h-Index hat. Junge Forscher in ihren ersten Publikationsjahren haben keine Chance auf einen hohen Wert. Des Weiteren ist der h-Index, ebenso wie der JIF fachlich nicht übertragbar, da in den unterschiedlichen wissenschaftlichen Domänen grundverschiedene Publikationstraditionen herrschen.

In Zeiten frei verfügbarer Publikationen im Netz bzw. in digitaler Form über digitale Bibliotheken stellt sich die Frage, ob die reinen Zitationszahlen das Maß aller Dinge sind. Schlögl u. Gorraiz (2012) schlagen beispielsweise vor, Zitationszahlen durch Downloadstatistiken zu ersetzen:

„Es stellt sich nun die Frage, ob es nicht geeignetere Datengrundlagen gibt. Anbieten würden sich hier vor allem Download-Daten. Diese fallen in den Bibliotheken selbst an und bringen dadurch das tatsächliche Nutzungsverhalten von wissenschaftlichen Zeitschriften besser zum Ausdruck. Außerdem sind Download-Daten aktuell, während Zitationsdaten erst mit einer mehr oder weniger großen Verspätung anfallen.“ (Schlögl u. Gorraiz, 2012, S. 87)

Als Abschluss ihrer Untersuchung stellen die beiden Autoren fest, dass die meist-zitierten Zeitschriften auch diejenigen sind, die die höchsten Downloadraten verzeichnen. Einen ähnlichen, aber abgeschwächten, Zusammenhang stellen sie auf der Articlebene und bei den Impactfaktoren fest.

Ist man sich der Nachteile der zuvor vorgestellten Verfahren bewusst, kann man mit Tools wie Publish or Perish<sup>5</sup> von Harzing (2010) auf Grundlage von Daten aus Google-Scholar alle o. g. Werte berechnen.

#### 4.4.4 Zentralität in Autorennetzwerken

Eine andere Art der Netzwerkanalyse ist die Berechnung der sogenannten Autorenzentralität. Die Annahme hinter dieser Berechnung ist, dass sich die Bedeutung eines Dokuments durch den Einfluss seines Autors bestimmen lässt. Dieser Einfluss wird hierbei mit der Zentralität des Autors in einem Netzwerk von Koautoren gleichgesetzt. Die Relevanz eines Dokuments steigt mit der Zentralität seiner Autoren in besagtem Netzwerk. Nutzt man die errechnete Zentralität der Autoren als Sortierkriterium, erhält man ein alternatives Relevanzranking für eine gegebene Dokumentmenge.

Der Hintergrund dieser Überlegung ist, dass in der Wissenschaft die Generierung von neuem Wissen in ein soziales Netzwerk von Wissenschaftlern eingebunden ist. Wissenschaftler arbeiten nicht isoliert von der sozialen Außenwelt, sondern im Austausch mit Kollegen. Die Ergebnisse dieser wissenschaftlichen Arbeit werden auch gemeinsam publiziert. Es bildet sich so ein großes, meist internationales Netzwerk von Koautoren (He, 2009; Freeman, 1977).

---

<sup>5</sup><http://www.harzing.com/pop.htm>

Die Zusammenarbeit in der Wissenschaft ist in ihrer Art und Weise nach verbunden mit dem Konzept der *Betweenness-Zentralität*, wie sie bei der Analyse von sozialen Netzwerken verwendet wird. Die *Betweenness* beschreibt den Grad zu welchem ein Knoten (in diesem Fall ein Wissenschaftler) auf einem kürzesten Pfad zwischen anderen Knoten liegt. Anders ausgedrückt kann dies als der Grad interpretiert werden, zu dem ein Knoten eine vermittelnde Rolle zwischen anderen Knoten einnimmt (Mutschke, 2009). Die Autorenschaft mit anderen Autoren sei dabei „a process in which knowledge flows among scientists“ (Yin u. a., 2006) und ein Autor mit hoher *Betweenness* nähme einen entsprechend stark frequentierten Punkt innerhalb dieses Netzwerks ein, der gleichzeitig unterschiedliche Gruppen im Netzwerk miteinander verbindet.

Neben empirischen Studien, die die Bedeutung des *Betweenness*-Maßen unterstreichen wie Yan u. Ding (2009), wurden die starken Zusammenhänge zwischen Zitationszahlen und *Betweenness* als Indikator für die weiter Verwendung dieses Maßes für das IR genutzt. Mutschke u. a. (2011) machten sich diese Bedeutung zunutze, indem sie die *Betweenness* als ein Rankingkriterium für die wissenschaftliche Informationssuche verwendeten. Sie beschreiben das Verfahren als ein mehrschrittiges Verfahren, das ausgehend von einer initialen Dokumentmenge, die auf Grundlage von Indextermen zusammengestellt wurde, ein Reranking auf *Betweenness* der Autoren anbietet (Mutschke u. a., 2011, S. 358).

#### 4.4.5 Web 2.0-Retrieval und Social-Ranking

Eine andere Situation kann in sogenannten Web 2.0, Social-Bookmarking oder anderen dem Social-Web zuzuordnenden Plattformen beobachtet werden. Hier haben sich in den letzten Jahren auch speziell für den wissenschaftlichen Bereich entsprechende Angebote platzieren können. Beispielhaft für den Bereich der digitalen Bibliotheken seien hier die beiden Plattformen Bibsonomy (Benz u. a., 2010) und CiteULike als von der Wissenschaftsgemeinde getragene Plattformen genannt. Der Inhalt dieser beiden Systeme wird primär von seinen Benutzern erzeugt und gepflegt (*User-generated Content*) und ist daher in Bezug auf die Erschließungsqualität mit traditionellen digitalen Bibliotheken nicht zu vergleichen, da hier ein Großteil der Daten von Laien und daher sehr heterogen erfasst wird. Es kommen keine einheitlichen *Knowledge Organization Systems* (KOS) wie Thesauri oder feste Klassifikationen zum Einsatz, sondern *Social Recommendations* und bevorzugt *Tagging*. Diese Art der Knowledge-Organization wird als *Folksonomie* bezeichnet (Peters, 2010).

Soziale Netzwerke und Web 2.0-Plattformen bieten eine Fülle an neuartigen, nicht-textuellen Features, die für das Retrieval genutzt werden können. Neben den offensichtlichen sozialen Aspekten wie Empfehlungen oder Popularitätswerten sind dies implizites und explizites Nutzer-Feedback sowie das Nutzungsverhalten. Wie Hotho u. a. (2006) zeigen, sind für solche auf Folksonomien basierende Systeme „ordinary ranking schemes such as  $TF*IDF$  [...]“ nicht ausreichend. Dies liege u. a. daran, dass die Textteile, die genutzt werden, um Webseiten, Bilder oder wissenschaftliche Dokumente zu beschreiben, nur sehr kurz sind und ein Volltext nicht verfügbar ist. Um dieses strukturelle Problem zu lösen, schlagen die Autoren zwei alternative Rankingverfahren vor, die sie *Adapted PageRank* und *FolkRank* nennen. Diese Verfahren sind besonders auf die Eigenarten und Strukturen

von Folksonomien ausgerichtet und konnten auf großen Folksonomie-Datensätzen evaluiert werden.

Der Adapted PageRank teilt mit dem klassischen PageRank die prinzipielle Idee eines Graph-basierten Ansatzes, wobei jeder einflussreiche Knoten den Einfluss anderer einflussreicher Knoten unterstützt, auf die er verlinkt bzw. mit denen er gleichzeitig kookkurriert. Ein Nutzer kann in diesem Netzwerk ein beliebiges Dokument oder andere Ressourcen taggen, wobei das Dokument bzw. die Ressource mehr an Einfluss gewinnt, je einflussreicher der Benutzer ist. Die drei unterschiedlichen Entitäten, die in diesem System verwendet werden, sind Dokumente bzw. Ressourcen, Benutzer und Tags. Die zuvor skizzierte gegenseitige Verstärkung des Einflusses kann symmetrisch auf Tags und Benutzer übertragen werden. Während der Adapted PageRank jeweils ein globales Ranking berechnet, kann der FolkRank ein personalisiertes Ranking und daraus Empfehlungen für den Benutzer generieren. Diese Empfehlungen können Dokumente, Tags oder auch andere Benutzerprofile sein, die von potentiellem Interesse für den jeweiligen Benutzer sein können.

Dieser Ansatz wurde von Zanardi u. Capra (2008) aufgegriffen, die Ähnlichkeit zwischen Benutzerprofilen maßen, indem sie Tag-Gewohnheiten von Nutzer verglichen und dieses Ähnlichkeitsmaß nutzten, um Dokumente zu ranken: „based on the inferred semantic distance of the query to the tags associated to such content, weighted by the similarity of the querying user to the users who created those tags“. Die Autoren selbst bezeichneten dies als *Social Ranking*. Einen vergleichbaren Ansatz gibt es von Bao u. a. (2007), die *Social Annotations* nutzen um die Popularität von Webseiten zu berechnen und so einen SocialPageRank implementieren. Die skizzierten Ansätze wurden kurz darauf von großen Websuchmaschinen wie Google oder Bing aufgegriffen, die bekannt gaben, dass sie Nutzungsdaten und Social Annotations von Facebook oder Twitter in den Retrieval- und Rankingprozess aufnehmen (Sullivan, 2010). Die genauen Details dieser Umsetzung bleiben im Verborgenen, aber die Grundannahme ist, dass die von Google sogenannten sozialen Kreise Einfluss auf das Ranking von Webseiten nehmen. Wird eine Webseite von einer verknüpften Person positiv markiert, so wird sie auch in meinem persönlichen Ranking höher bewertet, als bei anderen Personen, deren soziale Kreise diese Webseite nicht oder negativ markiert haben.

Speziell Zipfs Gesetz in Verbindung zu Folksonomien ist Bestandteil der Arbeiten von Peters (2010) und Peters u. Stock (2010). Hier wird aus Zipfs Gesetz schlussfolgernd und gezeigt, dass die in Folksonomien vergebenen Tags einer Zipf-typischen einer invers-logistischen oder auch einer Power-Law-Verteilung folgen. Sie schlussfolgern, dass Tags, die im Long-Tail (also selten für ein Dokument) vorkommen, abgeschnitten werden müssen. Nur Tags mit einem hohen Wiederholungsgrad sollten für weitere Verfahren verwendet werden. Die so verbleibenden Tags seien sogenannte „Power Tags“, die für ein verbessertes Retrieval in Folksonomien verwendet werden sollen. Ausgangspunkt bei Peters ist die Beobachtung, dass in Folksonomien zwar ein großes Potential im Bereich des gemeinschaftlichen Indexierens liegt, allerdings erhebliche Schwächen in der Suche und im Ranking existieren: „Recall seems to be quite good, but all systems lack precision“ (Hotho u. a., 2006). Peters u. Stock (2010, S. 84) schlagen vor, Beobachtungen und Gesetzmäßigkeiten aus der Informatik für das Information Retrieval zu nutzen.

## 4 Ranking und Relevanz

Sie unterscheiden dabei klar, wie mit den verschiedenen informatrischen Verteilungen umzugehen sei und berufen sich dabei auf Newman (2005):

„By neglecting the main difference between the two kinds of distributions, their value for information retrieval is ignored at the same time.“ (Peters u. Stock, 2010, S. 84)

Peters interpretiert die Power-Law-Verteilungskurve so, dass der linke Teil der Kurve als Widerspiegelung der kollektiven Übereinstimmung bei der Indexierung („majorities“), da hier die Begriffe mit der ähnlichen Frequenz liegen („power tags“), wogegen im Tail meist Einzelmeinungen („minorities“) mit niedriger Frequenz vertreten sind („tail tags“). Angewendet wird dieses Verfahren sowohl auf die Indexterme („power index tags“) als auch auf die Suchterme („power search tags“). Leider bleibt diese Arbeit eine Evaluierung des Ansatzes schuldig.

### 4.4.6 Sonstige Rankingverfahren

Neben den prominenten Ansätzen, wie sie in den vorherigen Abschnitten beschrieben wurden, existieren eine Reihe weniger bekannter Verfahren, auf die nur kurz eingegangen werden soll; eine ausführlichere Beschreibung ist den jeweiligen Originalquellen zu entnehmen.

#### Sponsored Search

In den Frühzeiten des WWW war keines der aus der IR-Forschung bekannten Verfahren wie das Vektorraum- oder das probabilistische, sondern ein kommerzielles Modell sehr erfolgreich. Die Suchmaschine Goto.com, die später in Overture umbenannt und letztlich von Yahoo aufgekauft wurde, setzte ein Rankingverfahren ein, das mit keinem der bisherigen Verfahren vergleichbar war. Die Position einer Webseite in der Ergebnisliste wurde (zumindest in den oberen Positionen) maßgeblich von der Höhe der Zahlungen der jeweiligen Webseitenbetreiber bestimmt (Fain u. Pedersen, 2006). Dies wurde unter dem Namen *Sponsored Search* oder *Sponsored Linking* bekannt und war bzw. ist eine beträchtliche Einkommensquelle für Suchmaschinenbetreiber. Zur Profitmaximierung wurde in frühen Implementationen dieses Ansatzes bewusst darauf verzichtet, bezahlte und algorithmisch bestimmte Resultate in der Ergebnisliste voneinander abzusetzen. Es war den Benutzern somit nicht klar, welches Verfahren für die jeweilige Listenpositionierung verantwortlich war.

Das Verfahren war so erfolgreich, dass die meistgesuchten Begriffe versteigert wurden, was zu hohen Summen für eine gute Platzierung in einer Trefferliste zu begehrten Begriffen wie bspw. „Casino“ führte (Raghavan, 2008). Große Suchmaschinenbetreiber wie Google oder Yahoo betreiben immer noch solche kommerziellen Rankings (z. B. Googles AdWords), doch werden erkaufte Treffer inzwischen entsprechend gekennzeichnet, bzw. örtlich von der algorithmisch erstellten Trefferliste getrennt dargestellt. Aus IR-Perspektive interessant wären Retrievaltests, die die Einschlägigkeit und Retrievalgüte von kommerziell und algorithmisch erstellen Trefferlisten vergleichen. Hierzu sind aber keine Studien bekannt.

### Relevance Ranking von Faktendaten

Die Suchmaschine LAILAPS (Lange u. a., 2010) ist eine Mehrdatenbank-Suchmaschine zur Recherche in Datenbanken der Biowissenschaften. Wie die Autoren feststellen, sind in diesen fachspezifischen Datenbanken einfache Verfahren wie TF\*IDF ungeeignet und auch probabilistische Verfahren konnten sich nicht bewähren. Daher werden die textuellen Anfragen der Nutzer um nicht-textuelle Attribute erweitert. Die Autoren nennen konkret neun unterschiedliche Attribute, die u. a. auch sehr domänenspezifische Einträge wie der zugehöriger Organismus oder die Sequenzlänge enthalten. Zusätzlich sind aber auch bekannte term-basierte Ansätze, wie Termhäufigkeit oder Überschneidungen zu Schlagwörtern im Datensatz angegeben. Die Kombination der einzelnen Attribute für die Rankingfunktion wird mit einem neuronalen Netz umgesetzt.

### Verlagsreputation als Rankingansatz

Jordy u. a. (1999) stellen einen interessanten Rankingansatz vor, der darauf basiert, die Reputation wissenschaftlicher Verlage zu messen. Das Verfahren basiert darauf, dass Buchbesprechungen dazu genutzt werden, um die Reputation eines Verlages zu schätzen. In der Studie wurden Bücher, die in den Verlagen de Gruyter, Greenwood, Doubleday, University of Georgia Press und Louisiana State University Press untersucht. Die Autoren merken an, dass, obwohl die meisten Buchbesprechungen positiv formuliert sind, doch klare Variationen in der Art und Weise wie ein Gutachter ein Buch bespricht, erkennbar sind. Die Autoren verwenden die aus der Studie gewonnenen Daten nicht für das Retrieval. Vielmehr vergleichen sie die impliziten Reputationseinschätzungen, die aus den Besprechungen gewonnen wurden, mit denen von Bibliothekaren und ob ein Zusammenhang zwischen Preis der Publikationen und deren Qualität bestand.

## 4.5 Automatisiertes vs. interaktives Retrieval

Im bisherigen Verlauf der Arbeit wurde eine Vielzahl unterschiedlicher Möglichkeiten beschrieben, wie das traditionelle Dokumentenretrieval durch textuelle und nicht-textuelle Dokumentfeatures angereichert werden kann. Allerdings wurde in den bisher vorgestellten Arbeiten meist nur ein spezieller Aspekt oder eine besondere Technik betont (Popularitätsmaße, Zentralität usw.). Dies ist vor allem dem Anstieg der Komplexität geschuldet, wenn mehrere dieser Verfahren miteinander kombiniert werden – sowohl auf algorithmischer als auch auf Seite des Interaktionsdesigns. Komplexere Verfahren sind meist schwerer zu vermitteln und damit zu bedienen, demnach ein klassischer Zielkonflikt.

Der Hintergrund, der im Folgenden umrissenen Debatte, lässt sich mit einem Zitat von Rainer Kuhlen gut einleiten, der in Zeiten ständig wachsender Informationsräume das folgende Informationsdilemma beschreibt :

„Auf der einen Seite verspricht die Informationsgesellschaft eine in der Menschheitsgeschichte bislang nicht gekannte Ausweitung der verfügbaren Informationsräume [...]. Auf der anderen Seite macht die Ausweitung der Informationsräume (konkret in der Ausgestaltung globaler Informationsmärkte) diese so komplex, daß niemand behaupten kann, sie auch nur annähernd zu beherrschen.“ (Kuhlen, 1999, S. 22)

#### 4 Ranking und Relevanz

Dies sei auch vollkommen natürlich und die Komplexität nicht zu umgehen, da diese nicht nur durch die reine Quantität ständig anwache, sondern auch durch die Verknüpfung oder Hypertextifizierung der Inhalte, wie Kuhlen es nennt.

Ein ähnliches Dilemma zeigt sich auch bei Mandl (2006), wenn er feststellt, dass Mehrwertdienste einen schweren Stand im Web-Retrieval haben:

„Zunehmend tauchen weitere Mehrwerte auf wie Personalisierung oder die geo-graphische Einschränkung von Suchergebnissen. Mehrwertkomponenten werden zwar im Web häufig angeboten, haben sich jedoch nicht etabliert. Ganz im Gegenteil, im Internet zeigt sich ein deutlicher Trend zur „Primitivisierung“ der Suchsysteme. Die Benutzer widersetzen sich allem, was über eine Eingabezeile und einer Ergebnisliste hinausgeht und so mancher Mehrwertdienst wird schlicht nicht wahrgenommen und einige entfernen die Anbieter sogar wieder.“ (Mandl, 2006, S. 65)

Mit der Öffnung von ehemals für geschlossene Nutzergruppen konzipierte Fachdatenbanken in Richtung webbasierter digitaler Bibliotheken und der Gewöhnung der Benutzer an Suchmaschinen mit nur einem einzigen Suchschlitz, ist zu vermuten, dass die Bereitschaft der Nutzer zur Nutzung komplexer Mehrwertdienste ähnlich gering einzuschätzen ist, wie dies in erfolgreichen Websuchmaschinen wie z. B. Google der Fall ist.

Verknüpfungen und Mehrwertdienste, wie von Kuhlen und Mandl beschrieben, sollen eigentlich die Informationsgewinnung vereinfachen, diese an einem Punkt zusammenbringen und so dem Benutzer weite Informationsräume erschließbar machen. Allerdings bringen sie in Form weiterer Interaktionskomponenten zusätzliche Komplexität in den Suchprozess hinein. Ein klassisches Problem des Interaktionsdesigns, das auch Norman (2008) erkennt und feststellt: „Simplicity is not the answer“. Es brauche neue Innovationen in Form technischer Informationsassistenten, um diesem Zielkonflikt zu begegnen, denn „Arbeit, in jeder Form, ist immer auf Assistenz angewiesen, personaler, institutioneller oder technischer Art. Also auch Informationsarbeit. [...]“ (Kuhlen, 1999, S. 24). Er spricht dabei von Assistenten oder Agenten – hierzu mehr im folgenden Abschnitt über den historischen Hintergrund der Debatte von Agentensystemen als Suchunterstützung im Information Retrieval und dem Paradigma der direkten Manipulation.

Während diese beiden Ansätze mehr einen historischen als praktischen Ansatz zur Diskussion bieten, werden darüber hinaus die beiden aktuellen Forschungsbereiche des *Learning-to-Rank* (LETOR) und des *Human-Computer-Information-Retrieval* (HCIR) vorgestellt. Während in den zuvor präsentierten Systemen jeweils nur einzelne Verfahren im Vordergrund standen, werden beim LETOR-Ansatz Verfahren des maschinellen Lernens angewendet, um die Komplexität zu beherrschen. Einen vollkommen konträren Ansatz verfolgen Forscher im Gebiet des HCIR, bei dem nutzerzentrierte und softwareergonomische Lösungen im Zentrum des Interesses stehen. Beide Ansätze lassen sich zurückführen auf eine Debatte zwischen Shneiderman und Maes, die im nächsten Abschnitt kurz umrissen werden soll.



### 4.5.1 Historischer Hintergrund: Agentensysteme vs. direkte Manipulation

Wieso werden Mehrwertdienste von Benutzern nicht akzeptiert bzw. ignoriert? Hinweise auf die Antwort zu dieser Frage könnte die Diskussion um die Gegensätzlichkeit von direkter Manipulation und proaktiven Agenten geben (Shneiderman u. Maes, 1997; Schaefer, 2008). Die Diskussion lässt sich zusammenfassen als Problematik, ob und inwiefern einem Benutzer die vollkommene Kontrolle über einen Prozess zugestanden wird, oder ob er gewisse Teile seiner Arbeit (in diesem Falle des Suchprozesses) an selbständig agierende Agenten bzw. Assistenten abgeben möchte. Offen ist die Frage, inwiefern durch den Entwurf von Mehrwertdiensten als Agenten die Nutzbarkeit nachprüfbar verbessert werden kann und dadurch die Akzeptanz dieser Mehrwertdienste steigt.

Der Begriff des Agenten ist in der Informatik mehrfach belegt: Während in der Software-Technik häufig von Multi-Agenten-Systemen die Rede ist, die meist kleine, autonome und proaktiv miteinander kommunizierende Automaten beschreiben (z. B. im Einsatzgebiet des Roboterfußballs), wird im Bereich der Mensch-Maschine-Interaktion gerne auf die Visionen von Kay (1990) oder Maes (1994) zurückgegriffen, die einen Agenten als virtuellen Butler oder persönlichen Assistenten einführen. Maes selbst beschreibt ihre Vision wie folgt:

„Basically, software agents are a new approach to user software, a new way of thinking about software that end-users have to use. In particular, the way in which agents differ from the software that we use today is that a software agent is personalized. A software agent knows the individual users habits, preferences, and interests. Second, a software agent is proactive. It can take initiative because it knows what your interests are. [...] A third difference with current software is that software agents are more long-lived. They keep running, and they can run autonomously while the user goes about and does other things. Finally, software agents are adaptive in that they track the user's interests as they change over time.“ (Shneiderman u. Maes, 1997, S. 49)

Die vier Kernmerkmale eines Software-Agenten sind laut Maes: Personalisierung, Proaktivität, Persistenz (dauerhafter Betrieb) und Adaptivität (ständiges Lernen und Erweitern des Profils).

Wooldridge u. Jennings (1995) gehen bei ihrer Definition von Agenten einen Schritt weiter und trennen die Definitionen von Agenten in einen „schwachen“ und einen „starken“ Teil auf:

„Perhaps the most general way in which the term agent is used is to denote a hardware or (more usually) software-based computer system that enjoys the following properties:

- autonomy: agents operate without the direct intervention of humans or others, and have some kind of control over their actions and internal state [...];
- social ability: agents interact with other agents (and possibly humans) via some kind of agent-communication language [...];

## 4 Ranking und Relevanz

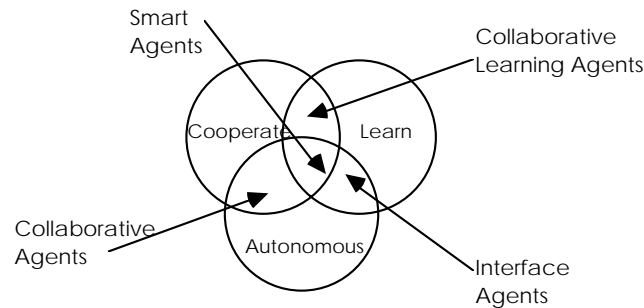


Abbildung 4.2: Klassifizierung von Agentensystemen (entnommen aus Nwana, 1996).

- reactivity: agents perceive their environment, (which may be the physical world, a user via a graphical user interface, a collection of other agents, the INTERNET, or perhaps all of these combined), and respond in a timely fashion to changes that occur in it;
- pro-activeness: agents do not simply act in response to their environment, they are able to exhibit goal-directed behaviour by taking the initiative.“ (Wooldridge u. Jennings, 1995, S. 3)

Diese „schwache“ Definition wird bei der „starken“ vor allem um eine anthropomorphe Sichtweise erweitert:

„For example, it is quite common in AI to characterise an agent using mentalistic notions, such as knowledge, belief, intention, and obligation [...]. Some AI researchers have gone further, and considered emotional agents [...]. (Lest the reader suppose that this is just pointless anthropomorphism, it should be noted that there are good arguments in favour of designing and building agents in terms of human-like mental states [...]) Another way of giving agents human-like attributes is to represent them visually, perhaps by using a cartoon-like graphical icon or an animated face [...] — for obvious reasons, such agents are of particular importance to those interested in human-computer interfaces.“ (Wooldridge u. Jennings, 1995, S. 5)

Für Nwana (1996) ist der Agentenbegriff eine Abstraktion, eine Idee oder ein Konzept, vergleichbar mit den Begriffen der Objektorientierung wie z. B. Methoden, Funktionen oder Objekte. Nwana definiert Agenten dabei nach ihren primären Verhaltensmerkmalen (Kooperation, Autonomie und Lernfähigkeit, s. Abbildung 4.2). Weiter geht Odell mit seiner These, Agenten als „objects with an attitude“ zu bezeichnen (Odell, 2000, S. 45). Ihm zufolge erlaube das Agentenkonzept die Beschreibung einer komplexen Softwareentität, die fähig ist, zu einem gewissen Grad autonom Aufgaben für den Benutzer zu erfüllen. Anders als aber z. B. Objekte, welche durch Methoden und Attribute definiert werden, werden Agenten durch ihr Verhalten definiert.

Die für die Mensch-Computer-Interaktion interessanten Agententypen wurden daher zuerst im Bereich der Interaktionsdesigns gesucht und so erlebten Interfaceagenten

(angelehnt an die „starke“ Definition mit anthropomorphen Visualisierungen) Ende der 1990er Jahre einen Hype und wurden u. a. in so bekannten Implementierungen wie der Büroklammer in der Textverarbeitung Microsoft Word oder dem Hund Rover in Microsoft BOB eingesetzt. Diese Agenten sollten dem Benutzer zur Seite stehen und ihn bei Problemen unterstützen, wurden aber von Kritikern und Kunden nicht akzeptiert. Die Entwicklung an BOB wurde eingestellt und die Interfaceagenten aus Word wieder entfernt. Diese Agenten gehen auf sogenannte Mixed-Initiative-Interfaces (Horvitz, 1999) zurück. Die Probleme lagen zum einen darin, dass Benutzer die anthropomorphen Repräsentationen der Agenten schlicht nicht ernst nahmen und auch große Defizite in Art und Weise der Präsentation und des Timings bestanden. Auch die zentralen Forderungen der Agententheorie nach einem ständig erweiterten Benutzermodell und der autonomen Reaktionsfähigkeit auf die Aktionen des Benutzers waren in den Microsoft-Produkten nicht gegeben.

Die Diskussion zwischen Shneiderman und Maes fußt auch zu einem großen Teil darauf, dass Shneiderman Interfaceagenten unterstellt, nur auf menschenähnlichen Repräsentationen zu basieren und mit fehleranfälliger Sprachsteuerung nutzbar zu sein. So schlägt Maes auch vor, dass:

„It’s as if you had someone else looking over your shoulder as you are using the application, noticing some of your preferences and habits and then offering to automate some of the tasks for you. So you still need a very good direct manipulation interface — visualization — all of these wonderful tools so that the user can personally interact with the application.“ (Shneiderman u. Maes, 1997, S. 53)

Vielmehr ist in der Shneiderman-Maes Diskussion die Frage von Bedeutung: „Wie viel Kontrolle kann und will der Benutzer an den Computer abgeben?“. Abbildung 4.3 zeigt eine Visualisierung der beiden konträren Positionen von Shneiderman und Maes, welche als Kontinuum angelegt ist. Auf der einen Seite steht Shneidermans Position, welche die vollständige Kontrolle des Benutzers in den Mittelpunkt stellt, die andere Seite wird von Maes und ihren autonomen, proaktiven Software-Agenten eingenommen. Nun sind beide Positionen für sich alleine schlecht umsetzbar und eher als theoretische Konstrukte zu sehen, denn kein erstzunehmendes User-Interface kann a) vollkommen auf Interaktion durch den Benutzer noch b) auf unterstützende, intelligente Teilkomponenten verzichten. Im Falle von a wäre das User-Interface per se überflüssig, im Falle von b die darunter liegende Intelligenz des Systems. Es lassen sich natürlich Systeme entwickeln, die nur das eine oder andere Extrem bedienen, allerdings wird ein ausgewogenes System dem Nutzer am besten dadurch helfen, indem es die Vorteile beider Extreme miteinander verbindet.

Für die Frage des Interaktionsdesigns der Agenten/Mehrwertdienste und deren Positionierung innerhalb des Kontinuums kann die Diskussion um Just-In-Time-Information-Retrieval (JITIR) interessante Aspekte einbringen. So sind diese von Rhodes u. Maes (2000) beschriebenen Agentenformen schon so angelegt, dass sie bewusst übersehen werden können (s. die JITIR-Diskussion um „ignorable interfaces“). Diese entkräften den direkten Vorwurf von Shneiderman zu der menschenähnlichen Repräsentation und dem störenden Element. Die Software kann – muss aber nicht – mit Hilfe der Agenten

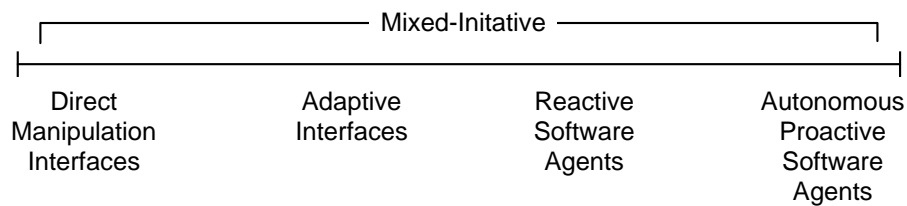


Abbildung 4.3: Einordnung unterschiedlicher Systemansätze dargestellt als Kontinuum zwischen den beiden Polen direkter Manipulation und autonomer Software-Agenten.

bedient werden. Dies ist analog vor den Forderungen der Barrierefreiheit für das WWW (Meurer, 2003) zu sehen. Auch hier herrscht das Paradigma, dass neue Technologien (z. B. AJAX) zwar durchaus eingesetzt werden können und der damit erbrachte Mehrwert auch anerkannt wird, gleichzeitig aber auch immer eine Rückfallmöglichkeit, die ohne diese neuen Technologien auskommt, geboten werden muss.

Historisch gesehen verlagerte sich nach den Rückschlägen für die Interfaceagenten die Diskussion weg von Agenten hin zu einfacher realisierbaren adaptiven Systemen bzw. Interfaces:

„The key idea [...] is that of adaptation to the individual user. Depending on their function and form, systems that adapt to their users have been given labels ranging from adaptive interfaces through user modeling systems to software agents or intelligent agents. Starting in the late 1990s, the broader term personalization became popular, especially in connection with commercially deployed systems. In order to be able to discuss the common issues that all of these systems raise, we will refer to them as user-adaptive systems (“UASs”).“ (Jameson, 2003, S. 305)

Obwohl Software-Agenten und adaptive Systeme nicht das Gleiche sind, werden die Begrifflichkeiten doch in manchen Fällen synonym verwendet. Allen Konzepten ist gemein, dass sie versuchen, individuell auf den Benutzer einzugehen, daher auch der sehr breitgefaste Begriff der Personalisierung, den Jameson beschreibt.

Um ein personalisiertes System zu erstellen, ist es zwingend notwendig, Informationen über den Benutzer, seine Vorlieben, Strategien, Ziele etc. zu sammeln und auszuwerten. Hier kann unterschieden werden zwischen Informationen, die der Benutzer selbst über sich preisgibt (Alter, Vorlieben aber auch „dieses Dokument hat mir gefallen“ etc.) und Informationen, die auf indirekte Weise ermittelt werden (implizite Evaluierung, vgl. Jameson, 2006; Goecks u. Shavlik, 2000). Aus Sicht der Nutzerzufriedenheit ist das zu aggressive Sammeln von Nutzerinformationen sicherlich zu vermeiden. Die gesammelten Informationen sollten zu einem Profil zusammengestellt werden, mit dem das System lernen kann. Dabei ist zu beachten, dass das so gewonnene Nutzermodell natürlich immer nur ein Modell sein kann und nie alle Benutzer bzw. Benutzergruppen gleichermaßen zufrieden stellen kann. Als Schlüssel zum Erfolg bleibt es, den Nutzer zu beobachten und

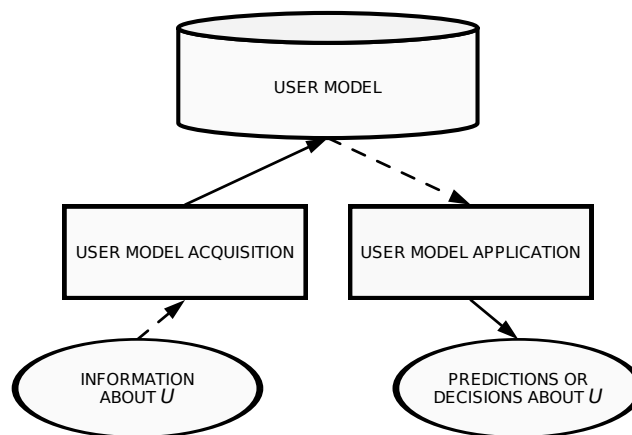


Abbildung 4.4: Generelles Schema zur Verarbeitung in einem Benutzer-Adaptiven-System (entnommen aus Jameson, 2003).

das Verhalten des Systems entsprechend anzupassen, eine Herausforderung vor dem auch der Ansatz des Learning-to-Rank steht.

#### 4.5.2 Learning-to-Rank

Ein großes Probleme in modernen Retrievalsystemen ist die Kombination unterschiedlicher Rankingverfahren und Dokumentenattribute. Dies gilt sowohl für textuelle wie nicht-textuelle Verfahren und Attribute. Moderne Datensätze zur Retrievalevaluation enthalten mehr als hundert verschiedene Features, die kombiniert werden können, um ein möglichst optimales Ranking zu ermöglichen (Trotman, 2005). So beinhaltet der Datensatz Microsoft Learning to Rank Dataset<sup>6</sup> 136 Features, die von Textattributen wie Dokumentlänge, Termfrequenz und BM25-Werten bis zu nicht-textuellen Eigenschaften wie PageRank, SiteRank und vielen weiteren reichen.

Die eigentliche Idee hinter dem LETOR-Ansatz ist es, standardisierte IR-Testkollektionen von Evaluationskampagnen wie TREC oder CLEF zu verwenden und einen überwachten oder semi-überwachten Algorithmus des maschinellen Lernens damit zu trainieren (s. Abb. 4.5). Menschliche Assessoren bewerten Dokumente auf ihre Relevanz mittels explizitem und implizitem Feedback (s. Abschnitt 5.1). Diese Relevanzbeurteilungen können verwendet werden, um einen Algorithmus des maschinellen Lernens zu trainieren. Das erlernte Modell wiederum wird dazu verwendet, eine Ergebnismenge zu ranken und so zur eigentlichen gerankten Ergebnisliste zu gelangen. Die Kollektionen selbst sind hierbei das Trainingsset, das IR-System ist die Evaluationsfunktion und bekannte IR-Performance-Maßzahlen wie Precision oder MAP werden als Gütemaß (*fitness*) verwendet. In diesem Aufbau soll der Algorithmus lernen, wie die Rankingfunktion angepasst werden muss, um die maximale Anpassung an die Relevanzbewertungen der Testkol-

<sup>6</sup><http://research.microsoft.com/en-us/projects/mslr/feature.aspx>

## 4 Ranking und Relevanz

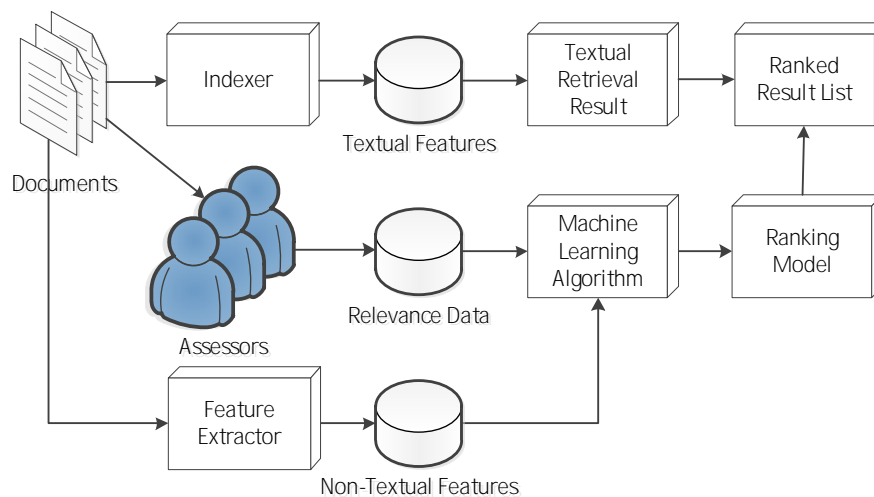


Abbildung 4.5: Ein idealisierter Workflow einer durch maschinelles Lernen angereicherten Suchmaschine unter Zuhilfenahme textueller und nicht-textueller Features.

lektionen zu erreichen. Das gleiche Verfahren lässt sich auch ohne eine standardisierte IR-Testkollektion implementieren, indem ex- oder implizites Nutzerfeedback (z. B. in der Form von Click-Through-Raten) verwendet wird. Dieses Verfahren kann damit auch auf große Websuchmaschinen-Datensätze angewendet werden (Liu, 2007).

Ausschlaggebende Punkte in jedem LETOR-System sind die jeweiligen Dokumentenfeatures, die für das Training des Algorithmus herangezogen werden. Die Features werden dabei auch als Faktoren (*factors*) oder Signale (*signals*) bezeichnet und können grob in drei Unterkategorien eingeteilt werden: statische, dynamische und anfrage-spezifische Features. Statische Features sind unabhängig von der jeweiligen Anfrage und stützen sich nur auf Informationen aus der Dokumentbasis. Beispiele hierfür sind z. B. der PageRank, Dokumentfrequenzen oder sprachliche Eigenschaften. Features, die abhängig von einer Anfrage sind, werden als dynamisch bezeichnet. Üblicherweise sind dies Werte wie TF\*IDF oder BM25, aber auch andere traditionelle Rankingfunktionen. Anfragespezifische Features sind nur von der Anfrage selbst abhängig.

Nahezu alle Rankingmethoden, Anwendungsdomänen und Dokumentenfeatures, die in den vorherigen Abschnitten vorgestellt wurden, können theoretisch in einem LETOR-System kombiniert werden. Tatsächlich konnten unterschiedliche LETOR-Implementationen zeigen, dass sie in der Lage sind, eine sehr hohe Retrievalqualität auch in großen Dokumentmengen mit unterschiedlichen, heterogenen Dokumentenfeatures zu erbringen. Anwendung finden diese Ansätze aber auch in Spezialgebieten wie der Identifikation von Experten (Moreira u. a., 2011) oder der Aggregation verschiedener Rankingverfahren.

Wie aber Chapelle u. a. (2011) beschreiben, sind viele Probleme und Nachteile des LETOR-Ansatzes ungelöst. Der Transfer eines einmal erlernten Modells hin zu einer anderen Dokumentmenge, die ggf. aus einer anderen Domäne stammt, generell nur über

ein zeit- und kostenintensives erneutes Training möglich. Weiterhin bestehen Probleme bzgl. der Diversität der Suchergebnisse.

Es verwundert nicht, dass große Suchmaschinenbetreiber wie Google nicht vollständig auf LETOR-Ansätze zurückgreifen, sondern trotzdem weiterhin auf händisch angepasste Algorithmen und Featurelisten zurückgreifen. In einem Blogpost<sup>7</sup> beschreibt Tom Costello, CEO der Suchmaschine Cuil, dass sie weiterhin an händischen Modellen festhalten, da diese immer noch reinen Modellen und Verfahren des maschinellen Lernens überlegen seien und die verwendeten Maße wie die Click-Through-Rate oder Verweildauer keine hinreichenden Kriterien für ein zufriedenstellendes Training seien. Dies sei darin begründet, dass diese Verfahren nur „learn what people say they like, not what people actually like“ (ebenda).

Bei der Diskussion ob ein rein maschinell-trainiertes Modell oder ein händisches Retrieval-Modell verwendet werden sollte, sind die Meinungen nicht eindeutig. Erschwerend kommt hinzu, dass bei einem maschinell-trainierten System im Nachhinein nicht zu klären ist, welche Faktoren in das jeweilige Ranking eingeflossen sind. Bei mehr als hundert verwendeten Dokumenteneigenschaften ist dies für Menschen nicht mehr überschaubar und erklärbar. Dies gilt dabei sowohl für den Endnutzer als auch für die Systementwickler. Gerade in Hinblick auf feste Trainingssets aus den Evaluationskampagnen müssen sich diese Ansätze aber den Vorwurf gefallen lassen „unfair“ zu arbeiten. Die LETOR-Systeme wurden speziell für diesen einen Anwendungsfall gebaut und trainiert, doch wird sich dies wahrscheinlich nicht übertragen lassen auf andere Anwendungsfälle, in denen z. B. keine sauberen Relevanzurteile vorliegen.

### 4.5.3 Human-Computer Information Retrieval

Eine aufstrebende Disziplin ist das Forschungsfeld des *Human-Computer Information Retrieval* (HCIR), das verschiedene Aspekte des IR und der Mensch-Computer-Interaktion (*Human-Computer Interaction*, HCI) zusammenbringt, um das Design von IR-Systemen zu verbessern (Marchionini, 2006). Ziel des HCIR-Ansatzes ist es Menschen aktiv in den Suchprozess einzubinden. Dabei soll mehr menschliche Intelligenz und letztlich Kontrolle in den Retrievalprozess eingebracht werden. Aus dem Gesichtspunkt des HCIR sind Menschen das eigentliche Zentrum im Designprozess eines IR-Systems. Ihnen sollte zu jedem Zeitpunkt eine Kontrollmöglichkeit über das System gegeben werden, sodass sie dadurch eine letztlich benutzungsfreundlichere Umgebung erhalten. Man beachte hierbei die Nähe zu Shneidermans Ansicht bzgl. des Paradigmas der direkten Manipulation. Traditionell werden unter Kontrollmöglichkeiten suchunterstützende Techniken wie interaktive Anfrageerweiterungen, facetiierte Suche und Navigation, Relevance-Feedback und grafische Darstellungen verstanden. All diese Techniken können für textuelle wie nicht-textuelle Dokumentenbestandteile aus Metadaten, wie kontrollierte Vokabulare oder andere KOS, verwendet werden. Eine einfache facetiierte Suchunterstützung, die Suchen über Schlagwörter, Autoren oder Publikationsjahre einschränken kann, ist in heutigen digitalen Bibliotheken (z. B. Sowiport oder das OPAC-System Blacklight, s.

<sup>7</sup><http://web.archive.org/web/20100822000214/http://www.cuil.com/info/blog/2009/06/26/so-how-is-bing-doing>

Kapitel 2) Standard. Gerade die facetthierte Suche ist eines der am häufigsten mit dem HCIR verknüpften Konzepte.

Tunkelang (2009) stellt in seiner Arbeit diesen verbreiteten Interaktionsmechanismus aus dem Bereich des HCIR vor. Die auf einer Facettenklassifikation basierende Suche (*Faceted Search*), soll den Vorteil einer solchen flexiblen Klassifikation besonders für große Mengen kleiner Dokumente ausspielen können. Dies ist etwa der Fall für die bibliothekarische Ordnung von Zeitschriftenartikeln. Häufig ist der Einsatz inzwischen aber auch vor allem in Web-Shops zu sehen (bspw. Amazon oder eBay). Weiterhin lassen sich mit der Facettenklassifikation sehr komplexe oder innovative Sachgebiete klassifizieren. Hierfür muss die Struktur nicht a priori vorliegen, es lassen sich im Nachhinein Unterklassen und oder Klassen aus Schnittmengen (Facetten) bilden. Der aktuelle Stand der Technik sieht allerdings nur eine Einbindung der facetthierten Suche als Filtermechanismus vor.

Andere bekannte Interaktionselemente, die dem HCIR zugeschrieben werden, sind Tag-Clouds oder Visualisierungen wie Diagramme oder Karten. In Systemen wie VIZGR (Hienert u. a., 2011c) werden diese Interaktionselemente eingesetzt, um Ergebnislisten interaktiv zu filtern. In anderen Kontexten wie dem Portal Sowiport werden Visualisierungen genutzt, um zeitliche und örtliche Verteilungen sowie Schlagwortempfehlungen aus unterschiedlichen KOS einzubinden (Hienert u. a., 2011a). Jedes visuelle Element in diesen Visualisierungen kann genutzt werden um ein Suchergebnis zu filtern oder eine neue Suche anzustoßen.

Zusammenfassend lässt sich sagen, dass die Anreicherung der Suche um alternative, interaktive Suchmechanismen im Zentrum der HCIR-Bestrebungen steht. Zurückgehend auf die Forderungen von Bates (2002) werden dem Benutzer hier alternative Suchstrategien angeboten, um das Nutzungserlebnis und letztlich die Effektivität der Suche zu steigern.

## 4.6 Zusammenfassung

Das traditionelle Dokumentenretrieval (Ad-hoc-Retrieval) hat einen Punkt erreicht an dem keine Verbesserung der verschiedenen Rankingverfahren wie TF\*IDF, BM25 oder den statistischen Sprachmodellen messbar ist. Allgemein sind diese inzwischen auf einem gemeinsamen hohen Niveau angekommenen. Die Ergebnisse der großen IR-Evaluationskampagnen stagnieren seit Jahren und so werden von namhaften Forschern des IR neue und alternative Ansätze gefordert, die eine weitere Verbesserung in dieser traditionellen und nach wie vor prominenten Disziplin des IR ermöglichen (Armstrong u. a., 2009; Buckley, 2009; Trotman u. Keeler, 2011). Verbesserungen werden hierbei explizit nicht nur auf bekannte Evaluationsmaße wie Precision und Recall beschränkt, sondern es sollen neue Kriterien in Betracht gezogen werden.

Eine Vielzahl an Mechanismen und Lösungen aus verschiedenen IR-Domänen wie dem Web-Retrieval, der Bibliometrie, der wissenschaftlichen Evaluationsforschung und des wissenschaftlichen Rankings, dem Machine-Learning und dem Human-Computer-Information-Retrieval wurden in diesen Kapitel zusammengefasst. Letztlich ist im Bereich der digitalen Bibliotheken in vielen Systemen immer noch ein verhältnismäßig alter Stand



der Technik implementiert. Selten gehen dortige Implementierungen über ein erweitertes boolesches Retrieval hinaus. Die kostspielig erzeugten oder eingekauften Metadaten zur Beschreibung der Daten werden nicht gezielt für das Retrieval genutzt. Dies ist von den jeweiligen (textuellen) Retrievalmodellen allerdings auch gar nicht vorgesehen.



## 5 IR-Evaluation

Um die Leistungsfähigkeit von verschiedenen Retrievalsystemen zu evaluieren, hat sich im Bereich des IR eine spezielle Form der Evaluation etabliert, die als das Cranfield-Paradigma bekannt geworden ist. Diese Art der Laborevaluation erlaubt es auf Grundlage von festgelegten Dokumentkorpora und Fragestellungen, kontrollierte Experimente durchzuführen. Gegenüber anderen Formen der Evaluation, wie z. B. beim interaktiven Retrieval, lassen sich Cranfield-Experimente beliebig oft wiederholen. Im folgenden Kapitel wird der allgemeine Hintergrund der IR-Laborevaluation nach dem Cranfield-Paradigma und deren Möglichkeiten zur Datenanalyse erläutert.

### 5.1 Das Cranfield-Paradigma

Das Cranfield-Paradigma, das auf die IR-Evaluationsforschung von Cleverdon (1960) zurückgeht, ist das heutzutage führende Paradigma bei der Evaluation von IR-Systemen, das einen starken Fokus auf empirische Untersuchungsmethoden legt:

„The Cranfield tests established the principle that arguments about the relative merits of different retrieval systems had to be empirically grounded, and, in this respect, they mark an historical change in consciousness from a philosophical and speculative approach to an experimental and empirical one.“ (Ellis, 1992, S. 50)

Essentielle Grundlage der Evaluation sind die Testkollektionen, welche sich immer aus drei Bestandteilen zusammensetzen:

- einem Dokumentkorpus,
- einer Zusammenstellung von Informationsbedürfnissen bzw. Fragestellungen, welche als Topics bezeichnet werden und
- einer Menge an Relevanzurteilen zu den jeweiligen Topics.

Testkollektionen werden traditionell in sogenannten Evaluationskampagnen wie z. B. der *Text Retrieval Conference* (TREC), oder dem *Cross Language Evaluation Forum* (CLEF) entwickelt, gepflegt und eingesetzt. Üblicherweise werden im Rahmen der Kampagnen die Korpora und die Topics von den Organisatoren der Kampagnen zur Verfügung gestellt. Die Teilnehmer der Kampagnen entwickeln dann unabhängig voneinander ihre Retrievallösungen und reichen die Ergebnisse ihrer Systeme bei den Organisatoren ein. Hierzu wird häufig das TREC-Dateiformat als gemeinsamer Standard verwendet, das

---

**Listing 5.1** Auszug aus einer `trec-top-file` Datei, die Retrievalergebnisse enthält.

---

```
76 0 GIRT-DE19900803086 0 1.234 str
76 0 GIRT-DE19990100010 1 0.999 str
76 0 GIRT-DE19980100598 2 0.998 str
...
99 0 GIRT-DE19950100589 9 0.904 str
```

---



---

**Listing 5.2** Auszug aus einer `trec-rel-file` Datei, die Relevanzurteile enthält.

---

```
76 0 GIRT-DE19900803086 1
76 0 GIRT-DE19990100010 0
76 0 GIRT-DE19980100598 1
...
99 0 GIRT-DE19950100589 0
```

---

von `trec_eval`<sup>1</sup> oder anderen Analyseprogrammen gelesen werden kann. Das Format ist sehr einfach aufgebaut (s. Listing 5.1, das ein Beispiel aus den folgenden Experimenten darstellt):

**Topic-ID:** Ein eindeutiger Bezeichner für das Topic. Dieser sollte aufsteigend sortiert sein. Inzwischen werden für die Topics eindeutige DOIs vergeben, so z. B. automatisch im Portal DIRECT, das in der CLEF-Kampagne eingesetzt wird.

**Anfrageiteration:** Bezeichnet die Iteration bei mehrstufigem Retrieval. Wird in den folgenden Ad-hoc-Experimenten ignoriert.

**Dokument-ID:** Bezeichner für das Dokument.

**Rank:** Errechnete Rankingposition des Dokuments. Die Zählung beginnt traditionell bei 0 und die Auflistung sollte aufsteigend sein.

**RSV-Wert:** Der vom System berechnete System-Relevanzwert (s. Abschnitt 4.2). Die Auflistung geschieht aufsteigend.

**Run-ID:** Bezeichner für das verwendete Verfahren.

Die Angabe nach diesem Standardformat erlaubt es, die einzelnen Systeme miteinander zu vergleichen. Hierzu werden die Dokumente aller Kampagnenteilnehmer in einem sogenannten *Pool* zusammengefasst. In diesem Pool sind alle Dokumente, die von den Teilnehmern in den `trec-top-file` Dateien abgeliefert wurden. Menschliche Assessoren (auch *Judge* oder *Rater* genannt) bewerten nun jedes Dokument bzgl. eines Topics hinsichtlich seiner Relevanz. Die Ergebnisse werden in Dateien festgehalten, die dem `trec-rel-file` Format entsprechen (Listing 5.2). Auch hier werden die Topic-ID, die Anfrageiteration (wird auch hier oft ignoriert) sowie die Dokument-ID angegeben. Zusätzlich enthält jede Zeile in der letzten Spalte noch das jeweilige Relevanzurteil des

---

<sup>1</sup>[http://trec.nist.gov/trec\\_eval/](http://trec.nist.gov/trec_eval/)

Assessors. Bei binären Relevanzurteilen sind dies die Werte 0 und 1 oder 0 bis  $n$  bei mehrstufigen Relevanzurteilen. Auf Grundlage der `trec-top` und der `trec-rel` Dateien können nun eine Reihe von Evaluationskennzahlen (s. Abschnitt 5.2) berechnet und das für die Kampagne leistungsfähigste System ermittelt werden.

Die zeitlich versetzte, nachträgliche Kombination von standardisierten Korpora, Topics und deren Relevanzurteilen ermöglicht eine Nachnutzung der Komponenten, die weitere Anwendungen im Rahmen der IR-Evaluation erlauben. Unabhängig von den Evaluationskampagnen kann jederzeit mit den gleichen Daten, aber einem anderen Retrievalsystem ein anderes Experiment durchgeführt werden, das dann auf den Wissensschatz der bereits durchgeführten Relevanzbewertungen zurückgreifen kann. Dabei wird davon ausgegangen, dass ein Großteil der potentiell relevanten Dokumente in den Pools enthalten war, da unterschiedliche Kampagnenteilnehmer mit unterschiedlichen Systemen zu ihrer Erstellung beigetragen haben und so eine gewisse Streuung und Breite des Pools gewährleistet ist. Trotzdem bleibt festzuhalten, dass bei diesem Verfahren meist nicht der gesamte Korpus zu jeder Fragestellung mit Relevanzurteilen versehen ist, sondern immer nur die Dokumente, die im Pool vorhanden waren. Ein vollständig neuartiges Retrievalsystem könnte also potentiell viele nicht bewertete (aber trotzdem relevante) Dokumente finden und trotzdem schlechte Evaluationskennzahlen produzieren, da die nicht bewerteten Dokumente in der normalen Evaluation als nicht relevant eingestuft werden. Dieses Problem wurde zwar von Autoren wie z. B. Büttcher u. a. (2007) aufgegriffen, die versuchten mit Verfahren des maschinellen Lernens diese Nachteile auszugleichen, doch sind die Ergebnisse nicht auf alle Evaluationsumgebungen übertragbar.

Der größte Vorteil der Cranfield-Evaluation ist die Möglichkeit, kontrollierte Experimente durchzuführen, die sich jederzeit wiederholen lassen, da alle Komponenten standardisiert und abrufbar vorliegen. Für die Sammlung von Relevanzurteilen und abgeschlossenen Evaluationen sind verschiedene Plattformen geschaffen worden, so z. B. `evaluatir.org`<sup>2</sup> oder das `DIRECT-System`<sup>3</sup>. Für Evaluationen im interaktiven Retrieval besteht diese Möglichkeit grundsätzlich nicht. Zwar können hier alle Interaktionsschritte aufgezeichnet und somit später nachgeprüft werden, aber es lässt sich nicht eine Variable des Experiments ändern und gleichzeitig die Reproduzierbarkeit des Experiments aufrechterhalten.

Auf dem IR-Forschungskontinuum nach Kelly (s. Abb. 5.1) ist das Cranfield-Paradigma ganz links auf dem Spektrum als „systemorientierte Evaluation“ eingruppiert. Im Gegensatz zur Untersuchung von interaktiven IR und der Untersuchung menschlicher Faktoren, steht das eigentliche Retrievalsystem im Vordergrund. Interaktionsabläufe oder der Kontext einer Informationssuche stehen nicht im Fokus dieses Evaluationsparadigmas. Dies führt zu kritischen Einwänden u. a. von Ingwersen u. Järvelin (2005, 2007), die diese Beschränkungen überwinden wollen.

Der Vorwurf gegenüber dem Cranfield-Paradigma ist, dass die Evaluation innerhalb eines künstlich geschlossenen Systems stattfindet (die *Cranfield Cave*, s. Abb. 5.2) in dem Relevanzurteile ohne den jeweiligen Kontext gefällt werden. Der Kontext in dem sich ein

<sup>2</sup><http://wice.csse.unimelb.edu.au:15000/evalweb/ireval/about/>

<sup>3</sup><http://direct.dei.unipd.it/>

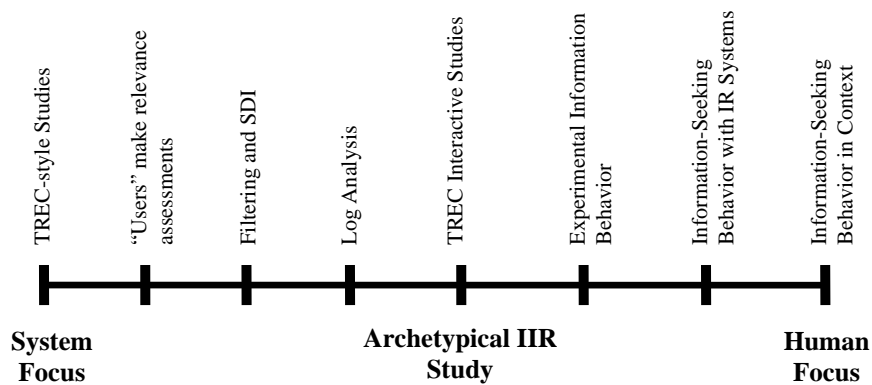


Abbildung 5.1: Das Forschungskontinuum des Information Retrieval mit den drei Foki Systemevaluation, Interaktives Retrieval und der Untersuchung von menschlichen Faktoren (entnommen aus Kelly, 2009, S. 10).

Suchender befindet sei aber für ein optimales Suchergebnis essentiell. Für diesen Ansatz, der den Anspruch erhebt, eine „holistische“ Sicht auf den Retrievalprozess zu werfen, gibt es bis heute allerdings kein so umfassendes Evaluationsframework, wie das des Cranfield-Paradigmas. Zwar existieren vereinzelt Insellösungen oder Testkollektionen für interaktive Retrievalevaluationen (Kelly, 2009, S. 197 ff.), doch gibt es keine vergleichbare Verbreitung, wie sie die klassischen IR-Testkollektionen im TREC-Stil vorweisen können.

Nicht für alle Anwendungsbereiche des IR wird das Cranfield-Paradigma als geeignete Evaluationsumgebung angesehen, so ist z. B. beim Web-Retrieval die Größe und die Unbeständigkeit der Datenbestände ein Problem. Obwohl auch bei der Evaluation von Web-Systemen mit Assessoren und Relevanzurteilen gearbeitet wird, können die so gewonnen Daten nicht zwangsläufig für andere Experimente wiederverwendet werden. Um die Evaluierung so einfach und praxisnah wie möglich zu gestalten, gibt es hier besondere Evaluationsumgebungen, wie das RAT-Toolkit (Lewandowski, 2011), das speziell für die Evaluation von Websuchmaschinen gebaut wurde.

Um die kost- und zeitintensive Evaluation mit Assessoren zu vereinfachen, ist in den letzten Jahren der Versuch unternommen worden, mit Hilfe von Crowd-Sourcing-Systemen wie z. B. dem Mechanical Turk von Amazon durchzuführen (Alonso u. a., 2010). Hierbei entstehen aber neue Probleme wie bei der Rekrutierung von geeigneten Assessoren: bspw. was für Aufgaben ihnen übertragen werden können, wie man ihre Beiträge zusammenführt und wie man Missbrauch verhindert (Doan u. a., 2011).

Ein anderer Trend ist die Evaluation mit Hilfe von Web-generierten Daten, z. B. die Klicks auf bestimmte Seitenelemente und Links, oder Seitenbesuche und deren Dauer zu zählen. Um die Qualität der Experimente ist es nach Crook u. a. (2009) allerdings nicht immer ausreichend bestellt. So weisen die Autoren auf die Gefahren dieser Art der Experimente (z. B. A/B-Tests) hin, da zwar in vielen Wissenschaftsbereichen solche Experimente sehr verbreitet und akzeptiert seien, im Bereich der Web-Science bzw. generell „online-betriebener“ Forschung nicht ausreichend entwickelt sei. Ihr Papier ist

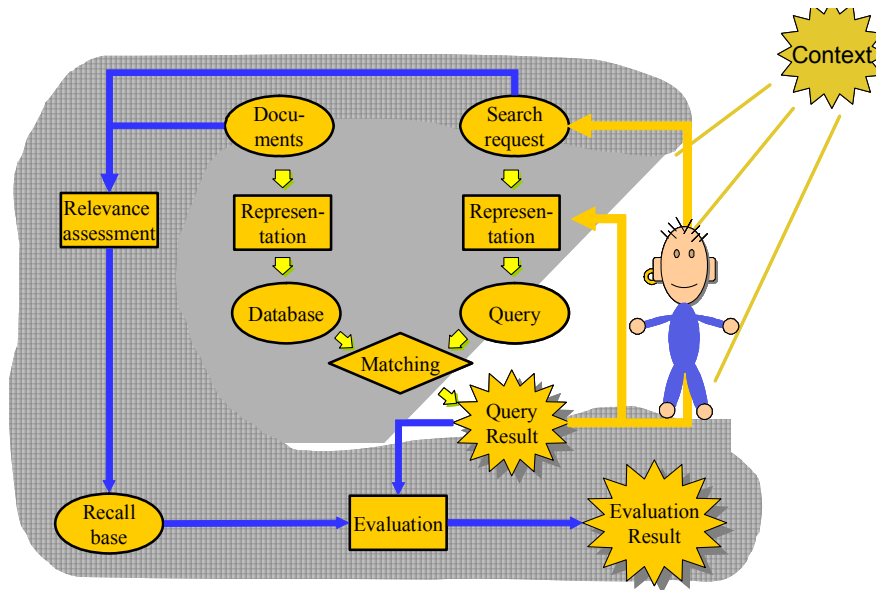


Abbildung 5.2: Die Cranfield Cave (entnommen aus Ingwersen u. Järvelin, 2007, S. 138)

ein starkes Plädoyer für die kontrollierten Experimente der klassischen IR-Evaluation.

Eine sehr ausführliche Diskussion des Cranfield-Paradigmas und der Laborevaluation im IR allgemein sowie die Vor- und Nachteile findet sich bei Sanderson (2010).

## 5.2 Kennzahlen in der IR-Evaluation

Bei der Evaluierung von Retrievalsystemen haben sich mit fortschreitender Entwicklung dieses Wissenschaftsgebiets unterschiedlichste Verfahren und Kennzahlen zur Begutachtung der Qualität eines solchen Retrievalsystems entwickelt.

Es wird bei der Evaluierung unterschieden zwischen ungerankten und gerankten Ergebnislisten. Im Folgenden werden drei Kennzahlen für Evaluierung ungeranker Ergebnislisten vorgestellt (Precision, Recall und das F-Measure) und fünf für gerankte Ergebnislisten (Cut-off-Precision, R-Precision, Average Precision, Binary Preference und Cumulative Gain). Um die jeweiligen Verfahren besser zu verstehen, beziehen sich die angebrachten Beispiele auf die erfundene Beispieldatenbank 5.1, die von einer fiktiven, insgesamt 100 Dokument beinhaltende, Datenbank ausgeht. Von diesen sind für eine ebenfalls fiktive Beispielanfrage 44 Dokumente relevant, wobei von diesen 30 Dokumente vom Retrievalsystem gefunden wurden und 14 nicht. Folglich sind die restlichen 56 Dokumente der Datenbank nicht für die Beispielanfrage relevant, trotzdem hat das Retrievalsystem in diesem Fall 12 davon als relevant markiert und in die Ergebnisliste mit aufgenommen. 44 nicht-relevante Dokumente wurden korrekterweise nicht vom System gefunden.

	Relevant	Nicht relevant
Gefunden	30	12
Nicht gefunden	14	44

Tabelle 5.1: Ergebnis einer Suchanfrage an eine hypothetische Datenbank mit 100 Dokumenten.

### Ungerankte Ergebnismengen

Zwei sehr oft ermittelte Werte, die zur Qualitätsbestimmung eines Retrievalsystems (aber z. B. auch von Klassifikationssystemen) herangezogen werden, sind Precision und Recall (van Rijsbergen, 1974). *Precision* (P) ist das Verhältnis der Anzahl der gefundenen Dokumente zur Anzahl der relevanten gefundenen Dokumenten und gibt die Genauigkeit an, mit der das Retrievalsystem arbeitet:

$$\text{Precision} = \frac{|\{\text{relevante Dokumente}\} \cap |\{\text{gefundene Dokumente}\}|}{|\{\text{gefundene Dokumente}\}|}. \quad (5.1)$$

Angewendet auf die Beispieldaten aus Tabelle 5.1 ergibt sich eine Precision von  $P = 30/(30 + 12) \approx 0,714$ , also ca. 71 %.

*Recall* (R) ist das Verhältnis der Anzahl der relevanten Dokumente zur Anzahl der relevanten, gefundenen Dokumenten. Recall kann auch als Trefferquote interpretiert werden:

$$\text{Recall} = \frac{|\{\text{relevante Dokumente}\} \cap |\{\text{gefundene Dokumente}\}|}{|\{\text{relevante Dokumente}\}|}. \quad (5.2)$$

Auch hier lässt sich der Recall mit Hilfe der Beispieldaten aus Tabelle 5.1 berechnen. Es ergibt sich ein Recall von  $R = 30/(30 + 14) = 0,681$ , also ca. 68 %.

Ein kombiniertes Maß aus Precision und Recall ist das sogenannte *F-Measure*. Hierbei wird das gewichtete, harmonische Mittel auf Precision und Recall berechnet:

$$F = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (5.3)$$

### Gerankte Ergebnislisten

In gerankten Ergebnislisten sind die obigen Maße, die sich jeweils auf die gesamte Ergebnisliste und damit auch über alle Recall-Abstufungen erstrecken, allerdings nur bedingt einsetzbar. Um dem Nutzungsverhalten und dem konkreten Anwendungsbezug für Suchmaschinen und digitale Bibliotheken gerecht zu werden, die jeweils mit gerankten Ergebnislisten arbeiten, wurden weitere Kennzahlen entwickelt, die Auskunft über die Retrievalqualität des Systems geben sollen. Da Nutzerstudien gezeigt haben, dass viele Benutzer eines Retrievalsystems nur die ersten Treffer einer gerankten Ergebnisliste betrachten (Jansen u. a., 2007; Hienert u. a., 2011a), wird zunächst die Kennzahl *Cut-off-Precision*, die auch als *Precision@k*,  $P(k)$  oder  $P@k$  bekannt ist, vorgestellt:



$$P(k) = P@k = \frac{|\{\text{relevante gefundene Dokumente mit dem Rang } \leq k\}|}{k}. \quad (5.4)$$

Hierbei wird nur ein Teil der vom System erzeugten Ergebnisliste zur Precisionberechnung herangezogen, bei Precision@10 (P@10) sind dies z. B. nur die ersten zehn Treffer der Ergebnisliste (Manning u. a., 2008, S. 161). Durch das zugrundeliegende Relevanzranking wird davon ausgegangen, dass die ersten zehn Treffer wahrscheinlich relevanter sind, gemessen am Informationsbedürfnis des Nutzers, sind als ein weiter am Ende der Ergebnisliste positioniertes Dokument. In den verschiedenen Evaluationskampagnen hat sich eine Abstufung der Cut-off-Precision in neun Bereiche etabliert. So sind im Evaluationswerkzeug `trec_eval` die neun Abstufungen P@5, P@10, P@15, P@20, P@30, P@100, P@200, P@500 und P@1000 enthalten. P@30 entspricht dabei bspw. einem typischen Ergebnislistenumfang von drei Seiten, ausgehend von 10 Treffern pro Ergebnisseite. Die Berechnung der Cut-off-Precision hat den Vorteil, dass man keinerlei Abschätzung über die Gesamtzahl der relevanten Treffer benötigt; gleichzeitig lässt sie sich aber schlecht mit anderen Evaluierungsmaßen wie z. B. dem F-Measure verbinden. Die Gesamtzahl der relevanten Dokumenten hat für eine Anfrage einen starken Einfluss auf die Cut-off-Precision.

Um dieses Problem zu umgehen, wurde die sogenannte *R-Precision* entwickelt (Manning u. a., 2008, S. 161). Bei dieser wird die Precision an der R-ten Position in einer gerankten Ergebnisliste, für eine Anfrage, die insgesamt  $R = |\{\text{relevante Dokumente}\}|$  besitzt, ermittelt. Um dieses Maß bestimmen zu können, muss natürlich die Gesamtzahl der relevanten Dokumente in einem Datensatz bekannt sein. Auf den Beispieldatensatz in Tabelle 5.1 angewendet, würde die R-Precision einer P@44 entsprechen, da es 30+14 relevante Dokumente gibt. Ein perfektes Retrievalsystem würde immer eine R-Precision von 1 erbringen, was bedeuten würde, dass alle relevanten Dokumente in der Ergebnisliste enthalten sind und ausschließlich die ersten R-Plätze der Ergebnisliste belegen würde. Die bei der Cut-off-Precision bemängelte schlechte Vergleichbarkeit unterschiedlicher Anfragen ist mit der R-Precision besser umzusetzen.

Ein weiteres zusammenfassendes Maß ist die *Mean-Average-Precision* (MAP), das sich als sehr stabiler und aussagekräftiger Wert für die Gesamtbewertung eines Retrievalsystems herausgestellt hat (Manning u. a., 2008, S. 160). Für die Bestimmung des MAP-Wertes muss zunächst die Average Precision für jede einzelne Anfrage ermittelt werden, wobei  $\text{rel}(k)$  eine Indikatorfunktion ist, die den Wert 1 zurückliefert, wenn das Dokument mit Rang  $k$  relevant ist; ist es nicht relevant liefert sie den Wert 0:

$$\text{AveP} = \frac{\sum_{k=1}^n (P(k) \times \text{rel}(k))}{|\{\text{relevante Dokumente}\}|}. \quad (5.5)$$

Der MAP-Wert für eine Menge an Anfragen  $Q$  ist letztlich der Mittelwert der Average Precision jeder einzelnen Anfrage:

$$\text{MAP} = \frac{\sum_{q=1}^{|Q|} \text{AveP}(q)}{|Q|}. \quad (5.6)$$

Speziell für den Einsatz mit bewerteter Relevanz (*graded relevance*) wurde von Järvelin u. Kekäläinen (2002) *Discounted Cumulative Gain* (DCG) vorgeschlagen. Im Gegensatz zu binären Relevanzbewertungen (s. Abschnitt 8.1.3), die nur zwischen relevant und nicht relevant unterscheiden, gibt es bei der bewerteten Relevanz weitere Abstufungen in der Relevanzbewertung. Angenommen, man verwendet eine dreistufige Relevanzskala, so sind die folgenden Relevanzbewertungen möglich: 2 – hoch relevant, 1 – teilweise relevant und 0 – nicht relevant. DCG liegt nun die Idee zugrunde, dass hoch relevante Dokumente einen höheren Einfluss auf die Berechnung der Kennzahl besitzen sollten, je höher die Rangposition ist. Ein hoch relevantes Dokument auf Rangposition eins sollte die einflussreichste Kombination sein.

Ein immanentes Problem bei der Bewertung mittels der o. g. Kennzahlen sind unvollständig bewertete Dokumentkollektionen. Dies kommt gerade bei großen Kollektionen und bei Evaluationskampagnen häufig vor, da durch die Bewertung eines Pools anstelle der gesamten Kollektion immer nur ein Ausschnitt bewertet werden kann (s. Abschnitt 8.1.3). Buckley u. Voorhees (2004) stellen die sogenannte *Binary Preference* (bpref) vor, um dieses Problem zu umgehen. Hierbei wird ermittelt, wie oft nicht relevante Dokumente vor relevanten Dokumente gerankt werden. Dokumente, die nicht beurteilt wurden, werden bei dieser Berechnung nicht weiter betrachtet. Die Berechnung erfolgt mittels:

$$\text{bpref} = \frac{1}{R} \sum_r 1 - \frac{|n \text{ höher gerankt als } r|}{R}. \quad (5.7)$$

Hierbei ist  $n$  die Anzahl der nicht relevanten Dokumente und  $r$  die Anzahl der als relevant eingestufteten Dokumente.  $R$  ist die Gesamtanzahl der relevanten Dokumente, analog zur Definition der R-Precision. Die bpref-Werte korrelieren meist mit MAP, zeigen aber ein stabileres Verhalten, wenn die Anzahl der bewerteten Dokumente abnimmt.

Neben den hier Vorgestellten gibt es noch eine ganze Reihe weiterer Kennzahlen, da diese aber in dieser Arbeit keine Verwendung finden, wird auf eine Vorstellung verzichtet. Sie sind aber in der Standardliteratur eingehend erklärt (u.a. Manning u. a., 2008; Voorhees u. Harmann, 2005). Die in dieser Dissertation genutzten Kennzahlen wurden ausgewählt, da sie es ermöglichen, jeweils unterschiedliche Aspekte eines IR-Systems zu betrachten. Während z. B. die Cut-off-Precision sehr nah an den tatsächlichen Bedürfnissen eines Benutzers ist (möglichst viele relevante Treffer unter den ersten 10 Ergebnissen), stellt MAP einen statistisch sehr stabilen Wert dar, der sich für den Vergleich über unterschiedliche Topics und Korpora hinweg eignet. Der Einsatz von bpref ist besonders dann, wenn viele neue und unbewertete Dokumente in der Ergebnisliste vorhanden sind, von Vorteil, da sich bei der Berechnung unbewertete Dokumente auf den Wert von bpref nicht negativ auswirken. Die Auswahl der Kennzahlen stellte eine Auswahl von sowohl statistisch sehr robusten Verfahren und nutzerorientierten Werten dar.

Des Weiteren sind die o. g. Verfahren nicht frei von Widerspruch, so wird z. B. MAP von Zobel u. a. (2009) als zu Recall-orientiert kritisiert, was nicht den Erwartungen der Benutzer entspräche, wobei gerade MAP von Buckley u. Voorhees (2000) als besonders stabil und aussagekräftig angesehen wird. Auch die Angabe von P@n wird in manchen

Kontexten als redundant eingestuft. Webber u. a. (2008) vertreten die Ansicht, dass eine zusätzliche Angabe einfacher Kennzahlen wie P@n in Kombination mit komplexeren Kennzahlen wie Discounted-Cumulative-Gain keinen Mehrwert für die Interpretation der Leistungsfähigkeit eines IR-Systems bietet. Letztlich hänge dies aber immer mit dem jeweiligen Evaluationsdesign zusammen und so könne bei einer großen Anzahl an Queries eine höhere Aussagekraft mit P@n erzielt werden als mit anderen Kennzahlen. Für den Einsatz einer großen Anzahl von Queries sprechen sich auch Buckley u. Voorhees (2000) aus. Eine IR-Evaluation mit weniger als 25 Queries wird als nicht aussagekräftig eingestuft, 50 Queries werden als untere Grenze empfohlen, um P@n zu verwenden (Voorhees, 2009).

### 5.3 Vergleich von gerankten Listen

Beim Vergleich der Rankingergebnisse zweier Systeme ist es nicht nur wünschenswert, die Einzelleistungen der Systeme mittels der zuvor vorgestellten Kennzahlen zu quantifizieren, sondern auch, die konkreten Rankings zweier oder mehrerer Systeme miteinander vergleichen zu können. Die Frage, die hier beantwortet werden soll ist, inwiefern die Ergebnisse zweier Systeme miteinander korrelieren. Mathematisch kann diese Frage mittel Kendalls  $\tau$  beantwortet werden (Voorhees, 1998). Kendalls  $\tau$  berechnet die Distanz zweier Rankings, indem die Anzahl an Vertauschungen von benachbarten Paaren gezählt wird, um ein Ranking in ein anderes zu überführen. Diese Anzahl der Vertauschungen wird normalisiert durch die Gesamtzahl von Paarvergleichen, sodass eine perfekte Korrelation den Wert 1 erhält, ein komplett gegensätzliches Ranking einen Wert von -1 und zwei vollständig zufällige Rankings einen Wert von 0.

Der Rankkorrelationskoeffizient Kendalls  $\tau$  ist wie folgt definiert:

$$\tau = \frac{C - D}{n(n-1)/2}, \quad (5.8)$$

wobei  $C$  die Anzahl der Paare ist, die konkordant bzw. übereinstimmend sind und  $D$  die Anzahl der Paare ist, die diskordant bzw. nicht übereinstimmend sind.  $n$  ist die Anzahl der betrachteten Dokumente pro Rankingverfahren. Für folgendes fiktives Beispiel seien die beiden Rankingverfahren  $R_1$  und  $R_2$  definiert, die für die Dokumente  $d_1 \dots d_5$  die folgenden Rankings liefern<sup>4</sup>:

$\{d_1, d_2, d_3, d_4, d_5\} \Rightarrow$		$d_3$	$d_4$	$d_1$	$d_2$	$d_5$	.
	$R_1$	1	2	3	4	5	
	$R_2$	1	2	5	3	4	

Verfahren  $R_1$  rankt die Dokumente in der Reihenfolge  $d_3, d_4, d_1, d_2$  und  $d_5$ , Verfahren  $R_2$  hingegen in der Reihenfolge  $d_3, d_4, d_2, d_5$  und  $d_1$ . Um nun zu überprüfen, inwiefern die Rangordnungen übereinstimmen, müssen die Rankings  $R_1$  und  $R_2$  einzeln verglichen werden. Hierzu wird bei einer maximalen positiven Korrelation von  $R_1$  und  $R_2$  Dokument

<sup>4</sup>Die Notation des Beispiels geht auf <http://www.univie.ac.at/ksa/elearning/cp/quantitative/quantitative-101.html> zurück.

$d_3$  in beiden Rankings den ersten Platz belegten,  $d_4$  den zweiten usw. Dies ist offensichtlich nicht bei allen Dokumenten der Fall. Für jedes Dokument von  $R_2$  wird nun verglichen, ob die auf sie folgenden Rangzahlen größer oder kleiner sind. Ein konkordantes Paar liegt dann vor, wenn eine größere Rangzahl auf eine kleinere folgt, ein diskordantes Paar dann, wenn eine kleinere auf eine größere Rangzahl folgt (im Folgenden fett markiert).

$d_3$ : 1-2 1-5 1-3 1-4  
 $d_4$ : 2-5 2-3 2-4  
 $d_1$ : **5-3 5-4**  
 $d_2$ : 3-4

Kendalls  $\tau$  berechnet sich dabei nun wie folgt ( $C = 8$ ,  $D = 2$ ,  $n = 5$ ):

$$\tau = \frac{8 - 2}{5(5 - 1)/2} = \frac{6}{10} = 0,6. \quad (5.9)$$

Über die Aussagekraft der  $\tau$  Werte gibt es keine eindeutige Interpretation, wie z. B. Sanderson u. Soboroff (2007) feststellen. Voorhees (2000, S. 712) gibt die Untergrenze von  $\tau \geq 0,9$  für die Äquivalenz von zwei Listen an. Liefern zwei Rankings Ergebnisse, die mindestens einen solchen  $\tau$ -Wert ergeben, kann von keinem signifikant unterschiedlichen Ranking gesprochen werden.

In der folgenden Auswertung wird mittels  $\tau$  quantifiziert, inwieweit sich Rankings voneinander unterscheiden bzw. wie kostspielig die Überführung eines Ranking in ein anderes ist.

## 5.4 Zusammenfassung

In diesem Kapitel wurde das sogenannte Cranfield-Paradigma zur kontrollierten Evaluation von IR-Systemen vorgestellt, wie es in internationalen Evaluationskampagnen wie TREC und CLEF verwendet wird. Es basiert auf drei Grundkomponenten: einem standardisierten Korpus von Dokumenten, einer Liste von Topics, die ein Informationsbedürfnis beschreiben und einer Menge von Relevanzurteilen. Zusammen formen diese drei Komponenten eine sogenannte Testkollektion. Innerhalb dieses Versuchsaufbaus lassen sich eine Reihe von Kennzahlen ermitteln, so z. B. Precision und Recall für unsortierte Ergebnismengen als auch die sogenannte Cut-off-Precision, MAP und viele weitere Verfahren für sortierte bzw. gerankte Ergebnislisten.

## 6 Informatrie

In den drei vorangegangenen Kapiteln wurde aufgezeigt, dass heutige digitale Bibliotheken in den beiden wichtigen Schlüsselkomponenten Ranking (Kapitel 4) und Anfrageerweiterung (Kapitel 3) Defizite aufweisen. Die üblicherweise implementierten Lösungen in diesem Bereich sind meist wenig innovativ, gemessen an den Möglichkeiten, welche sich aus der vorhandenen Datenfülle ergeben würde. Speziell in der Gegenüberstellung von textuellen und nicht-textuellen Mehrwertdiensten zeigt sich, dass primär textuelle Dienste angeboten werden. Ein Beispiel hierfür sind Search Term Recommender, die eingesetzt werden, um die sprachliche Vagheit bei der Anfrageformulierung mit Hilfe von Thesauri auszugleichen.

Um aus der Fülle der vorhandenen (Meta-)Daten innovative Mehrwertdienste zu entwickeln, bedarf es allerdings geeigneter Verfahren. Die Informatrie als Wissenschaftsdisziplin der quantitativen Analyse von Informationseinheiten jeglicher Art bietet eine Reihe von Analyseverfahren und -modellen an, um aus den bibliografischen Daten Informationen und Wissen zu extrahieren (s. Abb. 6.2).

„Informetrics is „the study of the quantitative aspects of information in any form, not just records or bibliographies, and in any social group, not just scientists.“ (Tague-Sutcliffe, 1992, S. 1, nach Stock u. Weber, 2006).

Neben der Informatrie als umfassenden Oberbegriff (Nacke, 1979) gibt es weitere Wissenschaftsdisziplinen, die sich mit der Analyse von Informationseinheiten befassen, u. a. sind diese nach Umstätter (2006):

**Bibliometrie:** Die Messung publizierter Dokumente, wie Bücher und Zeitschriften.

**Szientometrie:** Die Messung des Produktionsergebnisses der Wissenschaft.

**Webometrie:** Die Messung des Informationsangebotes im Internet.

Obwohl es intuitiv naheliegt, die vorhandenen Daten in digitalen Bibliotheken mit Hilfe von informatrischen, bibliometrischen oder szientometrischen Methoden zu untersuchen und dadurch Mehrwerte für das Retrieval zu schaffen, muss festgehalten werden:

„Praktische Anwendungen bibliometrischer Methoden [...] waren und sind selten. Am ehesten schien sich noch eine Bradford-Analyse als Hilfsmittel für die Auswahl von Zeitschriften-Abonnements für Spezialinstitute zu eignen [...].“ (Havemann, 2009, S. 58)

Tatsächlich ist die praktische Anwendung bibliometrischer Methoden im Bereich des Information Retrieval bisher selten. In Abschnitt 6.1.6 werden zwar einige Anwendungsgebiete vorgestellt, allerdings sind laut Glänzel (2003, S. 9 f.) tatsächlich nur drei praktische Anwendungsgebiete der Bibliometrie bekannt.

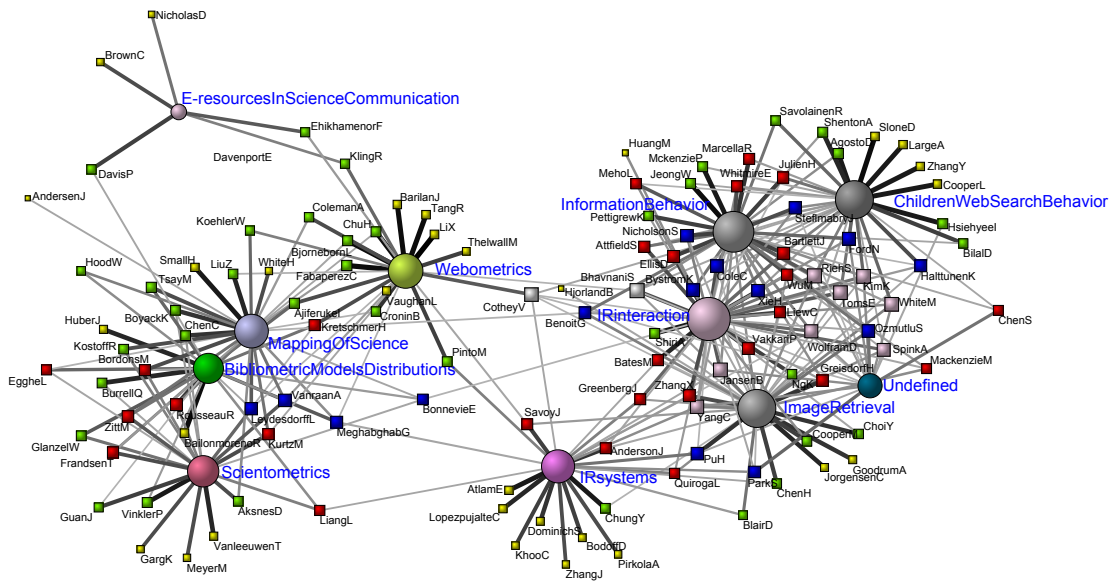


Abbildung 6.1: Eine Science-Map, die sogenannte ABCA Beziehungen (*author bibliographic-coupling analysis*) für die Jahre 2001–2005 zeigt. Es zeigen sich zwei klare Lager: die Szientometrie, bibliometrische Modelle und Webometrie auf der einen Seite und IR-Systeme, IR-Interaktion und weitere Retrievalthemen auf der anderen (Abbildung aus Zhao, 2011).

**Bibliometrie für Bibliometriker:** Dies ist das Gebiet der methodischen und Grundlagenforschung, die keinen praktischen Anwendungszweck kennt, sondern sich mit den zugrundeliegenden Gesetzmäßigkeiten und statistischen, mathematischen oder methodischen Herangehensweisen an die Bibliometrie beschäftigt.

**Bibliometrie für andere wissenschaftliche Disziplinen:** Die Gruppe der Wissenschaftler, die in anderen Disziplinen an bibliometrischen Fragen forscht, ist wohl die größte aber auch am meist durchmischte Gruppe. Bibliometrie wird hierbei natürlich mit dem Fokus auf die eigenen wissenschaftlichen Fragen und Probleme der Disziplin betrieben.

**Bibliometrie für die Wissenschaftspolitik:** Vermehrt wird Bibliometrie zur Wissenschaftsevaluation oder als Bewilligungskriterium für Forschungsgelder oder in Berufungsverfahren verwendet. Hierbei werden meist nationale oder institutionelle Wissenschaftsstrukturen (z. B. der Einfluss deutscher Forscher im Rest Europas) in vergleichender Art betrachtet.

Zwar sind noch einige Spezialfälle wie die Anwendung der Bibliometrie im Bereich der Wissenschaftsgeschichte bekannt (Scharnhorst u. Garfield, 2010), doch stellen diese Anwendungen Ausnahmen oder Ableitungen der drei o. g. Anwendungsfelder dar.

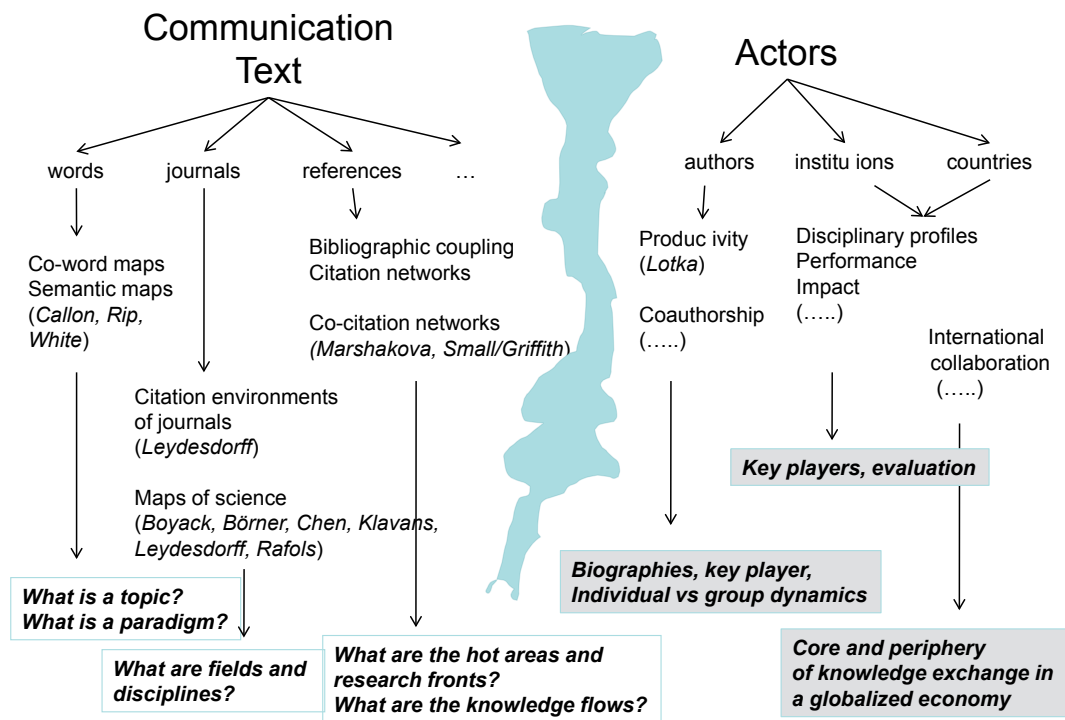


Abbildung 6.2: Gegenstände szientometrischer und bibliometrischer Untersuchungen, angeordnet nach ihren jeweiligen Hauptbeobachtungsobjekten: Texte und Autoren (entnommen aus Scharnhorst u. Garfield, 2010).

Diese Einschränkung verwundert, da doch die Bereitstellung zusätzlicher Funktionalität für das Retrieval auf Grundlage von erzeugten Metadaten von Hjørland (2000) als eine der zentralen Aufgaben von Dokumentaren und Informationswissenschaftlern angesehen wird. Ihre Funktion sei es, nicht nur die Dokumentkollektionen aufzubauen und zu unterhalten, sondern vor allem durch ihre Erschließung Mehrwerte für den Nutzer zu schaffen (Hjørland, 2000, S. 510–512). Die zentrale Aufgabe der Informationswissenschaft sei es, allgemeines Wissen und Prinzipien für den Umgang mit Informationen zu entwickeln und durch neue Technologien nutzbar zu machen. Er nennt explizit auch das Retrieval als einen solchen Mehrwert. So verwundert es, dass es offensichtlich wenige Berührungspunkte zwischen Disziplinen wie Szientometrie, Bibliometrie und Webometrie und Retrievalthemen wie IR-Systeme oder IR-Interaktion gibt, wie Zhao (2011) in einer Zitationsanalyse für die Jahre 2001–2005 zeigen konnte. Zhao arbeitet mit einer *author bibliographic-coupling analysis* (ABCA), deren Ergebnis in Abb. 6.1 dargestellt ist.

Die für diese Arbeit relevanten Untersuchungsgegenstände sind bibliografische Informationseinheiten, also wissenschaftliche Papiere, ihre Autoren, das von ihnen verwendete Vokabular bzw. Fachvokabulare und Journale, Konferenzen und Verlage, in denen diese Informationseinheiten publiziert werden (s. Abb. 6.2 und 6.3). Neben den gerade ge-

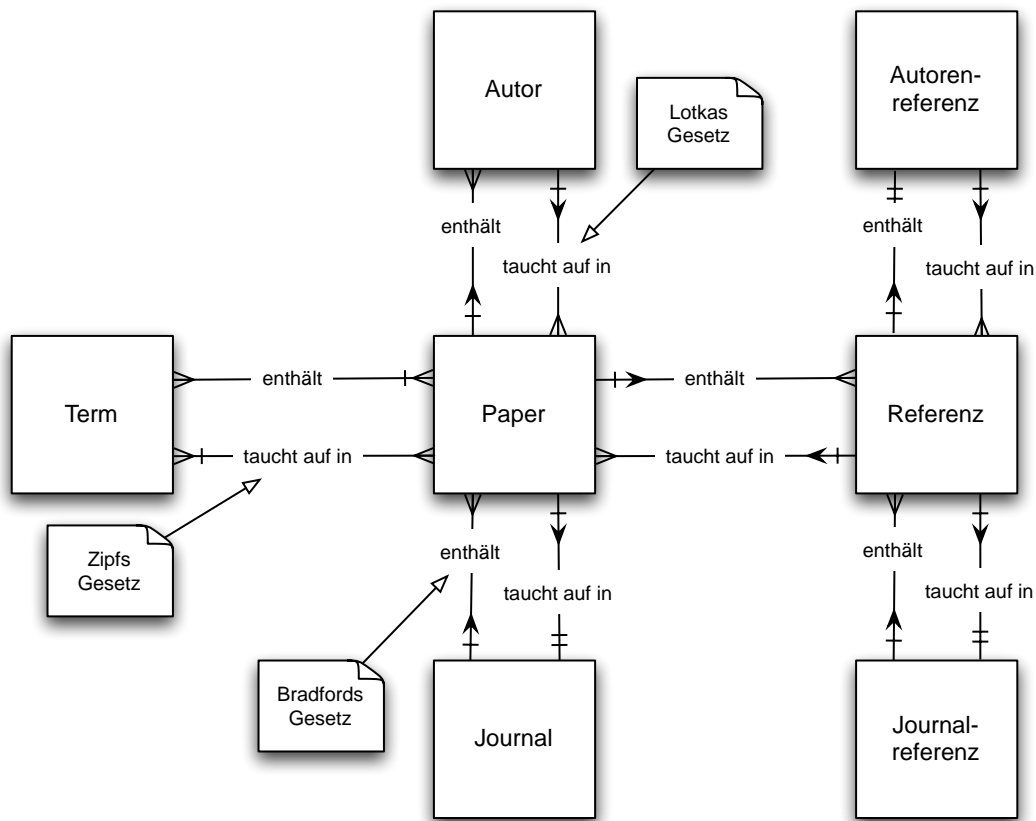


Abbildung 6.3: Grundsätzliche Entitäten bei der Betrachtung von Journalpapieren und ihre Beziehungen zueinander. Markiert sind die informetrischen Gesetze von Zipf, Bradford und Lotka (angelehnt an Morris u. Yen, 2004, übertragen in Krähenfuss-Notation).

nannten gibt es noch die Gruppe der Referenzinformationen, die in dieser Arbeit aber nicht vertieft betrachtet werden. Alle genannten Informationseinheiten oder Entitäten stehen mit der zentralen Informationseinheit, dem wissenschaftlichen Papier, in Verbindung. Morris u. Yen (2004) zeigen diese Verbindungen auf und nennen auch einige der Verbindungen, die gewissen Regelmäßigkeiten oder Gesetzmäßigkeiten folgen, z. B. Lotkas Gesetz zur Autorenschaft von Papieren. Diese Gesetzmäßigkeiten und deren bibliometrische/informetrische Messung werden im folgenden Kapitel vorgestellt.

Dass sich bibliometrische Beobachtungen prinzipiell für das IR nutzen lassen, wurde z. B. am Verfahren des Bradfordizing gezeigt (Mayr, 2009b). Bradfordizing bezeichnet dabei das Reranking einer Ergebnismenge auf Grundlage von Bradfords Gesetz (White, 1981). Artikel, die aus Zeitschriften mit hoher Publikationsfrequenz stammen, werden im Ranking bevorzugt und denen, die aus peripheren Zeitschriften stammen, vorangestellt. Auch neuere Ansätze des selben Autors zielen auf die Kombination von Bibliometrie und IR (White, 2007a,b).



## 6.1 Power-Law-Verteilungen

Lange bevor der Begriff Bibliometrie geprägt wurde, publizierte Lotka (1926) seine Untersuchung zur Produktivität von Autoren. Hierzu maß er die Zahl von Einträgen in zwei Fachbibliographien, einer physikalischen und einer chemischen. Er fand wenige Autoren mit vielen Einträgen und viele, die nur ein- oder zweimal publiziert hatten. Die Verteilung der Publikationen auf Autoren war extrem schief und folgte weitgehend einer Potenzfunktion. Zu einem ganz ähnlichen Ergebnis kam wenige Jahre später der Bibliothekar Samuel Bradford (1934) bei der Untersuchung der Verteilungen von Artikeln auf Zeitschriften in den Bibliographien zweier Spezialgebiete: Wenige Kernzeitschriften enthielten den größten Teil der Literatur, während eine ganze Reihe von Zeitschriften im betrachteten Zeitraum jeweils nur einen Aufsatz zum Thema herausbrachten. Dies ist der wesentliche Inhalt des nach ihm benannten Gesetzes der Streuung von Literatur (*Bradford's Law of Scattering*).

Beobachtungen aus der Bibliometrie, der Nutzung von Sprache oder der Wirtschaft zeigen, dass viele Verteilungs- und Entwicklungsphänomene – seien sie natürlichen oder menschlichen Ursprungs – einem Power-Law (dt. Potenzgesetz) entsprechen (u.a. Newman, 2005; Egghe, 2005). Allgemein lässt sich diese Art der Verteilung mit dem folgenden Ausdruck beschreiben:

$$f(x) = cx^{-\alpha}, \quad (6.1)$$

wobei  $c$  ein konstanter, normalisierender Faktor,  $x$  der Rang der aktuellen Informationseinheit und  $\alpha$  die Stärke des Gefälles ist. Power-Law-Funktionen sind monoton, d. h. wenn  $x$  seinen Wert verändert, wird  $f(x)$  stetig ansteigen oder abfallen. Wenn Power-Law-Funktionen genutzt werden, um Verteilungen zu beschreiben, ist der Exponent  $\alpha$  typischerweise positiv. Dies führt dazu, dass, wenn  $x$  steigt,  $f(x)$  abfällt. Verallgemeinert bedeutet dies, dass gezählte Items oder Ereignisse mit hohem Aufkommen typischerweise selten vorkommen. Man spricht bei den typischen Verteilungen auch von einem sogenannten *Long-Tail* oder der 80:20 Regel. Power-Law-Verteilungen sind unter verschiedenen Namen in der Literatur bekannt. In den folgenden Abschnitten sollen daher die Pareto-Verteilung und die Gesetze nach Lotka, Zipf und Bradford vorgestellt werden.

### 6.1.1 Pareto-Verteilung

Das nach Vifredo Pareto (1848–1923) benannte Gesetz der Pareto-Verteilung lautet frei zusammengefasst: 80 % aller erkennbaren Effekte resultiert aus 20 % aller Ursachen.

Zunächst ist festzuhalten, dass Pareto diese Gesetzmäßigkeit nicht selbst unter diesem Namen verbreitete. Der Begriff „Pareto's Law“ wurde vielmehr von Juran (1954) geprägt. Pareto selbst wird nachgesagt in seiner Arbeit nachgewiesen zu haben, dass 80 % des Eigentums in Italien in der Hand von 20 % der Italiener läge. In zahlreichen Arbeiten von anderen Autoren wurde u. a. beobachtet, dass 20 % der Angestellten eines Unternehmens 80 % der Produktivität erwirtschaften, dass 20 % der Kunden 80 % des Umsatzes erzeugen, dass 20 % eines Bibliotheksbestandes 80 % der Ausleihen erzeugen, dass 80 % der

$x$	$N$
150	400 648
200	234 185
300	121 996
400	74 041
500	54 419
600	42 072
700	34 269
800	29 314
900	25 033
1 000	22 896
2 000	9 880
3 000	6 069
4 000	4 161
5 000	3 081
10 000	1 104

Tabelle 6.1: Paretos Datensätze zur Einkommensverteilung (nach Hardy, 2010). Der Wert  $x$  beschreibt hierbei das Einkommen in £,  $N$  ist die Anzahl der Steuerzahler Großbritanniens im Jahr 1893/1894, die in die entsprechende Einkommensgruppe fallen.

Weltbevölkerung in 20 % der Städte leben usw. Man spricht daher umgangssprachlich auch von der 80:20 Regel.

Pareto selbst betrachtete mehr als 40 Datensätze zur Einkommensverteilung in Europa, Nord- und Südamerika. Er leitete daraus folgenden Zusammenhang ab:

$$\ln N = \ln A - \alpha \ln x. \quad (6.2)$$

Hierbei ist  $N$  die Anzahl der Personen, deren Einkommen größer als  $x$  ist.  $A$  und  $\alpha > 0$  sind Parameter, die geografisch und zeitlich variieren.

Ein Beispiel soll dieses Modell erläutern. Tabelle 6.1 zeigt einen Auszug aus Paretos Daten zu britischen Steuerzahlern aus den Jahren 1893/1894. Der Wert  $x$  gibt dabei die Einkommensgruppe und  $N$  die Anzahl der Steuerzahler an, die mindestens ein entsprechendes Einkommen hatten. So gab es 22 896 von insgesamt 400 648 Steuerzahlern, die mindestens 1 000 £ zu versteuern hatten. Mit Hilfe der Formel 6.2 kann nun berechnet werden, wie viele Personen ein Einkommen z. B. von 2 500 £ zu versteuern hatten:  $\exp(19,331 - 1,3379 \ln 2\,500) \approx 7\,067$ . Die beiden Parameter  $A = 19,331$  und  $\alpha = 1,3379$  wurden dabei aus Hardy (2010) entnommen.

### 6.1.2 Lotkas Gesetz

1926 veröffentlichte Alfred J. Lotka sein später oft zitiertes Papier mit dem Namen „The frequency distribution of scientific productivity“, in dem er die später nach ihm benannte Gesetzmäßigkeit formulierte:

„... the number (of authors) making  $n$  contributions is about  $1/n^2$  of those making one; and the proportion of all contributors, that make a single contribution, is about 60 per cent.“ (Lotka, 1926)

Eine beispielhafte Rechnung, bei der 100 Autoren jeweils einen Artikel verfassen, würde ergeben, dass es 25 Autoren gibt, die je 2 Artikel veröffentlichten ( $100/2^2 = 25$ ). 11 Autoren hätten jeweils 3 Papiere verfasst ( $100/3^2 \approx 11$ ) usw. Lotka selbst spricht hierbei von einer inversen Quadrat-Gesetzmäßigkeit (*inverse square law*). Seine These leitete er aus der Beobachtung zweier bibliografischer Datensätze aus der Chemie und Physik ab (*Chemical Abstracts* und *Geschichtstafeln der Physik*). Die prozentualen Anteile der Autoren, die 1, 2, 3, ...  $n$  Papiere veröffentlicht hatten, plottete er in einen Graphen mit doppelt logarithmierten Skalen. Den Grad der Steigung (*slope*) bestimmte er mit der Methode der kleinsten Quadrate und kam dabei auf eine Steigung von ca. -2. Lotka hatte damit eine Gesetzmäßigkeit über die Produktivität von Autoren im Wissenschaftsbetrieb aufgestellt.

Bis 1941 wurde die Arbeit zunächst überhaupt nicht zitiert und erst ab 1949 unter dem Namen Lotka's Law bekannt (durch Zipf, 1949). Spätestens durch die Arbeiten von de Solla Price (1963) wurde sie allerdings sehr populär. Auch wenn viele Autoren, darunter auch de Solla Price behaupteten, dass Lotkas Beobachtungen für unterschiedlichste Gebiete der Wissenschaft und deren Produktivität gültig seien, kann dies zunächst auf Grundlage zu geringer Größe der Datensätze und aufgrund statistischer Unzulänglichkeiten nicht mit Sicherheit behauptet werden. Wie Potter (1981, S. 22 ff.) zeigt sind, viele frühe Untersuchungen zu Lotkas Gesetz nicht in der Lage, einen statistischen Signifikanztest nach Kolmogoro-Smirnov zu bestehen. Hierbei wird die maximale Abweichung (*deviation*)  $D$  ermittelt:

$$D = \max |F_0(X) - S_n(X)|. \quad (6.3)$$

Potter beschreibt  $F_0(X)$  als die theoretische kumulative Frequenzfunktion und  $S_n(X)$  als die tatsächlich beobachtete kumulative Frequenzfunktion eines Datensamples aus  $n$  Beobachtungen. Signifikanz sei bis zu einem Schwellenwert von 0,01 gegeben, wobei die Kolmogoro-Smirnov-Statistik gleich  $1,63/n^2$  ist. Ist  $D$  größer als die Kolmogoro-Smirnov-Statistik, dann ist die Beispielverteilung nicht gleich der theoretisch angenommenen Verteilung. Lotka selbst betrachtete für die Chemical Abstracts 6981 Autoren, sodass Kolmogoro-Smirnov hier den Wert  $1,63/\sqrt{6981} = 0,0195$  annimmt. Für die Geschichtstafeln der Physik liegt der Wert bei  $1,63/\sqrt{1325} = 0,0448$ . Laut Potter erfüllen aber nur die ermittelten Werte für  $D$  der Geschichtstafeln der Physik die o. g. Bedingung, für die Chemical Abstracts könne Lotka selbst seine Gesetzmäßigkeit nicht einhalten (s. Tab. 6.2). Gleiches gilt auch für viele andere Studien, die bis in die 1980er Jahre durchgeführt wurden und die nur teilweise Lotkas Beobachtungen voll erfüllen konnten (laut Potter

Anz. der Papiere	Beobachtet	$S_n(X)$	Erwartet	$F_0(X)$	$ F_0(X - S_n(X)) $
1	0,5792	0,5792	0,6079	0,6079	0,0287
2	0,1537	0,7329	0,1520	0,7599	0,0270
3	0,0715	0,8044	0,0675	0,8274	0,0230
4	0,0416	0,8460	0,0380	0,8654	0,0194
5	0,0267	0,8727	0,0243	0,8897	0,0170
6	0,0190	0,8917	0,0169	0,9066	0,0149
7	0,0164	0,9081	0,0124	0,9190	0,0109
8	0,0123	0,9204	0,0095	0,9285	0,0081
9	0,0093	0,9297	0,0075	0,9360	0,0063
10	0,0094	0,9391	0,0061	0,9421	0,0030

Tabelle 6.2: Lotkas Daten zu Chemical Abstracts. Da  $D > 0,0195$  folgen die Daten aus Chemical Abstracts nicht Lotkas Gesetz (Tabellenwerte übernommen aus Potter, 1981).

u.a. Murphy, 1973, Schorr, 1975 oder Radhakrishnan u. Kernizan, 1979). Trotzdem ist in Tabelle 6.2 sehr gut zu sehen, dass für kleine Werte (Anzahl der veröffentlichten Papiere pro Autor) eine größere Schwankung zwischen beobachteten und erwarteten Werten zu sehen ist als für große Werte (s. Abschnitt 6.1.8).

Die Gründe hierfür sind vielfältig. Zunächst ist, wie bei allen bibliografischen Modellen, Lotkas Gesetz keine präzise statistische Verteilung, sondern eine Generalisierung, die auf zwei Datensätzen beruht. Weiterhin wurde von anderen Autoren der Fehler begangen, zu kleine Datensätze zu betrachten, die eine schlechte thematische oder zeitliche Abdeckung aufwiesen. Weiterhin wurden teilweise nur einzelne Quellen (u.a nur eine einzige Zeitschrift) betrachtet, was die Datenlage weiterhin verzerrt. Darüber hinaus untersuchte Lotka selbst in seinem Datensatz nur die Autoren mit den Anfangsbuchstaben A und B, weiterhin betrachtete er nur die Erstautoren eines Papiers, die Koautoren wurden von ihm ignoriert (Havemann, 2009, S. 18).

Es zeigt sich allerdings, dass Lotkas Annahmen nicht nur für Zeitschriftenartikel gelten, sondern auch für die Autorenschaft von Monografien Gültigkeit haben (McCallum u. Godwin, 1981). Gerade wenn man bei Monografien den Faktor Zeit mit in Betracht zieht, zeigen sich diese Phänomene. In der Studie von McCallum u. Godwin werden mehr als 2,5 Millionen Bibliothekskatalogeinträge der Universität von Illinois betrachtet. Die fünf wichtigsten Autoren sind danach Shakespeare, Milton, Goethe, Balzac und Dickens – allesamt Autoren, die zum Zeitpunkt der Studiendurchführung nicht mehr lebten und folglich nicht mehr publizierten. Trotzdem werden ihre Bücher durch ihre große Popularität und ihren Einfluss immer wieder neu aufgelegt und führen so zu einer ständig steigenden Anzahl ihrer Publikationen. Man kann sagen, dass die Produktivität eines Autors über seinen Tod hinaus andauert.

Das Problem der Produktivitätsmessung durch reine Publikationszahl ist nicht unumstritten und im Grunde das häufigste Argument gegen den *Science Citation Index* (SCI).

Umstätter (1999) nennt fünf Gründe, die hinter einer hohen Publikationsrate stehen können:

1. Der Autor beschreibt eine Methode, die sich auf viele Probleme anwenden lässt. Papiere solcher Art werden laut Umstätter allerdings auch sehr häufig zitiert.
2. Der Autor beschreibt eine Theorie, die sich auf viele Probleme anwenden lässt.
3. Der Autor beschreibt eine umstrittenen Behauptung, die hohe Attraktion für Zeitschriften und Leser hat.
4. Der Autor besitzt einen hohen Bekanntheitsgrad bzw. eine hohe berufliche Position.
5. Der Autor besitzt eine überdurchschnittliche Publikations- und Einreichungsaktivität. Er reicht seine zahlreichen Publikationen so oft bei unterschiedlichen Zeitschriften und Verlagen ein, bis letztlich das Papier veröffentlicht wird.

Trotz der Schwächen bleibt Lotkas Gesetz ein typisches und höchstprominentes Beispiel für eine sogenannte *Size-Frequency-Verteilung* (Egghe, 2005), welche die Anzahl der Quellen (*Sources*) mit einer bestimmten Nummer von Erzeugnissen (*Items*) in Verbindung setzt. Laut Potter (1981) ist es gleichzeitig ein starker Hinweis auf die Existenz universeller Gruppen von Autoren, die nach ähnlichen Produktivitätskriterien publizieren. Je größer und thematisch, zeitlich und international umfassender eine solche Gruppe wird, umso mehr scheint sie Lotkas Gesetz zu folgen, wohingegen kleinere Forschergruppen (z. B. einer bestimmten Unterdisziplin oder einer kurzen zeitlichen Epoche) mitunter nach anderen Regelmäßigkeiten publizieren. Er hält weiterhin fest, dass, um Lotka-typisches Verhalten nachweisen zu können, ein Zeitraum von mindestens zehn Jahren betrachtet werden sollte. Dies ist für die Gestaltung und Auswertung der Experimente in den späteren Kapiteln von Bedeutung.

Eine weitere Schwierigkeit sind die in Abschnitt 2.4.3 beschriebenen Probleme bei der eindeutigen Identifikation von Autoren in digitalen Bibliotheken oder Bibliografien. Die eindeutige Identifizierung wird z. B. durch nicht einheitliche Erschließung der Namen oder durch mehrdeutige Namen erschwert. Wie Scharnhorst u. Garfield (2010) feststellen, ist dies für die Analyse großer Datenbestände aber kein zwingendes Problem. Ein Beispiel hierfür ist die Arbeit von Fronczak u. a. (2007), die mit Hilfe der INSPEC Datenbank zeigten, dass auch ohne eine eindeutige Identifizierung der Autoren durch Hilfsmittel, wie die Detektion von Kohorten oder Langzeitautoren eine Verbesserung der Datenanalyse möglich ist.

### 6.1.3 Zipfs Gesetz

„The Principle of Least Effort means . . . that a person . . . will strive to solve his problems in such away as to minimize the total work that he must expend in solving both his immediate problems and his probable future problems.“  
(Zipf, 1949)

$r$	$f$	$C = r \cdot f$
10	2 653	26 530
20	1 311	26 220
30	926	27 780
40	717	28 680
50	556	27 800
100	265	26 500
200	133	26 600
300	84	25 200
400	62	24 800
500	50	25 000
1 000	26	26 000
2 000	12	24 000
3 000	8	24 000
4 000	6	24 000
5 000	5	25 000
10 000	2	20 000
20 000	1	20 000
29 899	1	29 899

Tabelle 6.3: Die Häufigkeitsverteilung aller Wörter aus Ulysses von James Joyce als Beispiel für Zipfs Gesetz. Zu sehen sind Rang  $r$ , Häufigkeit  $f$  und das nahezu konstante Produkt aus  $r$  und  $f$  (entnommen aus Glänzel, 2003, S. 42).

Obwohl dieses Zitat auf den ersten Blick zunächst wenig mit Informatric zu tun haben scheint, ist es doch einer der einflussreichsten Grundsätze der Informatric. Die von Zipf aufgestellte Gesetzmäßigkeit der Wortverteilungen in der natürlichen Sprache besagt, dass das Produkt aus Rang  $r$  und Frequenz  $f$  konstant  $C$  sein sollte. Die von Zipf aufgestellte Formel lautet

$$C = r \cdot f. \quad (6.4)$$

Zipf sortierte alle Wörter eines Textes nach Häufigkeit und trug auf der Ordinatenachse die jeweilige Häufigkeit auf. Auf die Abszissenachse trug er die Rangreihenfolge auf und erhielt so eine Hyperbel der Form von Abbildung 6.5. In Tabelle 6.3 ist ein Beispiel für Zipfs Untersuchungen aufzeigt, welches den Rang und die Häufigkeit unterschiedlicher Wörter in der Ulysses von James Joyce auflistet. Das zehnthäufigste Wort kommt danach 2 653 mal vor, das 20 000 häufigste allerdings nur ein einziges Mal. Bei diesem Beispiel ist  $C \sim 24 500$ .

Nach Umstätter (2003) sprach Zipf bei den beiden Achsen auch von Kräften: Die Ordinatenachse nannte er „force of unification“ und die Abszissenachse „force of diversification“. Die erste Kraft führt dazu, dass bestimmte Wörter im Englischen wie „the“, „it“, „is“ usw. oft vorkommen. Umstätter rechnet, dass jedes 10. bis 20. Wort in einem

englischen Text ein „the“ ist. Gleichzeitig werden aber auch viele Wörter benutzt, die im gesamten restlichen Text nicht mehr oder nur sehr selten vorkommen. Prinzipiell ist diese Gesetzmäßigkeit für alle Sprachen gegeben, dabei allerdings in unterschiedlichen Ausprägungen. So können im Deutschen die Äquivalente zum englischen Artikel „the“, „der“, „die“ und „das“ als genus-markierende Artikel aufgrund ihrer anderen Verteilung nur zu einem Drittel so häufig vorkommen.

Das Principle of Least Effort ist allerdings für jeden Text-erzeugenden Menschen das Gleiche. Es ist aus der Sicht der Informationstheorie nichts anderes, als ein ausgewogenes Verhältnis von Information und Redundanz, wie von Shannon u. Weaver (1963) beschrieben. Wendet man ein künstlich verkleinertes oder kontrolliertes Vokabular an, wie einen Thesaurus oder eine Klassifikation, kann dies starke Auswirkungen auf die *force of unification* bzw. die *force of diversification* haben, da somit bewusst mit einem natürlichen Spracheinsatz gebrochen wird. Vergleicht man bspw. das Sprachlernverhalten von Kindern, so sieht man, dass dieses fast natürlich dem Zipfschen Gesetz folgt: Ein Kind lernt pro Tag ca. drei Wörter und je nach Bildungsgrad später bis zu 10 000 insgesamt. Die Auswahl erfolgt dabei zum einen nach der Häufigkeit des Gebrauchs (permanente Wiederholung suggeriert Wichtigkeit) oder nach Neuigkeit und damit einem besonders hohen Informationsgehalt. So ergibt sich auch Zipfs Ansatz, dass der Zusammenhang zwischen Information · Redundanz = Konstante lautet.

#### 6.1.4 Bradfords Gesetz

Das sogenannte Bradfordsche Gesetz oder Bradfords's Law of Scattering wurde als Begriff von Brian C. Vickery (1948) geprägt (Umstätter, 2005b). Bradford selbst beschrieb seine Beobachtung wie folgt:

„... if scientific journals are arranged in order of decreasing productivity on a given subject, they may be divided into a nucleus of journals more particularly devoted to the subject and several groups or zones containing the same number of articles as the nucleus when the number of periodicals in the nucleus and the succeeding zones will be as  $1 : b : b^2 \dots$ “ (Bradford, 1934).

Die Grundaussage von Bradfords Gesetz lautet, dass sich eine immer gleiche Menge von Literatur auf gerankte Zeitschriften nach dem Prinzip  $1 : b : b^2 \dots$  bzw.  $n^0 : n^1 : n^2 \dots$  verteilt. Ein einfaches Beispiel macht diesen Sachverhalt deutlich: Während die Journale Nature oder Science von vielen Naturwissenschaftlern häufig gelesen werden, gibt es zahllose Zeitschriften, die nur wenige Leser mit einem entsprechenden Spezialinteresse haben. Gleiches gilt in umgekehrter Reihenfolge auch für die thematische Zusammensetzung der Zeitschriften, die von thematisch sehr breit bis hochspeziell reichen. Anders ausgedrückt heißt dies, dass ein Kern (der Nukleus) von wenigen Zeitschriften einen großen Teil der für ein bestimmtes Thema relevanten Literatur enthält (die sogenannten Kernzeitschriften), wobei hingegen eine große Zahl von Zeitschriften existiert, die nur sporadisch zu einer bestimmten Thematik publizieren. Möchte man nun eine vollständige Bibliographie zu einer Thematik erstellen, muss man einen großen organisatorischen, logistischen wie finanziellen Aufwand betreiben, um die Thematik abzudecken (Havemann, 2009, S. 16).

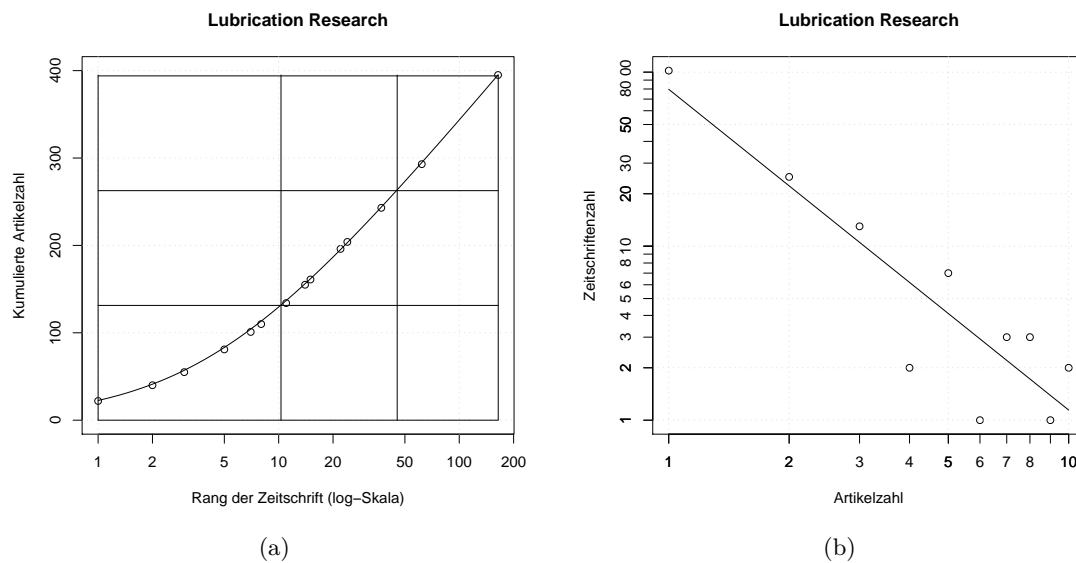


Abbildung 6.4: Verteilung der Zeitschriften nach der Zahl der Beiträge zur Schmiermittel-Forschung, (a) einfach-logarithmiert, die Linien zeigen gleich große Zonen an, (b) doppelt-logarithmische Darstellung (entnommen aus Havemann, 2009, S. 17–18).

Bradford selbst führte seine Analysen aus, indem er die Zeitschriften in der Rangfolge entsprechend der Anzahl der zur Thematik relevanten Artikel ordnete und jeweils gleich große Gruppen bzw. Zonen bildete. In Abbildung 6.4a ist dies durch die horizontalen und vertikalen Linien gekennzeichnet. Der Nukleus besteht aus 10 Zeitschriften, die insgesamt ein Drittel der 398 relevanten Artikel zur Schmiermittel-Forschung (Lubrication Research) beinhalten. Auf den kumulierten Achsen sieht man, dass die zweite Zone 35 weitere Zeitschriften beinhaltet und die dritte Zone ca. 122. Jede Zone beinhaltet folglich ungefähr das 3,5-fache der vorherigen Zone. Wählt man eine andere Darstellungsweise, die zunächst nicht auf kumulierte aber gleichzeitig doppelt-logarithmierte Achsen setzt, erhält man eine für Power-Laws typische Gerade (s. Abb. 6.4b). Die Einteilung, wie in diesem Beispiel in drei Zonen, ist rein willkürlich. Je nachdem wie viele Dokumente einer Zone zugerechnet werden, können beliebig viele Zonen definiert werden.

### 6.1.5 Lotkaische Informetrie und der Informationsproduktionsprozess

Alle zuvor beschriebenen Beobachtungen folgen einer einfachen Regel:

„Eine Wahrscheinlichkeitsverteilung mit einem Quadrat im Nenner, oder einer ersten Potenz im Nenner bei kumulierten Wahrscheinlichkeiten, hat nichts mit den normalen Gauß- oder Poisson-Verteilungen zu tun, oder mit irgendeiner anderen derartigen Kurve, die bei einem gewöhnlichen linearen Maß für zufällige Ereignisse herauskommt. Wenn die Anzahl der wissenschaftlichen Aufsätze



ähnlich verteilt wäre wie die Anzahl der Menschen verschiedener Körpergröße [...], würden wir hohe Punktzahlen weit seltener finden. Wissenschaftliche Aufsätze fallen nicht vom Himmel und sind deshalb auch nicht zufällig verteilt; im Gegenteil, bis zu einem gewissen Grad scheint es umso leichter zu sein, den nächsten zu schaffen, je mehr man schon hat.“ (Umstätter, 1999)

Historisch betrachtet, war es zunächst Davis (laut Potter, 1981, S. 25 f.), der frühe bibliometrische Studien miteinander verglich und feststellte, dass sich die beiden Datensätze in eine Lotka-typische Verteilung (Exponent 2) und eine Pareto-typische Verteilung (Exponent 1,5) unterscheiden ließen. Weitere solcher Beobachtungen führten letztlich zur Einsicht von Naranan (1971), dass viele bibliometrische Verteilungen letztlich das gleiche Phänomen beschreiben.

Mathematisch wurde dieser Zusammenhang und eine Überführung der unterschiedlichen Notationen z. B. von Adamic (2000) dargestellt. Adamic zeigt diesen Zusammenhang mit Hilfe eines AOL-Datensatzes von Log-Dateien. Sie zeigt dabei auch, dass die lange bekannten bibliometrischen Gesetzmäßigkeiten auch im WWW wiederzufinden sind. So zeigen sich die gleichen o. g. Phänomene in der Anzahl der Besucher einer Webseite (Adamic u. Huberman, 2000), die Anzahl der Webpage innerhalb einer Website (Huberman u. Adamic, 1999) oder die Anzahl der Links zu einer Webseite (Albert u. a., 1999). Auch neuere Arbeiten z. B. von Schmidt u. a. (2011) zeigen, dass es sich bei all diesen Beobachtungen, seien sie dem Bereich der Bibliometrie, Webometrie oder allgemein der Informetrie zuzuordnen, letztlich um gemeinsame Regelmäßigkeiten handelt. Bei der mathematischen Herleitung wird die Äquivalenz deutlich und die zugrundeliegenden Power-Laws werden als gemeinsames Konstrukt sichtbar.

Die Gemeinsamkeiten der Gesetzmäßigkeiten werden von Egghe (2005) nicht nur mathematisch hergeleitet, sondern auch konzeptionell untermauert. Wird in den jeweiligen Gesetzen jeweils von einer konkreten Entität wie bspw. Zeitschriften, Zeitschriftenartikeln, Autoren oder Quellen gesprochen, schlägt Egghe, vor von *Sources* (für Zeitschriften, Autoren etc.) und *Items* (welche von den Sources produziert werden, z. B. Artikel) zu sprechen. Ganz allgemein fasst er diese Art der mathematischen Analyse der o. g. Phänomene als lotkaische Informetrie (*Lotkaian Informetrics*) und den Prozess, der zu solchen Phänomenen führt, als Informationsproduktionsprozess (*Information Production Process*, IPP) zusammen. Beispiele für Sources und Items sind Autoren, die Artikel schreiben; Bücher in Bibliotheken, die Ausleihen erzeugen; oder Wörter, die in Texten Verwendung finden (s. Tab. 6.4).

Die dem IPP zugrundeliegende Source/Item-Beziehung kann auch als Type/Token-Beziehung bezeichnet werden. Es sollte dabei unterschieden werden zwischen der Size-Frequency-Funktion  $f$  (typischerweise Lotka) und der Rank-Frequency-Funktion  $g$  (typischerweise Zipf). Während bei der Size-Frequency-Funktion für jedes Anzahl  $n \in \mathbb{N}$  von Items die Anzahl  $f(n)$  der Sources bestimmt werden kann, ist auch gleichzeitig für den Rang  $r$  einer Source  $s \in S$  die Funktion  $g(r)$  definiert, die die Anzahl der Items, die zu einer Source gehören, liefert. Beide Funktionen  $f(n)$  und  $g(r)$  sind dual (Egghe, 2005, S. 10–12). Diese Dualität im IPP bedeutet, dass prinzipiell die Gesetzmäßigkeiten von Sources mit Hinblick auf ihre jeweiligen Items als auch die Gesetzmäßigkeiten von

Sources		Items
Autoren	→	Artikel
Zeitschriften	→	Artikel
Artikel	→	Zitationen (von/zu)
Artikel	→	Koautoren
Bücher	→	Ausleihvorgänge
Wörter (Typen)	→	Nutzung von Wörtern in Texten (Token)
Websites	→	Hyperlinks (ein-/ausgehend)
Websites	→	Webpages
Städte/Dörfer	→	Einwohner
Angestellte	→	Ihre Produktivität
Angestellte	→	Ihr Gehalt

Tabelle 6.4: Beispiele für Source/Item Beziehungen in einem IPP (entnommen aus Egghe, 2009, S. 2).

Items mit Hinblick auf deren Sources betrachten werden können. Sources und Items sind austauschbar. Die Size-Frequency-Funktion  $f(n)$  liefert die Anzahl der Sources mit  $n \in \mathbb{N}$  Items, wobei die Rank-Frequency-Funktion  $g(r)$  die Anzahl der Items in der Source mit dem Rang  $r$  liefert. Natürlichsprachlich ausgedrückt, würde dies für die Artikel/Referenzen einmal eine aktive „zitiert“ sowie eine passive „wird-zitiert“-Beziehung bezeichnen. Es ist folglich wichtig, die Art der Source/Item-Beziehung explizit anzugeben, um Verwirrungen vorzubeugen. Prominente Beispiele für solche Source/Item-Beziehungen sind in Tabelle 6.4 aufgelistet. Egghe verwendet in seinen Arbeiten den Parameter  $n$  für  $f(n)$ , da in dieser Arbeit jedoch allgemeiner von Power-Laws gesprochen wird, wird hier die Bezeichnung  $f(x)$  synonym verwendet.

### 6.1.6 Anwendung bibliometrischer Gesetze

Wie auch das Beispiel der Anwendbarkeit für informatriche Untersuchungen auf das Internet zeigt, ist zu bedenken, dass konkrete Anwendungen der Informatric nicht dem reine Erkenntnisgewinn dienen. Eine Studie von McCallum u. Godwin (1981) zeigte bspw., dass meist ein organisatorisches Interesse im Vordergrund stand. So sollte eine Entscheidungsgrundlage für die Neuauflage einer zweiten Ausgabe der anglo-amerikanischen Katalogregeln geschaffen werden. Hierzu wurde Bradfords Gesetz ausgenutzt, um bestimmen zu können, ob Periodika neu gegründet werden sollten oder nicht. Es bietet aber auch die Möglichkeit, herauszufinden, welche Zeitschriften für das eigene Forschungsinteresse besonders wichtig sind. Dies ist vor allem für Anfänger einer Disziplin hilfreich, da ihnen der nötige Überblick über ein Fach möglicherweise fehlt, um eigene Relevanzabschätzungen abgeben zu können.

Zipfs Gesetz wird beispielsweise eingesetzt, um Index-Kompressionen zu erreichen (Manning u. a., 2008, S. 85–89) oder um Stoppwortlisten zu generieren.

Technische Entwicklungen, wie z. B. Google Instant<sup>1</sup>, können einen Einfluss auf alte Gewohnheiten, die mit den bekannten Eigenschaften der informetrischen Verteilungen einhergingen, außer Kraft setzten, sodass diese nicht mehr zwangsläufig Geltung haben. So Ellis (2010) provokante Frage zum Einsatz von Google Instant: „Will Google kill the Long Tail?“. Diese und weitere Auswirkungen und Anwendungen informetrischer Analysen auf Suchmaschinen und das Information Retrieval werden in Kapitel 7 betrachtet.

### 6.1.7 Eigenschaften von Power-Laws

Nachdem nun die Power-Laws als charakteristisches Merkmal hinter der Lotkaischen Informetrie und den IPPs aufgezeigt wurden, sollen im folgenden Abschnitt einige Eigenschaften der Power-Laws beschrieben werden. Zunächst werden diese auch als skalenfrei (*scale-free*) bezeichnet. Dies sagt aus, dass eine Erhöhung um einen bestimmten Faktor für den Wert von  $x$  eine entsprechende Verringerung (oder Erhöhung) für  $f(x)$  bedeutet. Allerdings hat Newman (2005) nachgewiesen, dass die Power-Law-Verteilung, die einzige skalenfreie Verteilung ist, sodass die beiden Bezeichnungen synonym verwendet werden können. Im Rahmen dieser Arbeit wird allerdings die Begrifflichkeit der Power-Laws bevorzugt verwendet.

Abbildung 6.5 zeigt zwei typische Visualisierungen für eine lehrbuchhafte Power-Law-Funktion ( $c = 10\,000$ ,  $a = 2$ ). Zum einen ist die charakteristische, stark abfallende Kurve in der ersten Visualisierung mit nicht-skalierten Achsen zu sehen. Man erkennt sehr gut den Long-Tail, der gegen die y-Achse konvergiert. Die gleiche Funktion kann auch mit einer doppelt logarithmierten Skala gezeichnet werden, was durchaus üblich ist. In dieser Ansicht erkennt man eine Gerade, was als ein erstes Indiz für eine Power-Law-Verteilung der gemessenen Daten angesehen werden kann. In Abschnitt 6.1.9 wird neben der grafischen Identifikation auch ein mathematisches Verfahren vorgestellt, um Power-Law-Verteilungen zu erkennen und den Exponenten  $\alpha$  zu errechnen.

### 6.1.8 Vergleich von empirisch ermittelten und formalen Power-Laws

Vergleicht man empirische Power-Law-Verteilungen, also gemessene Werte wie Stadtpopulationen oder Telefonanrufe (vgl. die vorherigen Abschnitte), mit mathematischen bzw. formalen Power-Laws so wird schnell klar, dass empirisch hergeleitete Power-Law-Verteilungen nur eine Approximation ihrer mathematischen Beschreibungen sind. Durch die zwangsläufige Vereinfachung eines umfangreichen Sachverhaltes in die Form eines mathematisch simplen Power-Law-Modells ist eine Abweichung nicht zu vermeiden. Die durch Modelle beschriebenen Sachverhalte sind meist komplexerer Natur oder ihre Umstände mathematisch schwieriger zu beschreiben. Dies gilt auch für die in dieser Arbeit betrachteten Anwendungsfälle der empirisch gemessenen Verteilungen im IPP. Wie in allen Modellbildungsprozessen, nimmt man mit der Approximation eine Abweichung von der wirklichen Verteilung in Kauf. Können einfache Power-Laws die gemessenen Datenlage nicht ausreichend genug wiedergeben oder die Repräsentation eine systematische Abweichung aufweisen, sollte eine Anpassung des einfachen Power-Law-Modells in

<sup>1</sup><http://www.google.com/instant>

## 6 Informetrie

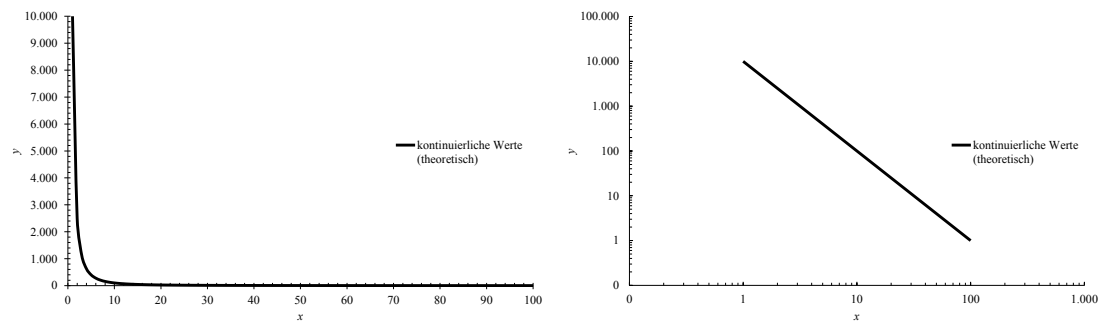


Abbildung 6.5: Plot derselben Power-Law-Funktion mit den Werten  $c = 10\,000$  und  $\alpha = 2$ . Während der erste Plot eine natürliche, nicht-skalierte Achse besitzt, wurden im zweiten Plot beide Achsen logarithmisch skaliert. Die dargestellte Funktion ist ein typischer Vertreter für ein theoretisches Power-Law, das auf kontinuierlichen Werten basiert.

Betracht gezogen werden. Milojević (2010b) fasst insgesamt vier solcher Abweichungen zusammen, die in Abbildung 6.6 dargestellt und nachfolgend beschreiben werden.

Zunächst beschreiben Csányi u. Szendroblaci (2004, S. 036131-2) eine typische Abweichung, die sich durch die Existenz zweier unterschiedlicher Exponenten für Formel 6.1 auszeichnen. In doppelt logarithmierter Darstellungen sind nicht eine, sondern zwei Geraden zu sehen.

In Glänzel (2007, S. 94) wird die Pareto-Verteilung zweiter Ordnung („Pareto distribution of the second kind“) oder die sogenannte Lomax-Verteilung vorgestellt, welche vorliegt wenn für eine positive Zufallsvariable  $X$  gilt:

$$G(x) = P(X \geq x) = \frac{N^\alpha}{(N + x)^\alpha}, \quad \text{für alle } x \geq 0, \quad (6.5)$$

wobei  $N$  und  $\alpha$  positive reelle Zahlen sind. Alternativ wird laut Glänzel auch die Parametertransformation  $\alpha = \alpha/(\alpha - 1)$  angewendet. Praktisch angewendet bedeutet dies, dass die Lomax-Verteilung zunächst für kleine Werte von  $x$  nahezu konstant erscheint (im Plot erkennt man eine fast horizontale, schwach abfallende Linie), wobei für große Werte von  $x$  eine reguläre Power-Law-Verteilung sichtbar ist.

Eine weitere Variante sind die in Newman (2005) vorgestellten Power-Laws mit einem exponentiellen Cut-off. Hier wird für kleine Werte von  $x$  eine normale Power-Law-Verteilung beobachtet, sodass im Plot zunächst eine gerade Linie zu erkennen ist. Für größere Werte von  $x$  fällt die Gerade allerdings ab und ähnelt dem Plot einer exponentiellen Funktion. Der exponentielle Anhang für große  $x$  fällt also schneller als das Power-Law.

Eine von Milojević (2010a) selbst beschriebene Variante ist die der Log-Normal/Power-Law-Verteilungen, bei denen für große  $x$  eine normale Power-Law-Verteilung zu sehen ist, für kleine  $x$  allerdings logarithmische Normalverteilung vorliegt. Diese Normalverteilung zeigt sich im doppelt logarithmierten Plot durch eine Rundung, die zusätzlich ein Maxi-

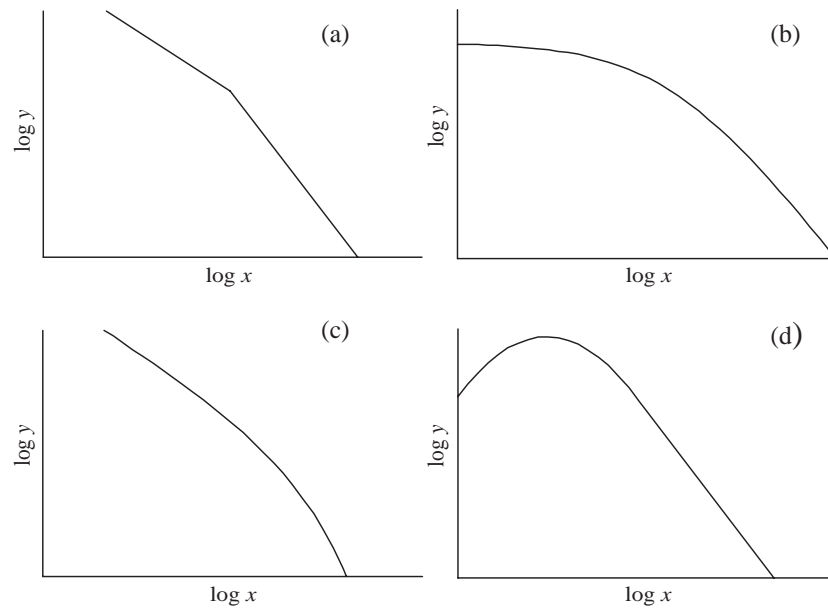


Abbildung 6.6: Unterschiedliche Abweichungen von Power-Laws: (a) doppelte Power-Laws, (b) Lomax oder Pareto Verteilung zweiter Ordnung, (c) Power-Laws mit einem exponentiellen Cut-off und (d) einer Log-Normal/Power-Law-Verteilung (entnommen aus Milojević, 2010b, S. 2419).

mum besitzen kann – gegensätzlich zu normalen Power-Law-Verteilungen oder den drei vorgestellten Varianten, die kein solches Verhalten zeigen und alle monoton fallend sind.

Neben den genannten vier Variationen gibt es einen weiteren Faktor, der einen Unterschied zwischen formalen und empirisch ermittelten Verteilungen erklärt. Die formale Definition geht von einem kontinuierlichen Zahlenraum für  $x$  aus, tatsächlich gemessen werden können aber nur diskrete Werte  $x \in \mathbb{N}$ , die zu einer Verzerrung in der Darstellung führen. Diese Verzerrungen sind im Vergleich der beiden Abbildungen 6.5 für theoretische, kontinuierliche Werte und 6.7 für empirisch ermittelte diskrete Werte zu sehen. Bei theoretisch, künstlich generierten Werten ist der sogenannte *Fat-Tail* zu sehen, der auf dem Phänomen beruht, dass es wesentlich mehr Werte zum Ende der Verteilung gibt. Die gedachte Linie ist aber sowohl hier als auch bei empirisch gemessenen diskreten Werten klar zu erkennen. Bei den abgebildeten diskreten Werten handelt es sich um einen Plot der Verlagsinformationen aus der Datenbank Bibsonomy (verwendet wurde der Bibsonomy-Dump-2011-01-01). Klar zu erkennen ist sowohl der Fat-Tail als auch das für empirisch ermittelte Verteilungen typische Zittern im Tail.

Es ist üblich, für das kontinuierliche Modell die Variable  $x$  und für die diskrete Variante die Variable  $k$  zu verwenden, wie dies auch in den Abbildungen an den Achsenbeschriftungen zu sehen ist. Weiterhin kann gezeigt werden (Egghe, 2005, S. 378 ff.), dass die kontinuierliche Funktion

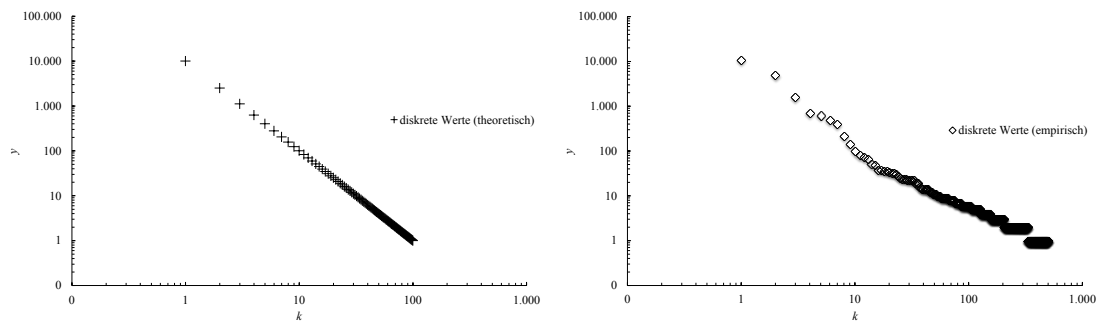


Abbildung 6.7: Plot zweier diskreter Power-Law-Verteilungen, basierend auf einer theoretisch und einer empirisch ermittelten Werteverteilung zu Verlagsinformationen in der Datenbank Bibsonomy.

$$f(j) = \frac{C}{j^\alpha}, \quad (6.6)$$

die gleichbedeutend mit Formel 6.1 ist, für  $j \in [1, \rho_m]$  für  $\rho_m = \infty$  mathematisch auf die folgende diskrete Form abgebildet werden kann:

$$f_d(n) = \frac{K}{n^a}. \quad (6.7)$$

Hierbei ist  $n = 1, 2, \dots, n_{max}$  und  $K$  konstant. Man beachte die unterschiedliche Benennung des Exponenten, die  $\alpha$  für die kontinuierliche Funktion und  $a$  für die diskrete Variante lautet. Diese Unterscheidung stammt aus der Arbeit von Egghe (2005), sie ist aber in anderen Arbeiten nicht unbedingt üblich, sodass in dieser Arbeit sowohl für die kontinuierliche als auch die diskrete Variante die Variante  $\alpha$  verwendet wird.

Eine weitere Diskrepanz ist anhand der Beispielrechnung zu Lotkas Gesetz aus Abschnitt 6.1.2 und einer an Milojević (2010b) angelehnten Rechnung zu erklären. Für den Exponenten  $\alpha = 2$  und dem normalisierenden Faktor  $c = 100$  würden 6,25 Autoren vier Papiere verfasst haben ( $f(4) = 100 * 4^{-2} = 100/4^2 = 6,25$ ). In einem wirklichen Beispiel muss der Wert 6,25 natürlich auf eine natürliche Zahl abgebildet werden, üblicherweise wird auf den Wert 6 abgerundet. Die Rundungsproblematik kann aber nicht den wahren Kern des Problems beschreiben, da empirisch ermittelte Power-Laws nicht einfach nur gerundete Varianten eines theoretischen Power-Laws sind. Vielmehr muss die probabilistische Eigenart der empirischen Verteilungen betrachtet werden, welche besagt, dass der ermittelte Wert von 6,25 nur den Mittelwert aller Verteilungsmöglichkeiten darstellt. Mittels einer Poisson-Verteilung mit dem Mittelwert 6,25 ist zwar der wahrscheinlichste Ausgang der Wert 6 mit einer Wahrscheinlichkeit von knapp 16 %, aber selbst der Wert 0 ist möglich, der Wert 5 mit einer Wahrscheinlichkeit von 15 % fast genauso wahrscheinlich wie der Wert 6. Man spricht von statistischem Rauschen (*noise*), das zwar immer vorhanden ist, aber am deutlichsten im Tail zum Vorschein kommt. Die übliche Power-Law-Verteilung liefert große Werte für kleine  $k$  (wenige Autoren  $k$  mit vielen

Papieren  $f(k)$ ), wobei das Rauschen kaum wahrnehmbar ist. Für große  $k$  variiert der Wert von  $f(k)$  zunehmend und wird sichtbar. Die Schwankungen sind als Zittern im Tail zu erkennen (s. Abb. 6.7). Man sollte dabei nicht den Fat-Tail mit Zittern verwechseln. Beide sind in Abbildung 6.7 zu sehen. Der Fat-Tail ist eine normale Eigenschaft jeder Power-Law-Verteilung, da es wesentlich mehr Werte zum Ende der Verteilung gibt, die in Plot übereinander abgebildet werden und dadurch fetter wirken. Es handelt sich dabei um ein reines Darstellungsproblem.

Eine Möglichkeit, dem Zittern im Tail entgegenzuwirken, ist das sogenannte *Binning*, bei dem die Werte zu Gruppen zusammengefasst werden und so eine Darstellung mit geringeren Fehlern ermöglichen. Eine solche Fehlerbereinigung in der Darstellung und die Ermittlung der jeweiligen Variante bzw. die weiteren Feinheiten in der Bestimmung einer reinen Power-Law-Verteilung liegen außerhalb des Schwerpunktes dieser Arbeit. Eine Vielzahl an Methoden zur Bestimmung beschreiben z. B. Clauset u. a. (2009). Es ist zudem gängige Praxis bei typischerweise unsauberen Verteilungen im IPP und in der Informationswissenschaft von Power-Laws zu sprechen, auch wenn diese bei genauerer Betrachtung nicht als reine Power-Laws angesehen werden dürften, sondern eine der obigen Varianten bzw. einer vollkommen anderen Verteilung zugerechnet werden müssten. Einer der Gründe hierfür ist die Einfachheit des Modells, das im Prinzip durch einen einzigen Wert beschrieben werden kann, den Exponenten. Der folgende Abschnitt beschäftigt sich daher mit den Verfahren zur Ermittlung des Exponenten.

### 6.1.9 Ermittlung des Power-Law-Exponenten

Erlaubt der Plot einer gemessenen Verteilung durch Approximation einer Linie in der doppelt logarithmierten Darstellung, eine erste Einschätzung als Power-Law-Verteilung, kann der Exponent  $\alpha$  ermittelt werden. Die einfachste Variante, den Exponenten zu bestimmen ist der Einsatz des Standardverfahrens zur Ausgleichsrechnung, der Methode der kleinsten Quadrate. Während dieses Standardverfahren bei künstlich erzeugten Verteilungen ohne das statistische Zittern im Tail zum richtigen Ergebnis führt, ist es nicht geeignet für realistische Daten, wobei hier auch das im vorherigen Abschnitt vorgestellte *Binning* helfen kann, den Fehler zu minimieren.

Newman (2005) stellt eine einfache Methode vor, um aus den Daten den Power-Law-Exponenten  $\alpha$  ermitteln zu können, ohne dabei auf Hilfsverfahren wie Binning zurückgreifen zu müssen:

$$\alpha = 1 + n \left[ \sum_{i=1}^n \ln \frac{x_i}{x_{min}} \right]^{-1}. \quad (6.8)$$

Die Werte für  $x_i, i = 1 \dots n$  sind die gemessenen Werte von  $x$  und  $x_{min}$  bezeichnet den kleinsten  $x$ -Wert. Allerdings kann das Verfahren nach Newman nur auf die kontinuierliche Variante angewendet werden und scheitert ebenfalls bei diskreten Werten. Clauset u. a. (2009) und Egghe (2005, S. 387 ff.) stellen insgesamt drei Verfahren vor, um den Exponenten auch auf der Grundlage von diskreten Werten ermitteln zu können. Dort sind auch die Herleitungen zu den folgenden Formeln zu finden.

Die einfachste und schnellste Methode ist danach die sogenannte Quick-and-Dirty-Methode. Bei ihr wird  $n_{max} = \infty$  definiert,  $A$  beschreibt die Gesamtzahl der Items und  $T$  die Gesamtzahl der Quellen (wobei  $A > T$ ). Die beiden fehlenden Parameter zur Lösung der Gleichung 6.7 sind somit der Exponent  $\alpha$  und die Konstante  $K$ , die wie folgt ermittelt werden können:

$$\alpha = \frac{\ln\left(\frac{f(1)}{f(2)}\right)}{\ln 2} \quad (6.9)$$

sowie

$$K = f(1). \quad (6.10)$$

Die Annahme, dass  $K$  gleich der Anzahl von Quellen für ein Item ist, erleichtert die Bestimmung. Er kann direkt aus den vorhandenen Daten ausgelesen werden. Gleiches gilt für den Exponenten  $\alpha$ . Allerdings ist kritisch anzumerken, dass bei dieser Quick-and-Dirty-Methode Messfehler in den ersten beiden Werten immense Auswirkungen haben, wobei die ersten beiden Werte meist wesentlich stabiler sind, als dies z. B. im Tail der Fall ist.

Zwei zuverlässigere Methoden sind das Linear-Least-Square-Verfahren, das bereits von Lotka selbst angewendet wurde (Lotka, 1926) und eine Maximum-Likelihood Estimation, die wie folgt berechnet wird:

$$\frac{\sum_{n=1}^{n_{max}} f(n) \ln n}{\sum_{n=1}^{n_{max}} f(n)} = -\frac{\xi'(\alpha)}{\xi(\alpha)}. \quad (6.11)$$

Mittels  $\frac{-\xi'(\alpha)}{\xi(\alpha)}$  kann der Exponent  $\alpha$  errechnet werden. Bei  $\xi$  handelt es sich um die Riemann-Zeta-Funktion. Durch Einsetzen aller  $n$  diskreten Datenpunkte erhält man einen Wert, der in einer vorberechneten Wertetabelle nachgeschlagen werden kann. Beide Verfahren beruhen auf solchen Wertetabellen zur Bestimmung des Exponenten. Diese in dieser Arbeit wiederzugeben erscheint nicht zielführend. Für die praktische Nutzung ist die Arbeit von Rousseau u. Rousseau (2000) zu empfehlen. Die entsprechenden Formeln, Wertetabellen und Verfahren wurden von beiden Autoren in dem frei verfügbaren Programm LOTKA implementiert.

## 6.2 Kookkurrenzmodelle

Neben der Analyse von IPP-Beziehungen mittels der lotkaischen Informatrie ist die Kookkurrenzanalyse eine bekannte Methodik für informatrische Untersuchungen. Diese kommt allerdings nicht nur in der Informatrie zum Einsatz, sondern z. B. auch in Disziplinen wie der natürlichen Sprachverarbeitung (*natural language processing*, NLP). Sie wird in der Standardliteratur ausgiebig erläutert, z. B. bei Manning u. Schütze (1999). Im IR wird die Kookkurrenzanalyse und ihre Modelle hauptsächlich eingesetzt, um mit ihrer



Hilfe Term-Term Beziehungen zu analysieren und so z. B. Synonyme oder Phrasen zu lokalisieren. Hierbei sind die Begriffe der Kollokation und der Kookkurrenz voneinander zu trennen. Die Kollokation (*collocation*) beschreibt in der Linguistik das vermehrte Auftreten grammatikalisch miteinander verbundener und daher benachbarter Wörter, ohne dabei auf die Gründe des Auftretens genauer einzugehen. Von Kookkurrenz (*co-occurrence*) wird im weiteren Sinne gesprochen, wenn das allgemeinere Phänomen des Auftretens von Wörtern im gleichen Kontext gemeint ist. Die Kollokation ist in ihrer Definition enger als die eher breite Beschreibung der Kookkurrenz.

Auch in der Anwendung für die Informatik zeigt sich, dass die Kookkurrenzanalyse ein wesentlich breiteres Anwendungsspektrum besitzt; so gibt es dort Analysen zu Kozitationen (*co-citations*), Koautorenschaften (*co-authorships*), Koklassifikationen (*co-classifications*) oder Koworten (*co-words*). Die so gewonnenen Daten können genutzt werden, um Karten der wissenschaftlichen Landschaft zu zeichnen oder andere Graphen, die einen Überblick über die Beziehungen von Wissenschaftlern oder Disziplinen zueinander erlauben. Neben der Anwendung in Form von Visualisierungen jeglicher Art wird die Kookkurrenzanalyse genutzt, um bibliometrische Rankings aufzubauen. Weitere Anwendungen von bibliometrischen Kookkurrenzen werden bei Tijssen u. Van Raan (1994) genannt. Generell lassen sich die unterschiedlichen Analyseverfahren auf drei Beziehungsmuster zurückführen: *Item-Item*-, *Entität-Entität*- oder *Item-Entität*-Beziehung. Ein Item ist in diesem Zusammenhang ein Dokument und eine Entität ist ein zum Dokument gehörendes Attribut, wie z. B. das zugehörige Journal oder Autoren.

Die Nutzung der Kookkurrenzanalyse ist nicht unumstritten. So beschreibt Ferber (1996), dass sie für die Anwendung in der Anfragenerweiterung nur einen begrenzten Wert habe, wenn ein falsches Kookkurrenzmaß gewählt werde. Auch die Größe und Struktur der analysierten Datenmenge sei entscheidend für eine erfolgreiche Anwendung solcher Maße. Aus linguistischer Sicht betrachtet Kuropka (2005) die Kookkurrenzanalyse und kommt zu dem Schluss, dass besonders die Annahme, dass Terme unabhängig voneinander auftreten (in Verbindung zur Bag-of-Words-Annahme) nicht korrekt sei. Dies werde bei sprachlichen Konstrukten wie Inflektionen, Wortgruppen oder Komposita deutlich. Er spricht sich daher für eine stärkere Anwendung von statistischen Modellen wie dem Latent-Semantic-Indexing oder statistischen Sprachmodellen aus. Trotz allem werden auch gerade die recht simplen Verfahren, wie z. B. der Jaccard-Index, noch häufig angewendet und können, entsprechende Vorverarbeitung und Datenpflege vorausgesetzt, zu guten Ergebnissen führen. Eine recht aktuelle Umsetzung der klassischen Kookkurrenzanalyse ist die sogenannte normalisierte Web-Distanz (*Normalized Web Distance*) oder Google-Distanz (Cilibrasi u. Vitányi, 2010). Hier sollen durch die Verwendung der sehr großen Suchindizes kommerzieller Suchmaschinen wie Google oder Yahoo die eingangs erwähnten Schwächen, die z. B. durch die Verwendung von zu kleinen Datensätzen entstehen können, umgangen werden. Eine gute Übersicht über die effiziente Implementierung von verschiedenen Verfahren der Kookkurrenzanalyse findet sich bei Büchler (2006).

Im Folgenden werden drei Möglichkeiten zur Kookkurrenzanalyse vorgestellt, die in zuvor beschriebenen Arbeiten zum Einsatz kamen: (1) der Jaccard-Index (z.B. bei Schaer u. a., 2012c,b), (2) die Likelihood-Ratio (z.B. bei Petras, 2006) und (3) Machine-Learning-Verfahren (z.B. bei Schaer u. a., 2011).

### 6.2.1 Jaccard-Index

Der Jaccard-Index (auch Jaccard-Koeffizient) wurde ursprünglich von Jaccard (1901) vorgeschlagen und für seine botanischen Studien verwendet. Im Nachhinein wurde er aber für viele andere Zwecke verwendet, so z. B. die Kozitationsanalyse (Hamers u. a., 1989). Allgemein wird durch den Jaccard-Index die Ähnlichkeit von Mengen ausgedrückt.

Um die Kookkurrenz unterschiedlicher Dokumenteneigenschaften miteinander vergleichen zu können, kann der Jaccard-Index verwendet werden. Hierzu wird der Koeffizient zwischen der Menge der gemeinsamen Element und der Vereinigungsmenge gebildet. Angewendet auf die Dokumentmenge  $D$  und zweier Dokumentattribute  $x$  und  $y$  können die beiden Untermengen  $D_x$  und  $D_y$  gebildet werden. Hierbei kann  $x$  z. B. ein Term sein und  $y$  ein Autorenname.  $D_x$  ist nun die Menge aller Dokumente, die den Term  $x$  enthalten und  $D_y$  die Menge aller Dokumente, die den Autorennamen  $y$  beinhalten. Vereinfacht können auch die Dokumentfrequenzen  $df_x$ ,  $df_y$  und  $df_{xy}$  bestimmt werden, die jeweils die Anzahl der Dokumente mit den Attributen  $x$ ,  $y$  bzw. deren gleichzeitiges Auftreten wiedergeben:

$$J(x,y) = \frac{|D_x \cap D_y|}{|D_x \cup D_y|} = \frac{df_{xy}}{df_x + df_y - df_{xy}}. \quad (6.12)$$

Der Jaccard-Index normalisiert die Attribut-Attribut-Ähnlichkeiten zwischen 0 und 1, wobei 0 keinerlei Ähnlichkeit bedeutet und 1 die maximale Ähnlichkeit ist. Wird  $Jm(x,y) = 1 - J(x,y)$  verwendet, wird aus dem Jaccard-Index eine Metrik, die anstelle der Ähnlichkeit den Abstand zwischen den Attributen beschreibt (Schaer u. a., 2012b).

Da die Anzahl der Elemente der beiden Mengen  $D_x$  und  $D_y$  meist sehr groß ist, kann es in der Berechnung zu sehr kleinen Jaccard-Indexwerten kommen. Es kann daher ratsam sein, die großen Unterschiede durch das Logarithmieren von Zähler und Nenner auszugleichen:

$$J_{\log}(x,y) = \frac{\log(df_{xy})}{\log(df_x + df_y - df_{xy})}. \quad (6.13)$$

### 6.2.2 Likelihood-Ratio

Plaunt u. Norgard (1998) als auch Buckland u. a. (1999) verwendeten einen statistischen Likelihood-Ratio-Ansatz, um die Beziehungen zwischen Begriffen der natürlichen Sprache und eines kontrollierten Vokabulars zu analysieren. Sie konnten damit die Wahrscheinlichkeit voraussagen, mit der ein kontrollierter Begriff zu einem beliebigen Suchbegriff in Verbindung steht. Das eingesetzte Verfahren basiert dabei auf der Auszählung von Häufigkeiten in denen ein Ereignis  $A$  und ein Ereignis  $B$  vorkommen. In dem konkreten Anwendungsfall waren dies die Wörter  $A$  und  $B$ , jeweils aus dem Titel/Abstract und den vergebenen Schlagworten einer Arbeit.

Die vollständige Herleitung der folgenden Gleichungen ist bei Plaunt u. Norgard (1998) ausführlich dargestellt. Zur Veranschaulichung wird hier nur die endgültige Gleichung gezeigt. Dabei ist  $-2 \log \lambda$  das Gewicht, nach dem die Assoziationsstärke zwischen zwei

Wörtern A und B bemessen wird, indem man  $k_1$ ,  $k_2$ ,  $n_1$  und  $n_2$  kennt:

$$-2 \log \lambda = 2 [\log L(p_1, k_1, n_1) + \log L(p_2, k_2, n_2) - \log L(p, k_1, n_1) - \log L(p, k_2, n_2)], \quad (6.14)$$

wobei

$$\log L(p, k, n) = k \log p + (n - k) \log(1 - p) \quad (6.15)$$

sowie

$$p_1 = \frac{k_1}{n_1}, p_2 = \frac{k_2}{n_2} \text{ und } p = \frac{k_1 + k_2}{n_1 + n_2}. \quad (6.16)$$

Zur Auflösung der Gleichung werden noch die Werte  $k_1$ ,  $k_2$ ,  $n_1$  und  $n_2$  benötigt, die durch die möglichen Kombinationen der Vorkommnisse der beiden Wörter  $A$  und  $B$  ermittelt werden:  $AB$ ,  $A\bar{B}$ ,  $\bar{A}B$  und  $\bar{A}\bar{B}$ , wobei  $\bar{\phantom{x}}$  die Abwesenheit des Wortes kennzeichnet. Durch das Auszählen der Häufigkeit des gemeinsamen Auftretens werden  $k_1$ ,  $k_2$ ,  $n_1$  und  $n_2$  wie folgt bestimmt:  $k_1 = AB$ , die Häufigkeit des gemeinsamen Auftretens von  $A$  und  $B$ ,  $n_1 = AB + \bar{A}B$ , die Gesamthäufigkeit der Beobachtungen.  $k_2$  und  $n_2$  sind die Werte der Häufigkeit unter der Voraussetzung, dass  $B$  nicht präsent ist:  $k_2 = A\bar{B}$  und  $n_2 = A\bar{B} + \bar{A}\bar{B}$ .

Von einer gegebenen Trainingsmenge können nun freie Terme aus Titel oder Abstract und kontrollierte Terme aus dem zur Verschlagwortung genutzten Thesaurus miteinander in Beziehung gesetzt werden. Hierzu wird ein Wörterbuch aus Kowörtern aufgebaut, welches für jede Kombination die Stärke der Assoziation (den berechneten Wert  $-2 \log \lambda$ ) enthält. Mit Hilfe des Wörterbuchs kann die Wahrscheinlichkeit der Ähnlichkeit nachgeschlagen werden.

### 6.2.3 Machine-Learning und automatische Klassifikationssysteme

Neben der vorgestellten Berechnung der Assoziationsstärke mittels Kookkurrenzanalysen wie dem Jaccard-Index oder der Likelihood-Ratio können auch Verfahren des maschinellen Lernens und der automatischen Klassifikation verwendet werden. Diese Verfahren wurden bereits im Kontext des Learning-to-Rank in Abschnitt 4.5.2 beschrieben. In der klassischen Informatik sind Verfahren des maschinellen Lernens eher ungewöhnlich. Häufiger eingesetzt werden hingegen Verfahren wie *Support Vector Machines* (SVM), als ein Hilfsmittel zur Generierung zusätzlicher Informationen, die dann von klassischen Informatikverfahren analysiert werden können. So setzten Fu u. Aliferis (2010) maschinelles Lernen ein, um automatisch zukünftige Zitationsraten abschätzen zu können und diese dann zu analysieren, doch sind solche Anwendungen von Machine-Learning die Ausnahmen.

Die verwendeten Verfahren können der automatischen Klassifikation zugerechnet werden, die mit einem überwachten Lernprozess (*supervised learning*) arbeiten. Hierbei ist eine Grundmenge an Zuordnungen bereits bekannt (z. B. Verschlagwortungen aus einem Thesaurus für eine Menge von Freitexten), deren Struktur dann von dem Algorithmus gelernt wird. Andere Ansätze, wie z. B. die latente semantische Analyse (*Latent Semantic*

*Analysis*, LSA), sind der Gruppe des unüberwachten Lernens (*unsupervised learning*) zuzurechnen, bei denen die Struktur bzw. ein zugrundeliegendes Modell zunächst vom Algorithmus erzeugt und gelernt wird. Die Berechnungen, die diesen beiden Verfahren zugrundeliegen, werden z. B. in kommerziellen Indexierungsprogrammen wie dem Recommind Mindserver eingesetzt. Alternative Implementierungen arbeiten auch mit Klassifikatoren, die das Verfahren des naiven Bayes einsetzen, so z. B. das freie Klassifikationstoolkit Apache Mahout (Sean u. a., 2011) oder die Statistiksoftware R (Karatzoglou u. a., 2006).

### 6.3 Zusammenfassung

Wie eingangs beschrieben sind die beiden Disziplinen IR sowie die Informatric in vielen Anwendungsfällen disjunkt. Obwohl beide Disziplinen sich mit dem gleichen Forschungsobjekt (in unserem Falle Dokumente einer wissenschaftlichen digitalen Bibliothek) auseinandersetzen, haben sie doch einen anderen Blickpunkt darauf. Aus Sicht der Informatric ist das IR oft nur eine Hilfswissenschaft, die genutzt wird, um den Datensatz, der für die informatricische Analyse verwendet wird, zu verarbeiten (Ingwersen, 2012a). Eine wirkliche Verbindung der beiden Disziplinen, z. B. in Form von informatricisch-motivierten Mehrwertdiensten für das Retrieval wie u. a. von White (2007a) oder Mayr (2009b) propagiert, ist selten.

In den beiden vorherigen Abschnitten wurden zwei Verfahren der Informatric vorgestellt: Die Analyse von Verteilungsinformationen im Informationsproduktionsprozess (IPP) und deren Verbindung zu sogenannten Power-Laws und die Berechnung von Konkurrenz beliebiger Entitäten einer Dokumentmenge. Diese sollen als methodische und konzeptionelle Grundlage für informatricisch-motivierte Mehrwertdienste dienen, die den Nutzer im Retrievalprozess unterstützen. Diese Mehrwertdienste werden im folgenden Teil II vorgestellt.

## **Teil II**

# **Verfahren und Materialien**



# 7 Nicht-textuelle Mehrwertdienste für das Dokumentenretrieval

Im folgenden Abschnitt werden zwei sogenannte nicht-textuelle Mehrwertdienste bzw. Verfahren zur Suchunterstützung vorgestellt, die in den späteren Kapiteln evaluiert werden. Zunächst wird ein Rerankingverfahren vorgestellt, das auf Grundlage von Häufigkeitsverteilungen in Metadaten und deren informatrischen Analyse eine alternative Sortierung für Ergebnislisten ermöglicht. Des Weiteren wird mittels der gleichen Metadaten, aber mit anderen Auswertungsverfahren, eine Anfrageerweiterung implementiert. Beide Verfahren sollen typische Probleme bei der Suche in digitalen Bibliotheken umgehen bzw. abmildern.

## 7.1 Anfrageerweiterung

Wie in Abbildung 1.2 dargestellt, ist die Anfrageerweiterung eine zentrale Komponente innerhalb des Retrievalprozesses, neben der Beeinflussung des Rankings, die gleich mehrere Probleme des klassischen Dokumentenretrievals zu beheben versucht (s. Kap. 3).

### 7.1.1 Nicht-textuelle Anfrageerweiterung

Die generelle Idee hinter der Anfrageerweiterung mit bibliometrischen Entitäten ist es, zusätzliches Wissen, das in den Metadaten der Dokumente bzw. in den Entität-Dokumentbeziehungen verborgen ist, nutzbar zu machen (s. Abb. 6.3). Ein Beispiel hierfür können die Namen von Autoren sein, die eine Reihe von relevanten Dokumenten verfasst haben. Wenn man die Beziehung zwischen der ursprünglichen Suchanfrage und den dazu einschlägigen Autorennamen extrahieren kann, kann man diese Namen zur Anfrageerweiterung verwenden. Dies setzt voraus, dass Autoren, Journale oder andere bibliografische Entitäten einem gewissen Muster in ihrer Thematik oder ihren Beziehungen untereinander folgen.

Im folgenden Ansatz wird ein Verfahren der Kookkurrenzanalyse, der in Abschnitt 6.2.1 vorgestellte Jaccard-Index, verwendet. Mit diesem Verfahren kann die semantische Distanz bzw. die semantische Verbindung zwischen beliebigen, freien Termen und gleichzeitig auftretenden (kookkurrierenden) Entitäten bestimmt werden. Freie Terme sind z. B. solche, die in Titel oder Abstract einer Arbeit vorkommen. Die anderen Entitäten können Autoren, Journale oder Thesaurusterme aus der gesamten Dokumentmenge sein. Mittels des Jaccard-Index kann die Ähnlichkeit bzw. semantische Beziehung zwischen verschiedenen Entitäten eines Datenbestandes quantifiziert werden. So können möglichst nahe Autoren oder Thesaurusterme zu einem beliebigen Thema, das durch frei wählbare Begriffe beschrieben wird, bestimmt werden.

Angewendet auf das Beispiel einer Suche nach „Gesundheit im Ruhestand“ sind dies u. a. die Autoren „Hauser, Richard“ und „Bäcker, Gerhard“. Beide sind Sozialwissenschaftler, die Arbeiten im Bereich der Sozial- und Rentensysteme sowie Gesundheitsfragen geschrieben haben. Weiterhin wird bei dieser Anfrage eine Auswahl an Thesaurustermen vorgeschlagen, die z. B. die Terme „Sozialpolitik“ oder „Ältere Menschen“ beinhalten. Für das konkrete Verfahren der Anfragerweiterung werden die höchstbewerteten (gemessen am Jaccard-Index)  $n = 4$  Attribute bzw. Entitäten zur Anfrage hinzugefügt. Dieser Wert basiert auf vergleichbaren Arbeiten von Petras (2005), die eine Erweiterung mit vier Termen als ausreichend beschrieb (der sogenannte Search Term Recommender, s. Abschnitt 7.1.2).

Im Sinne der Theorie der Polyrepräsentation sollen die zusätzlichen Suchbegriffe in verschiedenen Konstellationen gesucht werden. Für das folgende Beispiel werden die Felder Title (TI), Abstract (AB), kontrollierte Thesaurusterme (CT) und Autorennamen (AU) verwendet. Generell werden im weiteren Verlauf der Arbeit aber auch andere Felder genutzt. Die hier gewählte Konstellation dient nur als Beispiel für die generelle Herangehensweise. Zur einfacheren Lesbarkeit werden für die folgende Beispielanfrage nur jeweils zwei Begriffe/Entitäten verwendet.

Die Beispielanfrage nach „Gesundheit im Ruhestand“ würde wie folgt in eine einfache Suche in Titel und Abstract überführt werden:

```
TI/AB = (Gesundheit OR Ruhestand)
```

Die erweiterte Anfrage würde nun folgendermaßen aussehen (zusammengesetzte Namen und Begriffe sind in Anführungszeichen gesetzt):

```
TI/AB = (Gesundheit OR Ruhestand OR Sozialpolitik OR "Ältere Menschen")
```

```
OR CT = (Gesundheit OR Ruhestand OR Sozialpolitik OR "Ältere Menschen")
```

```
OR AU = ("Hauser, Richard" OR "Bäcker, Gerhard")
```

Die erweiterten Thesaurusterme wurden an die ursprünglich Anfrage angehängt und die aus der Kookkurrenzanalyse extrahierten Autorennamen sind in der letzten Zeile enthalten.

### 7.1.2 Pretest: Anfragerweiterung mit Thesaurustermen

Eine bekannte Variante, um Anfragerweiterungen mit kontrollierten Thesaurustermen durchzuführen, ist der sogenannte Search Term Recommender (STR). Inspiriert durch Arbeiten von Petras (2006) soll dieser dem Sprachproblem des IR entgegenwirken (s. Abschnitt 3.1). In Arbeiten von Schaer u. a. (2010b) wurde ein STR auf Grundlage der kommerziellen Klassifikationssoftware Reconnind Mindserver (inzwischen Decisiv Categorizer<sup>1</sup>) implementiert. Hierzu wurden ca. 370 000 Dokumente der sozialwissenschaftlichen Datenbank SOLIS als Trainingsmaterial verwendet. Der Mindserver konnte auf die Titel und Abstracts sowie die intellektuell vergebenen Schlagworte aus dem Thesaurus Sozialwissenschaften zugreifen. Die Texte aus Titel und Abstract bildeten die Menge der freien Terme und die Thesaurusterme bildeten die kontrollierten Terme.

<sup>1</sup>[http://www.reconnind.de/produkte/decisiv\\_categorization](http://www.reconnind.de/produkte/decisiv_categorization)



run	R	RP	P@5	P@10	P@15	P@20	P@30	P@100
B	0,206	0,177	0,574	0,492	0,445	0,414	0,363	0,198
B+MS	0,252	0,206	0,456	0,427	0,402	0,381	0,363	0,240
B+J	0,263	0,205	0,520	0,460	0,437	0,425	0,399	0,251

Abbildung 7.1: Vergleich zweier Implementierungen der Anfragerweiterung mittels eines STR: Baselinesystem B mit Erweiterungen aus der proprietären Klassifikationssoftware Mindserver (B+MS) und eine Implementation auf Basis des Jaccard-Index (B+J).

Aufgabe der Mindservers war es, die Beziehungen zwischen den freien und kontrollierten Termen zu erlernen.

Der ursprüngliche Einsatzzweck war, neue Dokumente, die zu SOLIS hinzugefügt werden, automatisch mit Schlagworten aus dem Thesaurus Sozialwissenschaften zu versehen. Die Funktionalität des Mindservers wurde für den STR ausgenutzt, indem ihm anstelle von Dokumenten bestehend aus Titel und Abstract, Pseudodokumente bestehend aus den Anfragetermen einer Suche zur Klassifikation vorgelegt wurden. Die Software verschlagwortete die Pseudodokumente genauso wie normale Dokumente und lieferte zu einer beliebigen (kurzen) Eingabe von Anfragebegriffen, passende Thesaurusterme. Diese konnten zur Anfragerweiterung genutzt werden. Dieser Ansatz wurde sehr erfolgreich in bspw. Schaer u. a. (2010b) oder Mutschke u. a. (2011) evaluiert. Allerdings ist dieses Verfahren von proprietärer und nicht quelloffener Software abhängig.

Von Schaer u. a. (2012b) wurde eine alternative Implementation des STR entwickelt, die im Gegensatz zur ersten Implementation ohne die proprietären Programmbestandteile auskommt. Sie basiert intern auf dem in Abschnitt 6.2.1 beschriebenen Jaccard-Index. Durch die neugewonnene Möglichkeit, nun den Quellcode selbstständig verändern und optimieren zu können, konnte die Trainingszeit für den STR von ehemals ca. 3 Tagen (Mindserver) auf nur noch ca. 3–4 Stunden reduziert werden. Gleichzeitig, wie der Pretest in Abbildung 7.1 zeigt, sind die Ergebnisse des auf dem Jaccard-Index basierenden STR (B+J) fast immer besser, als die der Mindserver-Variante (B+MS). Dies gilt sowohl für den Recall, als auch für die Cut-off-Precision an den Stellen P@5, P@10, P@15, P@20, P@30 und P@100. Einzig die R-Precision des Mindservers ist mit 0,206 minimal besser als die der Jaccard-Variante mit 0,205. Im Pretest wurden die GIRT-Testkollektion sowie die CLEF-Topics 76–175 verwendet. Der genaue Versuchsaufbau und die verwendeten Testkollektionen werden in Kapitel 8 beschrieben.

### 7.1.3 Pretest: Anfragerweiterung mit Autorennamen

Eine erste Evaluation des Ansatzes wird ausführlich in Schaer u. a. (2012c) beschrieben. Das dort beschriebene Experiment soll im Folgenden als Pretest für die Effektivität des Verfahrens stehen. In diesem Pretest werden vier verschiedene Anfragevarianten gegeneinander evaluiert, die zusammen mit einem regulären TF\*IDF-Ranking in der Solr-Umgebung (s. Kap. 5) implementiert wurden. Wie bereits im vorherigen Abschnitt

run	MAP	RP	P@10	P@20	P@100
B	0,139	0,182	0,442	0,353	0,172
B+TE	0,153	0,229*	0,430	0,399	0,217*
B+AE	0,122*	0,184	0,400*	0,350	0,175
B+TE+AE	0,170*	0,239*	0,478	0,427*	0,218*

Tabelle 7.1: Retrievalergebnisse drei unterschiedlicher Anfrageerweiterungen: Das Baselinesystem B, ohne jegliche Anfrageerweiterung; die Baseline erweitert um Thesaurusterme (B+TE); die Baseline erweitert mit Autorennamen (B+AE) sowie eine Kombination aller drei (B+TA+TE).

wurde für den Pretest auf die GIRT-Testkollektion zurückgegriffen, bei der die CLEF-Topics 76–125 verwendet wurden, um die Effektivität des zuvor beschriebenen Verfahren zu demonstrieren.

Zunächst wurde eine Baseline (B) implementiert, die aus einer unverarbeiteten Anfrage, die aus den Topiciteln generiert wurde, besteht. Hierzu wurden lediglich bekannte Stopwörter aus den Topiciteln entfernt und das Stemming (Porter Stemmer) der Suchmaschine Solr verwendet. Diese Baseline ist der Referenzwert für alle anderen Implementationen dieses Pretests. Als Expansionsvarianten wurden sowohl eine Erweiterung mit kookkurrierenden Thesaurustermen des Thesaurus Sozialwissenschaften (B+TE) und eine Erweiterung mit hoch-assoziierten Autorennamen (B+AE) implementiert. Als vierte Variante wurde eine Kombination aller drei Varianten (B+TE+AE) umgesetzt. Hierbei wurden der ursprünglichen Anfrage B noch Thesaurusterme (TE) und Autorennamen (AE) hinzugefügt.

Für den Pretest wurden die Evaluationsmaße MAP, R-Precision, P@10, P@20 und P@100 berechnet (s. Kap. 5). Die Ergebnisse sind in Tabelle 7.1 zusammengefasst. Der aus der Literatur bekannte Ansatz der Erweiterung mittels Thesaurustermen (B+TE) führte zu einer deutlichen, jedoch statistisch nicht signifikanten Steigerung des MAP-Wertes (+9 %) und eine signifikante Verbesserung in der R-Precision und P@100. B+AE ist gemessen an MAP (-12 %) und P@10, schlechter als die Baseline, wohingegen die R-Precision, P@20 und P@100 zur Baseline vergleichbare Werte liefern. Das kombinierte Verfahren kann in allen gemessenen Werten die besten Ergebnisse erzielen. Statistisch signifikante Veränderungen gegenüber der Baseline sind mit \* markiert. Die Signifikanzprüfung wurde mittels eines t-Tests mit  $\alpha = 0,05$  durchgeführt.

Zusammenfassend kann gesagt werden, dass die Erweiterung mit Autorennamen zu einer Verschlechterung gegenüber der Baseline führt, einem Query-Drift. In der Kombination aller drei Varianten können aber signifikante Verbesserungen sowohl bei MAP, R-Precision, P@20 und P@100 gemessen werden. Wählt man anstelle der einfachen Baseline die bereits erweiterte Baseline B-TE, kann der MAP-Wert der kombinierten Variante immer noch eine signifikante Verbesserung verzeichnen (+11 %). Diese positiven Ergebnisse, insbesondere bzgl. der Kombination der Varianten, lässt eine weitere Untersuchung der Effekte einer Anfrageerweiterung mit nicht-textuellen Elementen aussichtsreich erscheinen.

## 7.2 Alternatives Ranking – Reranking

Die Grundidee des hier vorgestellten Verfahrens basiert darauf, die Ordnung einer ursprünglich nach Relevanz sortierten Ergebnisliste nachträglich zu verändern. Hierzu wird ein Verfahren implementiert, das ein alternatives, nicht-textuelles Relevanzkriterium nutzt. Die nachträgliche Veränderung eines vorherigen Rankings wird als Reranking bezeichnet. Wie in den vorherigen Kapiteln beschrieben, werden von Benutzern wissenschaftlicher digitaler Bibliotheken unterschiedliche Kriterien zur Bewertung der Relevanz eingesetzt. In Bezug auf Nutzerstudien in digitalen Bibliotheken (s. Abschnitt 2.5) sind dies z. B. die Aktualität eines Dokuments, das Renommee einer Fachzeitschrift oder die Produktivität eines Autors.

### 7.2.1 Ausgangslage: TF\*IDF-Implementation der Suchmaschine Solr

In der praktischen Umsetzung wird hierbei auf die Implementierung des Vektorraummodells und der TF\*IDF-Gewichtung der freien Suchmaschine Solr zurückgegriffen, die auf der Programmbibliothek Lucene basiert (Bialecki u. a., 2012). Der tatsächliche Rankingalgorithmus von Solr und Lucene ist eine Variante des erweiterten booleschen Retrievals und des Vektorraummodells. Die Gewichtungsformel 7.1 unterscheidet sich dabei nur gering von den ursprünglichen Formeln für die TD\*IDF-Berechnung. Die Abkürzungen  $d$  für Dokument,  $q$  für Anfrage und  $t$  für Anfrageterm sind äquivalent zu der Formel 4.7 auf Seite 60:

$$\text{score}_{\text{solr}}(q,d) = \text{coord}(q,d) \cdot \text{qNorm}(q) \cdot \sum_{t \in q} (\text{tf}(t,d) \cdot \text{idf}(t) \cdot \text{boost}(t) \cdot \text{iNorm}(t,d)). \quad (7.1)$$

Die Unterschiede liegen in den folgenden Faktoren, die die Rankingformel beeinflussen (in Anlehnung an die offizielle Lucene-JavaDoc-API<sup>2</sup> und McCandless u. Hatcher, 2010):

**coord(q,d)** ist ein Gewichtungsfaktor, der abhängig ist von der Anzahl der Anfrageterme  $t \in q$ , die im jeweiligen Dokument  $d$  vorkommen. Dokumente, die mehrere der Anfrageterme enthalten, erhalten so einen höheren Wert. Bei  $\text{coord}(q,d)$  handelt es sich um einen sogenannten „search time factor“, der zur Laufzeit ausgewertet wird.

**qNorm(q)** ist ein Normalisierungsfaktor, der angewendet wird um die jeweiligen Werte für  $\text{score}_{\text{solr}}(q,d)$  vergleichbar zu machen. Dieser Faktor greift selbst nicht in den Rankingprozess ein, da alle gerankten Dokumente mit dem gleichen Faktor multipliziert werden. Allerdings sorgt er dafür, dass Werte für unterschiedliche Anfragen  $q$  vergleichbar sind. Es handelt sich auch hierbei um einen Faktor, der zur Laufzeit berechnet wird. Normalerweise wird der Faktor wie folgt berechnet:

$$\text{qNorm}(q) = \frac{1}{\sum_{t \in q} (\text{idf}(t) \cdot \text{boost}(t))^2}. \quad (7.2)$$

<sup>2</sup>[http://lucene.apache.org/core/3\\_6\\_0/api/core/org/apache/lucene/search/Similarity.html](http://lucene.apache.org/core/3_6_0/api/core/org/apache/lucene/search/Similarity.html)

**boost(t)** ist ebenfalls ein Faktor, der zur Laufzeit ermittelt wird. Es kann für jeden Anfrageterm  $t \in q$  ein eigener  $\text{boost}(t)$ -Wert angegeben werden.

**iNorm(t,d)** bündelt ein Reihe von Gewichtungsfaktoren, die zur Indexierungszeit ausgewertet werden:

- Dokumentabhängige Faktoren ( $\text{boost}(d)$ ) können ein spezielles Dokument in der Wertung bevorzugen oder benachteiligen, wobei
- Feldabhängige Faktoren ( $\text{boost}(f)$ ) dies für einzelne Felder tun können.
- Längenabhängige Faktoren ( $\text{lengthNorm}$ ) berechnen zum Zeitpunkt der Indexierung die Länge von Feldinhalten, sodass z. B. Einträge in einem längeren Titel tendenziell weniger gewichtet werden, als in einem sehr kurzen.

All die o. g. Faktoren werden letztlich zu einer einzigen Fließkommazahl zusammengerechnet und mit in den Index abgelegt. Unter der Voraussetzung, dass ein Dokument aus mehreren Feldern besteht, wird  $\text{iNorm}(t,d)$  wie folgt berechnet:

$$\text{iNorm}(t,d) = \text{boost}(d) \cdot \text{lengthNorm} \cdot \prod_{f \in d_t} \text{boost}(f). \quad (7.3)$$

Zusammengefasst bedeutet dies, dass bei der Berechnung des tatsächlichen Wertes für  $\text{score}_{\text{solr}}(q,d)$  eine Vielzahl von Faktoren zusammenwirken, die sowohl zur Anfrage- bzw. Laufzeit als auch zum Zeitpunkt der Indexierung berechnet werden.

Das in dieser Arbeit verwendete alternative Rankingverfahren setzt an den vorhandenen Schnittstellen der Rankingformel 7.1 an, indem gezielt die Gewichtungsfaktoren, insbesondere der laufzeitabhängige Faktor  $\text{boost}(t)$ , beeinflusst werden. In der *Lucene/Solr Query Language* (Smiley u. Pugh, 2009), mit derer sich Suchanfragen und die Gewichtungsfaktoren bei der Anfrageformulierung beeinflussen lassen, werden die Gewichtungen wie folgt umgesetzt:

$$(\text{Sport}^2 \text{ OR } \text{Jugendarbeit}^{0.5})$$

Zu sehen ist die Anwendung des booleschen Operators **OR** sowie der Einsatz des laufzeitabhängigen Boost-Operators  $\wedge$ . Im Beispiel werden die beiden Anfrageterme **Sport** und **Jugendarbeit** disjunktiv verknüpft, wobei allerdings der Begriff **Sport** doppelt gewichtet ( $\wedge^2$ ) und der Begriff **Jugendarbeit** nur halb gewichtet wird ( $\wedge^{0.5}$ ). Liegen keine weiteren Gewichtungsfaktoren vor, die zum Anfrage- oder Indexierungszeitpunkt Einfluss genommen haben, ergäbe sich für die vereinfachte TF\*IDF-Formel 7.4 und die drei Dokumente  $d_1$ ,  $d_2$  und  $d_3$  das Ranking, wie es in Tabelle 7.2 dargestellt wird.

$$\text{score}_{\text{simple}}(q,d) = \sum_{t \in q} (\text{tf}(t,d) \cdot (1 + \frac{N}{\text{df}(t)}) \cdot \text{boost}(t)). \quad (7.4)$$

Obwohl  $d_3$  einen tf-Wert von 0 hat, weil das Wort **Jugendarbeit** gar nicht darin vorkommt, liegt der errechnete Score doch bei 13,32. In dem fiktiven Beispiel würde dies trotz allem für den höchsten Rankingwert ausreichen.

	tf(S)	tf(J)	$N$	df(S)	df(J)	TF*IDF	score
$d_1$	1	6	100	6	10	$(1 \cdot 2,22 \cdot 2) + (6 \cdot 2 \cdot 0,5)$	10,44
$d_2$	2	1	100	6	10	$(2 \cdot 2,22 \cdot 2) + (1 \cdot 2 \cdot 0,5)$	10,88
$d_3$	3	0	100	6	10	$(3 \cdot 2,22 \cdot 2) + (0 \cdot 2 \cdot 0,5)$	13,32

Tabelle 7.2: Beispielberechnung für TF\*IDF für die Anfrage nach Sport OR Jugendarbeit, abgekürzt mit S (Sport) und J (Jugendarbeit).

### 7.2.2 Erweiterung des Standardrankings

In der klassischen Version des Bradfordizing nach White (1981) wird ein sogenanntes zonenbasiertes Reranking vorgeschlagen, das erstmalig von Mayr (2010) evaluiert wurde. Beide Arbeiten gehen davon aus, dass die bibliometrische Gesetzmäßigkeit der Streuung von Literatur (Bradford's Law of Scattering, s. Kap. 6) genutzt werden kann, um das Retrieval zu beeinflussen. Bradfordizing bezeichnet dabei das Reranking einer Ergebnismenge auf Grundlage von hohen Publikationsfrequenzen sogenannter Kernzeitschriften. Artikel dieser Kernzeitschriften werden im Ranking bevorzugt und denen, die aus peripheren Zeitschriften stammen, vorangestellt. Hierzu werden verschiedene Zonen definiert, die alle eine ungefähr gleich große Anzahl an Dokumenten beinhalten. Die innerste Zone ist der Kern, die äußeren Zonen die Peripherie bilden.

Ein Nachteil des zonenbasierten Ansatzes ist, dass er innerhalb der Zonen kein besonderes Ranking vorsieht. Theoretisch werden alle Dokumente des Kerns gleich stark gewichtet. Dies ist zwar eine Verbesserung gegenüber dem booleschen Retrieval, aber kann nicht mit der Flexibilität eines Relevanz-Rankings konkurrieren, das kein (grobes) zonenbasiertes Ranking von Blöcken, sondern ein (feineres) dokumentenbasiertes Ranking ermöglicht.

#### Zonenloses Reranking auf Basis von Bradfordizing

Um diesen Nachteil auszugleichen, wird eine zonenlose Variante (Schaer, 2011) vorgeschlagen, die eine Kombination von bewährten TF\*IDF-Rankingfaktoren und einem durch das Bradfordizing inspirierten Gewichtungsfaktor darstellt. Die zuvor vorgestellte Standardtechnologie Solr soll gezielt ausgenutzt und erweitert werden, um dieses Rankingverfahren umzusetzen und zu evaluieren. Mit Hilfe der Gewichtungsfaktoren wird ein alternatives Ranking etabliert.

Angewendet auf das grundlegende Solr-Ranking wird die Rankingformel 7.1 am Faktor  $boost(t)$  gezielt verändert und die Anfrage mittels des folgenden Mechanismus erweitert:

- Zunächst wird auf Basis einer Anfrage  $q$  eine initiale (unsortierte) Dokumentmenge  $D_q \subseteq D$  zusammengestellt, wobei  $D_q$  die Menge der zu  $q$  passenden Dokumente ist.  $D$  ist die Gesamtmenge der Dokumente.
- Anschließend werden aus  $D_q$  die Anzahl der Dokumente extrahiert, die einer bestimmten Entität  $e$  zugehörig sind, z. B. einem bestimmten Autor, einer Zeitschrift

o.ä., und der Häufigkeit ihres Auftretens nach sortiert. Dies wird als *ef* bezeichnet, die Frequenz der jeweiligen Entität (in Anlehnung an die Termfrequenz *tf*).

- Aus der Häufigkeit wird nun der Gewichtungsfaktor ermittelt, der für jede einzelne Entität mittels einer Umformulierung der ursprünglichen Anfrage umgesetzt wird.

Für die Beispielanfrage nach **Sport OR Jugendarbeit** ( $|D_q| = 1\,728$ ) ergeben sich die Häufigkeitswerte (in diesem Fall der Autorennamen, gekürzt auf die zehn häufigsten Autorennamen) wie sie in Tabelle 7.3 aufgelistet sind. Insgesamt gibt es 1\,594 eindeutige Autorennamen, die die Menge der Entitäten  $E$  innerhalb der Menge der Dokumente  $D_q$  formen. Für jedes Element  $e \in E$  wird die Anfrage  $q$  wie folgt erweitert (in Pseudocode):

---

**Listing 7.1** Pseudocode zur Anfrageumformulierung

---

```
var type = "author"
var q = "Sport OR Jugendarbeit"
var q_new = ""

foreach e in E {
  var boost = freq(e)
  q_new += "($q AND $type:$e^$boost)"
  if (e != E.last()) {q_new += " OR "}
}
```

---

Angewendet auf die Beispielanfrage nach **Sport OR Jugendarbeit** ergibt sich so die folgende umformulierte Anfrage in der entsprechenden Solr-Syntax:

---

**Listing 7.2** Umformulierte Anfrage in Solr-Syntax nach dem Verfahren aus Listing 7.1

---

```
(( "Sport OR Jugendarbeit" ) AND author:"Feuchthofen, Jörg E."^11) OR
(( "Sport OR Jugendarbeit" ) AND author:"Konegen-Grenier, Christiane"^10) OR
(( "Sport OR Jugendarbeit" ) AND author:"Buttler, Friedrich"^8) OR
(( "Sport OR Jugendarbeit" ) AND author:"Bultmann, Torsten"^7) OR
(( "Sport OR Jugendarbeit" ) AND author:"Herrmann, Helga"^7) OR
(( "Sport OR Jugendarbeit" ) AND author:"Manning, Sabine"^7) OR
(( "Sport OR Jugendarbeit" ) AND author:"Geißler, Karlheinz A."^6) OR
(( "Sport OR Jugendarbeit" ) AND author:"Knoll, Joachim H."^6) OR
(( "Sport OR Jugendarbeit" ) AND author:"Lenhardt, Gero"^6) OR
(( "Sport OR Jugendarbeit" ) AND author:"Lenske, Werner"^6) OR
...
(( "Sport OR Jugendarbeit" ) AND author:"Zühlsdorf, Petra"^1)
```

---

Es werden alle Dokumente, die sowohl die beiden Suchterme **Sport** und **Jugendarbeit** als auch zum Autoren **Feuchthofen, Jörg E.** gehören, um den Faktor 11 höher gewichtet, die Dokumente der Autorin **Konegen-Grenier, Cristiane** um den Faktor 10 usw. Im Detail kann der Gewichtungsfaktor unterschiedlich ausgeprägt sein und trägt so zu einem anderen Verhalten des Rankings bei, wie in den folgenden Abschnitten beschrieben wird.

Position	Autorennamen	Häufigkeit in $D_q$	$boost_2$	$boost_4$
1	Feuchthofen, Jörg E.	11	121	14 641
2	Konegen-Grenier, Christiane	10	100	10 000
3	Buttler, Friedrich	8	64	4 096
4	Bultmann, Torsten	7	49	2 401
5	Herrmann, Helga	7	49	2 401
6	Manning, Sabine	7	49	2 401
7	Geißler, Karlheinz A.	6	36	1 296
8	Knoll, Joachim H.	6	36	1 296
9	Lenhardt, Gero	6	36	1 296
10	Lenske, Werner	6	36	1 296
...	...	...	...	...
1 594	Zühlsdorf, Petra	1	1	1
	$e \in E$	$ef$	$ef^2$	$ef^4$

Tabelle 7.3: Berechnung der Gewichtungsfaktoren für das zonenlose Bradfordizing basierend auf der Häufigkeit der Entitätsfrequenz  $ef$  und deren Potenzierungen mit den Exponenten 2 und 4.

### Mischzonen Reranking

Zentrales Element des o. g. Verfahrens ist die Zeile 6 des Listings 7.1, also der Gewichtungsfaktor `boost`, der auf eine beliebige Entität, die mit der ursprünglichen Anfrage gleichzeitig auftritt, angewendet wird. Der Gewichtungsfaktor kann durch beliebige Operationen verändert werden. Es hat sich allerdings in der konkreten Anwendung gezeigt, dass eine Potenzierung des Häufigkeitswertes von  $e \in E$  am ehesten den für dieses Experiment gewünschten Effekt erzielt. Durch die Wahl des Exponenten kann ein unterschiedliches Verhalten des Rankings erwirkt werden. In Pretests zeigt sich, dass die Exponenten 1, 2 und 4 jeweils zu einem anderen erklärbaren Verhalten führen.

Das folgende Listing 7.3 zeigt die Einbindung der drei vorgeschlagenen Exponenten, die jeweils den Faktor `boost` aus der Potenzierung (`pow`) der Basis (`freq(e)`) mit einem der drei Exponenten (`exp`) umsetzen.

Die erste Variante einer Potenzierung mit dem Exponenten 1 stellt die defensivste Variante dar, da hierbei keine besondere Gewichtung auf `boost` einwirkt. Für die komplette Rankingformel heißt dies, dass die jeweilige Frequenz der Entitäten alleine den Gewichtungsfaktor bestimmt und dieser Einfluss auf das Ranking im Zweifel recht gering ist.

Um auf der anderen Seite die einzelnen Entitäten voneinander abzugrenzen und ein eher klassisches, zonenbasiertes (und daher trennendes) Reranking zu ermöglichen, wird ein möglichst großer Gewichtungsfaktor gewählt, der dafür sorgt, dass ausreichend Abstand den endgültigen Rankingwerten (`score`) existiert. Experimentell hat sich hier die Potenzierung mittels des Exponenten 4 als solch ein Wert herausgestellt.

**Listing 7.3** Pseudocode zur Anfrageformulierung mit wechselnden Gewichten

```

...
arr exponentes = [1,2,4]

foreach exp in exponentes {
  foreach e in E {
    var boost = pow(freq(e),exp)
    q_new += "($q AND $type:$e^$boost)"
    if (e != E.last()) {q_new += " OR "}
  }
}

```

Position	Autorennamen	Häufigkeit in $D_q$	Gesamtanzahl	$boost_{EF*IDF}$
1	Feuchthofen, Jörg E.	11	14	43,83
2	Konegen-Grenier, Christiane	10	14	39,85
3	Bultmann, Torsten	7	9	29,24
4	Herrmann, Helga	7	15	27,68
5	Manning, Sabine	7	15	27,68
6	Buttler, Friedrich	8	80	25,82
7	Lenske, Werner	6	10	24,79
8	Knoll, Joachim H.	6	16	23,56
9	Lenhardt, Gero	6	20	22,98
10	Geißler, Karlheinz A.	6	23	22,62
...	...	...	...	...
1594	Zühlsdorf, Petra	1	1	5,13
	$e \in E$	ef	df	$ef \cdot \frac{N}{df}$

Tabelle 7.4: Beispielberechnung des EF\*IDF-Gewichtungsfaktors für die Anfrage **Sport OR Jugendarbeit**.

Eine Mischform der beiden Ansätze bildet die Potenzierung mit dem Exponenten 2. Hierbei hat sich experimentell eine gute Durchmischung der term-basierten TF\*IDF-Rankingfaktoren und der Entitätenfrequenzen gezeigt.

### EF\*IDF-Reranking

Während in den vorherigen Varianten nur die Häufigkeit des Auftretens der Entitäten in der initialen Dokumentmenge  $D_q$  für den Gewichtungsfaktor von Bedeutung ist, kann das Verfahren auch für komplexere Gewichtungsverfahren angewendet werden. Ein reines Anwenden der Häufigkeitswerte beim Bradfordizing entspricht in der Grundüberlegung einem Ranking allein auf Grundlage der Termfrequenz (TF), wohingegen mächtigere Verfahren wie TF\*IDF eher in Modellen wie den Pennant-Diagrammen (s. Abschnitt



4.2) zur Anwendung kommen. Angewendet auf das Berechnungsverfahren der Gewichtungsfaktoren wird (1) die Entitätenfrequenz (ef) für die Häufigkeit des Auftretens von Entitäten, die in der zu der Suchanfrage  $q$  passenden Dokumentmenge  $D_q$  enthalten sind, (2) sowie die bekannte Dokumentfrequenz (df), die die Anzahl der Dokumente in der gesamten Dokumentmenge  $D$ , die die Entität  $e$  enthält, benutzt:

$$ef(e, D_q) = \text{freq}(e, D_q), \quad (7.5)$$

$$df(e, D) = \text{freq}(e, D), \quad (7.6)$$

$$EF * IDF = ef \cdot \frac{N}{df}. \quad (7.7)$$

Der Gewichtungsfaktor ist nicht nur abhängig von der Häufigkeit des gleichzeitigen Auftretens der Entität, die zum Ranken verwendet werden soll, in der aktuellen Dokumentmenge  $D_q$ , sondern auch von deren Auftreten in der Gesamtmenge  $D$ . In der Tabelle 7.4 sieht man, wie der Gewichtungsfaktor in Abhängigkeit von ef und df berechnet wird. Es zeigt sich ein eindeutig anderes Ranking.

### 7.2.3 Pretest: Vergleich von zonenbasierten und -losem Bradfordizing

Die Tabelle 7.5a „zeigt die gemittelten Precision-Werte für 125 CLEF-Topics“ (Mayr, 2009b, S. 156) des zonenbasierten Bradfordizing. „Im Durchschnitt beträgt die Precision in der Kernzone 0,28. Die Precision sinkt in den folgenden Zonen deutlich ab (Zone 2 = 0,23 und Zone 3 = 0,19). Die baseline-Precision fasst die durchschnittliche Precision über alle Zeitschriftenartikel zusammen“ (Mayr, 2009b, S. 156 f.). Da in der Arbeit von Mayr nicht mit Evaluationsmaßen wie P@10 oder MAP gearbeitet wurde, ist ein direkter Vergleich der zonenbasierten und zonenlosen Ansätze nicht eins zu eins möglich. Allerdings kann grob folgender Vergleich gelten:

- $P_{core}$  entspricht ungefähr dem Durchschnitt von P@30 und P@100, und
- $P_{Z2}$  ist ungefähr durch P@100 angenähert.

Dies ist durch die durchschnittliche Größe der einzelnen Topics begründet, die bei 145 Artikeln lag (Mayr, 2009b, S. 142). Folglich hatte jede Zone eine Größe von ca. 50 Artikeln.

An den genannten Schnittpunkten sieht man eine weitgehende Übereinstimmung der Precisionwerte (z. B. 2004:  $P_{Z2}=0,185$  und  $P@100=0,178$ ). In den einschlägigen Cut-off-Precisionwerten wie P@5 und P@10 sieht man allerdings auch einen wesentlich höheren Wert als im rein zonenbasierten Modell. Im Durchschnitt liegt die kontinuierliche Rankingversion sogar leicht vor dem zonenbasierten Ansatz, was aber der nicht genauen Vergleichbarkeit und den unterschiedlichen Durchschnittsberechnungen geschuldet sein mag.

CLEF-Jahrgang	Topics	$P_{core}$	$P_{Z2}$	$P_{Z3}$	$P_{baseline}$
2003	25	0,294	0,218	0,157	0,221
2004	25	0,226	0,185	0,134	0,179
2005	25	0,310	0,240	0,174	0,239
2006	25	0,288	0,267	0,244	0,265
2007	25	0,278	0,256	0,217	0,248
$\emptyset$	25	0,279	0,233	0,185	0,230

(a) Durchschnittliche Precision für Zeitschriftenartikel in den Bradford-Zonen (Core, Z2 und Z3) und Baseline-Precision im CLEF-Datensatz (N=125 Topics) – Zonenbasiertes Bradfordizing entnommen aus Mayr (2009b, S. 157)

CLEF-Jahrgang	Topics	P@5	P@10	P@15	P@20	P@30	P@100	P@200
2003	25	0,424	0,384	0,349	0,338	0,311	0,253	0,157
2004	25	0,312	0,256	0,221	0,212	0,223	0,178	0,098
2005	25	0,384	0,368	0,357	0,350	0,333	0,246	0,156
2006	25	0,360	0,288	0,269	0,298	0,300	0,271	0,185
2007	25	0,320	0,288	0,285	0,290	0,279	0,265	0,191
$\emptyset$	25	0,355	0,313	0,294	0,294	0,282	0,233	0,154

(b) Vergleichsergebnisse des Rerankingverfahrens aus Listing 7.1

Tabelle 7.5: Vergleich des zonenbasierten Bradfordizing nach Mayr (2009b, S. 157) und des zonenlosen Modells aus Listing 7.1.

### 7.3 Zusammenfassung

Im vorangegangenen Abschnitt wurden zwei Verfahren zur Verbesserung des Retrievals, sogenannte Mehrwertdienste, vorgestellt. Die Verfahren ermöglichen es auf Grundlage von textuellen und nicht-textuellen Merkmalen, die gleichzeitig mit den Suchtermen auftreten, ein alternatives Ranking oder eine Anfrageerweiterung durchzuführen. Beide Verfahren konnten in Pretests ihre generelle Funktionsfähigkeit zeigen. Für eine weitergehende Analyse werden im folgenden Kapitel die verwendeten Materialien für eine umfassende IR-Evaluation nach dem Cranfield-Paradigma vorgestellt.

## 8 Materialien

Im folgenden Kapitel wird die Umgebung für die Evaluation der in Kapitel 7 beschriebenen Verfahren zusammengefasst. Es werden die drei essentiellen Bestandteile von IR-Testkollektion, der Dokumentkorporus, die vordefinierten Topics und die Relevanzurteile anhand der beiden Testkollektionen GIRT und iSearch genauer vorgestellt.

### 8.1 IR-Testkollektionen

In dieser Arbeit werden zwei bekannte IR-Testkollektionen verwendet: der GIRT4-Korpus mit den in der CLEF-Initiative gesammelten Topics und Relevanzurteilen des *Domain-specific Tracks* sowie die iSearch-Testkollektion in der Version 1.0. Die Wahl fiel auf diese beiden Testkollektionen, da diese die folgenden Kriterien erfüllen:

- Die Testkollektion sollte Dokumente aus dem Umfeld der digitalen Bibliotheken enthalten.
- Die Testkollektionen sollten eine gewisse Mindestgröße ( $\geq 100\,000$  Dokumente) besitzen.
- Die Dokumente sollten einen fachwissenschaftlichen Bezug haben.
- Die Qualität der Dokumentenerschließung sollte einen gewissen Mindeststandard erfüllen (Titel, Abstract, Autoren), damit die in der Arbeit entwickelten Verfahren anwendbar sind.
- Die Testkollektionen sollten frei und kostenlos verfügbar sein, um eine Nachnutzung durch andere Forscher zu ermöglichen.

Zwar gibt es eine Reihe anderer Testkollektionen, wie die bekannten TREC-Datensätze (Disc 4 und 5) oder die INEX-Kollektionen, doch kann keine die oben genannten Kriterien erfüllen, aufgrund des Fehlens von wissenschaftlichen Dokumenten (TREC) oder zu geringer Größe (INEX in der Version 1.4 beinhaltet nur 12 107 Dokumente).

#### 8.1.1 Testkorpora

GIRT (*German Indexing and Retrieval Test database*) ist ein Evaluationskorporus, der beim ehemaligen Informationszentrum Sozialwissenschaften (IZ, inzwischen GESIS) entwickelt wurde. Kluck (2003) beschreibt die ursprüngliche Motivation zur Erstellung dieses Korpus wie folgt:

„General-purpose news documents require very different search criteria than those used for reference retrieval in databases of scientific literature items, and also offer no possibility for comparable test runs with domain-specific terminology. There was thus a strong demand for resources that would allow research into retrieval on German texts; groups wanted to test the insights accruing from the TREC experience and the capability of morphological components derived from English texts on German language material.“ (Kluck, 2003)

Die ersten Versionen von GIRT wurden in den frühen TREC-Kampagnen eingesetzt und mit der Version 4 zuerst in der CLEF-Kampagne 2003 für das Cross-Language-Retrieval verwendet.

SOLIS ist das sozialwissenschaftliche Informationssystem, das seit 1978 aufgebaut und entwickelt wird. Die Datenbank SOLIS wird von GESIS gemeinsam mit Kooperationspartnern wie dem „Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit in Nürnberg, den Herausgebern der Zeitschrift für Politikwissenschaft und dem Wissenschaftszentrum Berlin für Sozialforschung hergestellt“<sup>1</sup>. Die Inhalte der Datenbank sind neben Nachweisen zu Monographien und Sammelwerken auch die Fachartikel aus rund 300 Zeitschriften, die in unterschiedlichen Graden erfasst werden – je nach Einschlägigkeit für die fachliche Abdeckung der Datenbank. Hierbei sind vor allem die sozialwissenschaftlichen Fachgebiete Soziologie, Bildungsforschung, Politikwissenschaft, Historische Sozialforschung, Sozialpolitik, Demographie, Sozialpsychologie, Ethnologie, Methoden der Sozialwissenschaften, Arbeitsmarkt- und Berufsforschung und Kommunikationswissenschaften von Bedeutung. Die Liste der erfassten Zeitschriften kann online eingesehen werden und wird laufend erweitert<sup>2</sup>. Neben der alphabetischen Liste der Zeitschriftennamen enthält die Liste auch eine Relevanzbewertung für die Sozialwissenschaften (Zimmer u. Rehse, 2011).

iSearch (*integrated Search*) ist ein relativ neuer Evaluationskorpus, der erstmalig 2010 vorgestellt wurde (Lykke u. a., 2010). Wie auch GIRT enthält er wissenschaftliche Literaturnachweise, in diesem Fall aus der Fachdomäne der Physik und primär mit dem Ziel der Evaluation einer integrierten Suche über verschiedene Informationsquellen hinweg. Hierzu werden in iSearch sowohl Metadaten (Untermenge PN) und Volltexte (Untermenge PF) aus dem Preprint-Server arXiv als auch Bibliothekskataloginformationen (Untermenge BK) der dänischen Nationalbibliothek zusammengetragen. Da im Gegensatz zu SOLIS und GIRT keine Qualitätskontrolle und kontrollierte Erfassung der Daten durchgeführt wurde, sind sowohl quantitativ wie qualitativ Unterschiede sichtbar. Durch die Eingabe der Metadaten durch bibliografische Laien sind sowohl Unstimmigkeiten in der Erfassung der Autoren, Konferenz- und Zeitschriftennamen zu beobachten. Zusätzlich wird auf den Einsatz von kontrollierten Vokabularen und anderen Erschließungswerkzeugen verzichtet.

Der GIRT4-DE-Korpus wurde zusätzlich in einer erweiterten Variante GIRT4-DE-EXT eingesetzt. Hierzu wurden die Metadaten nach der Prozedur, wie in Abbildung

<sup>1</sup><http://www.gesis.org/unser-angebot/recherchieren/solis/>

<sup>2</sup><http://vt-www.gesis.org/pdf/Zeitschriften.pdf>

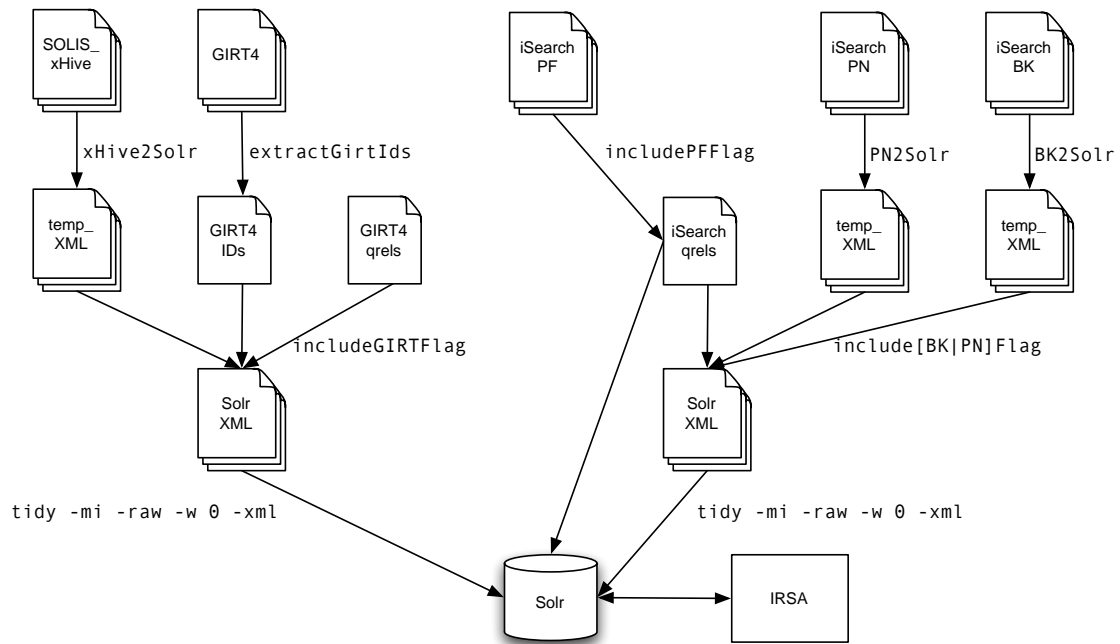


Abbildung 8.1: Überblick über den Transformationsprozess für die Überführung von GIRT4-DE und SOLIS in GIRT4-DE-EXT sowie der Aufbereitung der iSearch-Kollektion. Die Kantenbeschriftungen bezeichnen die verwendeten Konvertierungsskripte, die in Anhang D beschrieben werden.

8.1 skizziert, erweitert. Dies war möglich, da GIRT4 eine Untermenge von SOLIS ist. Eine Übersicht über die von SOLIS nach GIRT4-DE-EXT übertragenen Metadaten ist in Tabelle 8.1 zusammengefasst. Wenn in den späteren Kapiteln vom GIRT-Korpus die Rede ist, ist damit immer GIRT4-DE-EXT gemeint, wenn nicht gesondert ausgezeichnet. Die in der Tabelle 8.1 gegenübergestellten Merkmale der verschiedenen Korpora sind im Einzelnen:

**ID:** ein eindeutiger Identifier,

**AUTHOR:** Autorennamen,

**EDITOR:** Herausgebernamen,

**TITLE:** Titel des Dokuments,

**SOURCE:** Angaben zum Publikationsorgan, z. B. Konferenzbände oder Zeitschriften,

**ISSN:** ISSN für Publikationen in Zeitschriften,

**ISBN:** ISBN für Publikationen in Monografien oder Sammelbänden,

**PUBYEAR:** Publikationsjahr,

Korpus	ID	AUTHOR	EDITOR	TITLE	SOURCE	ISSN	ISBN	PUBYEAR	KEYWORDS	CLASSIFICATION	ABSTRACT	FULLTEXT	METHOD	LOCATION	PUBLISHER	PAGES
iSearch-PN	●	○	-	○	○	-	-	-	-	-	●	-	-	-	-	-
iSearch-BK	●	○	○	○	-	-	-	-	○	○	○	-	-	-	-	-
iSearch-PN	●	○	-	○	-	-	-	-	-	-	○	○	-	-	-	○
GIRT4-DE	●	●	○	○	○	-	-	●	●	●	●	-	●	-	-	-
GIRT4-DE-EXT	●	●	●	○	○	●	●	●	●	●	●	-	●	○	○	●
SOLIS	●	●	●	○	●	●	●	●	●	●	●	-	●	○	○	●

Tabelle 8.1: Featurematrix zu den verwendeten Korpora. Durch die drei Symbole -, ○ und ● werden die jeweiligen Eigenschaften der Korpora beschrieben: - = Eigenschaft nicht vorhanden ; ○ = nicht in einer standardisierten Form erfasst; ● = standardisiert erfasst.

**KEYWORDS:** freie oder kontrollierte Schlagwörter,

**CLASSIFICATION:** Klassifikationsinformationen,

**ABSTRACT:** Autorenreferate oder intellektuelle Abstracts,

**FULLTEXT:** Volltext des eigentlichen Dokuments,

**METHODS:** Angaben zur Forschungsmethode,

**LOCATION:** Ortsangaben z. B. zum Herausgeber oder Veranstaltungsort,

**PUBLISHER:** Verlagsangaben,

**PAGES:** Anzahl oder von-bis Seitenangaben.

Durch die drei Symbole -, ○ und ● werden die jeweiligen Eigenschaften der Korpora beschrieben, wobei - bedeutet, dass diese Eigenschaft nicht vorhanden ist bzw. sich diese nicht anwendbar ist; ○ beschreibt ein Vorhandensein der Information, allerdings nicht in einer standardisierten Form; bei der Angabe von ● liegt die Information standardisiert vor, z. B. in Form eines kontrollierten Vokabulars.

In Abbildung 8.2 ist die unterschiedliche zeitliche Abdeckung der beiden Testkollektionen zu sehen. Während für GIRT der Schwerpunkt der Dokumente in der Mitte der 1980er

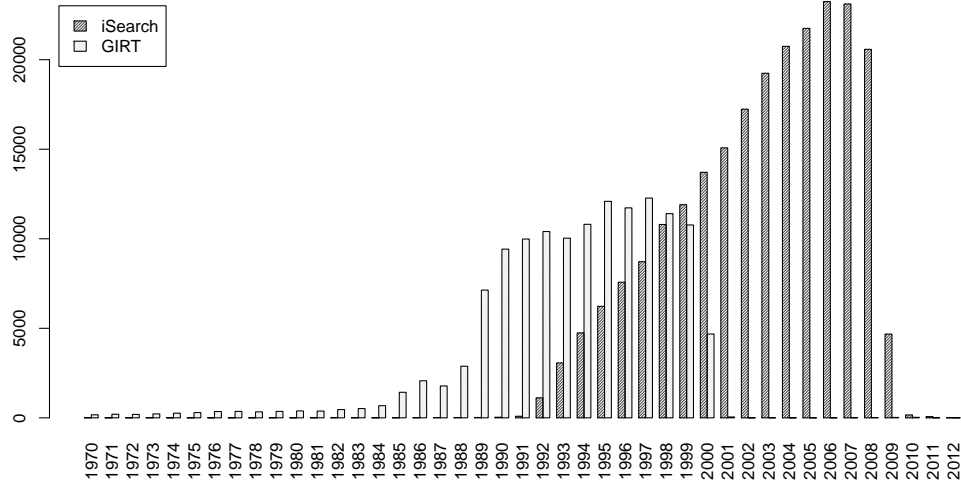


Abbildung 8.2: Anzahl der Dokumente pro Erscheinungsjahr für iSearch und GIRT.

Korpus	Dokumentart	Größe in kB	Dokumente	kB/Dok.
iSearch-PN	Literaturnachweise	365 568	291 245	1,25
iSearch-BK	Bibliothekskatalog	10 358	18 442	0,56
iSearch-PF	Volltexte	55 807 721	143 569	388,71
GIRT4-DE	Literaturnachweise und Forschungsprojektdaten	341 382	151 319	2,26
GIRT4-DE-EXT	Literaturnachweise	1 170 896	135 214	8,66
SOLIS	Literaturnachweise	2 844 559	377 540	7,53

Tabelle 8.2: Überblick über die im folgenden Experiment verwendeten Korpora. Alle Korpora bestehen aus einer Anzahl an bibliografischen Metadaten aus den beiden wissenschaftlichen Domänen Physik (iSearch) und den Sozialwissenschaften (GIRT und SOLIS).

Jahre bis ins Jahr 2000 reicht, beginnt für iSearch der stetige Zuwachs an Dokumenten ab Anfang der 1990er Jahre.

In Tabelle 8.2 sind die unterschiedlichen Kollektionen in ihrer Größe gegenübergestellt. Alle Korpora bis auf iSearch-PF enthalten ein intellektuell erstelltes Abstract. PF besteht ausschließlich aus PDF-Volltexten. Die Datensätze unterscheiden sich stark in der Größe, so ist ein durchschnittliches Dokument des erweiterten GIRT4-DE-EXT-Datensatzes nahezu viermal so groß wie der original GIRT4-DE-Datensatz. GIRT4-DE-EXT ist mit 135 214 Dokumenten kleiner als iSearch mit 453 256 Dokumenten. Allerdings wird klar, dass GIRT4-DE-EXT mit 8,66 kB pro Dokument die umfangreichsten Metadaten besitzt. Bei iSearch-PF sind insgesamt 55 GB an PDF-Volltexten enthalten, die im Durchschnitt 388,71 kB pro Dokument groß sind.

Es werden die Unterschiede in der Erschließungsqualität klar: GIRT ist im Vergleich zu iSearch wesentlich konsistenter und umfangreicher erschlossen. Es fällt auf, dass weniger und schlechter erschlossene Metadaten für iSearch vorliegen.

### 8.1.2 Topics

Topics sind Fragestellungen im Rahmen einer Evaluationskampagne, die von Retrievalsystemen bearbeitet werden sollen. Traditionell besteht ein Topic aus einem stichwortartigem Titel, ein bis zwei zusammenfassenden Sätzen zur Fragestellung und einer ausführlichen Beschreibung, die auf spezielle Gesichtspunkte oder Einschränkungen der Fragestellung eingeht. Erstmals wurde diese Art der Fragestellung in der Evaluationskampagne TREC eingesetzt, später dann von weiteren Kampagnen wie CLEF oder NTCIR aufgegriffen und erweitert.

Die Listings 8.1 und 8.2 zeigen zwei typische Fragestellungen aus den Korpora GIRT und iSearch. Während bei GIRT neben

**num:** einer laufenden Topicnummer,

**DE-title:** Titel,

**DE-desc:** eine zusammenfassende Beschreibung und

**DE-narr:** ausführliche Beschreibung

enthalten sind, ist das iSearch-Exzerpt ein Beispiel für eine umfänglichere Topicbeschreibung. Hier werden neben der

**topic\_id:** laufenden Topicnummer auch

**author\_id:** der Autor,

**current\_information\_need:** eine Beschreibung des Informationsbedürfnisses,

**work\_task:** die aktuelle Arbeitsaufgabe,

**background\_knowledge:** Hintergrundwissen des Fragestellers,



---

**Listing 8.1** Beschreibung eines Topics aus dem iSearch-Korpus mit Angabe einer laufenden Topicnummer, dem Autor der Topicbeschreibung, dem Informationsbedürfnis, der aktuellen Arbeitsaufgabe, dem Hintergrundwissen des Fragestellers, der idealen Antwort und einer Liste von möglichen Suchtermen.

---

```

<topic>
  <topic_id>015</topic_id>
  <author_id>090</author_id>
  <current_information_need>I am looking for examples of intracellular
    recordings on cells. Especially I am looking for recordings of the
    intracellular electrical potential of living neurons. I want to
    investigate which methods that previously have been used to study the
    intracellular environment of living cells. And again especially
    methods that has been used to record the intracellular electrical
    potential.</current_information_need>
  <work_task>In my Master project I am involved in a project where the
    ultimate goal is to design a device capable of measuring the
    intracellular potential of living cells in an acute brain slice. These
    potential will be recorded using nano electrodes, which is 3
    dimensional electrodes where the tip has a radius below 100 nm.
    Therefore it would be interesting to me to see which other methods
    that has been used to make measurements of the intracellular
    environment of the cells. These results will be compared to the
    results obtained with our new approach.</work_task>
  <background_knowledge>I know that the intracellular potential previously
    has been recorded by a method called patch clamp. In this method the
    cell is punctured and therefore the cell does not survive. Furthermore
    I know that researchers have penetrated the cell membrane with various
    types of nanowires and carbon nano tubes without killing the cell.
    However I do not know if any of these penetrating nanowires has been
    used to measure on the intracellular environment of the cells. The
    extracellular electrical potential of living firing neurons has been
    recorded by many researchers using micro electrodes, therefore it
    would be interesting to see if any had recorded the intracellular
    potential.</background_knowledge>
  <ideal_answer>The ideal answer would be an overview over the various
    recordings that have been used to measure on the intracellular
    environment. Also articles from which this overview can be made by me
    are interesting. In these articles it should be stated whether the
    cell survive the measurement, what was recorded and how well the
    recording performed. It is especially interesting to me whether
    recordings of the intracellular potential of living cells has been
    performed without killing the cell.</ideal_answer>
  <search_terms>Intracellular; Electrical potential; Electrodes;
    Neurons</search_terms>
</topic>

```

---

---

**Listing 8.2** Beschreibung eines Topics aus dem GIRT4-DE-Korpus mit Angabe einer laufenden Topicnummer, des Titels, einer Kurzbeschreibung und einer ausführlichen Beschreibung von Kriterien zur Relevanzbewertung.

---

```
<top>
  <num>99</num>
  <DE-title>Sport und Jugendarbeit</DE-title>
  <DE-desc>Finde Dokumente, die über den Erfolg und die Probleme der
    Einbindung von Jugendlichen in sportliche Aktivitäten als Beitrag zur
    Arbeit mit Jugendlichen berichten.</DE-desc>
  <DE-narr>Relevante Dokumente liefern Informationen zu Jugendlichen und
    Kindern im Zusammenhang mit sozialer Jugendarbeit und der Funktion des
    Sportes. Die soziale Arbeit mit jungen Menschen und deren Einbindung
    in regionale sportliche Aktivitäten als Erfolgskonzept der
    Jugendsozialarbeit ist hier von besonderem Interesse.</DE-narr>
</top>
```

---

**ideal\_answer:** die ideale Antwort aus Sicht des Fragestellers sowie

**search\_terms:** eine Liste von möglichen Suchtermen angegeben.

Allgemein wird in iSearch nicht von einem anonymen Fragesteller und einem neutralen Informationsbedürfnis ausgegangen, wie dies in TREC und CLEF der Fall war, sondern ein konkreter Benutzer und dessen Informationsbedürfnis. Zusätzlich wird der erweiterte Fragekontext in das Topic eingebaut. Dies erlaubt zwar auf der einen Seite eine sehr viel weitergehende Analyse des Kontextes und der Nutzerzentriertheit, erschwert aber auch die Nachnutzung der Topics durch andere Assessoren. Ein iSearch-Topic ist folglich auf einen speziellen Assessor ausgerichtet, wohingegen ein GIRT-Topic durch seine neutrale Beschreibung theoretisch auch von anderen Assessoren bewertet werden könnte. Die starke Zentrierung auf einzelne Nutzer der iSearch-Topics erklärt sich aber zentral durch den geplanten Einsatz in der Fachwelt des Task-based, interaktiven und kontextbezogenen Retrievals, z. B. der *Information Interaction in Context* (IiX)-Konferenz.

### 8.1.3 Relevanzurteile

Am Beispiel der CLEF-Evaluationskampagne aus dem Jahr 2005 stellen Kluck u. Winter (2006) die Motivation und Herangehensweise bei der Erstellung von Topics dar. Zusätzlich werden von ihnen aber auch die allgemeinen und besonderen Probleme bei der Topicerstellung und den anschließenden Relevanzbewertungen (den sogenannten Assessments) beschrieben: Zunächst ist festzuhalten, dass bei der Fülle von Dokumenten eines Korpus keine abschließende und umfassende Relevanzbewertung möglich ist, da auch die Evaluationsorganisatoren keine Übersicht über hunderttausende Dokumente besitzen. Dies ist zum einen der Größe des Korpus geschuldet, aber auch seiner wissenschaftlichen Ausrichtung, die besonders die sozialwissenschaftlich-abstrakten Schreibweisen und Formulierungen als problematisch darstellt. Im Gegensatz zu den klassischen Ad-hoc-Testkollektionen aus TREC und CLEF, die mit Dokumenten aus Tageszeitungen, Journalen oder Agenturarchiven arbeiten, handelt es sich bei GIRT um wissenschaftliche Fachtexte. Das

Korpus	Topics	rel	% rel	Ø rel	nrel	% nrel	Ø nrel	rel+nrel
GIRT								
2003	25	2 117	19 %	84,7	9 020	81 %	360,8	11 137
2004	25	1 663	17 %	66,5	8 073	83 %	322,9	9 736
2005	25	2 682	20 %	107,3	10 506	80 %	420,2	13 188
2006	25	3 759	25 %	150,4	11 189	75 %	447,6	14 948
2007	25	3 689	23 %	147,6	12 599	77 %	504,0	16 288
2008	25	2 290	15 %	91,6	12 503	85 %	500,1	14 793
total	150	16 200	20 %	108,0	63 890	80 %	425,9	80 090
iSearch								
PDF	44	1 343	29 %	30,5	3 213	71 %	73,0	4 556
ABS	48	1 021	30 %	21,3	2 336	70 %	48,7	3 357
Books	45	424	43 %	9,4	568	57 %	12,6	992
Meta	51	2 386	28 %	46,8	6 130	72 %	120,2	8 516
total	65	6 156	27 %	94,7	16 372	73 %	251,9	22 528

Tabelle 8.3: Überblick über die Anzahl der Relevanzbewertungen (rel = Anzahl der relevanten Bewertungen, nrel = nicht relevante Bewertungen, % = Anteil an der Gesamtzahl der Bewertungen, Ø = durchschnittliche Anzahl an Bewertungen pro Topic) für die beiden Korpora GIRT und iSearch.

hier verwendete Vokabular unterscheidet sich immens von der alltäglichen Wortwahl in den Ad-hoc-Testkollektionen. Gleiches gilt auch für den Korpus iSearch, der rein naturwissenschaftlich-physikalisch ausgerichtet ist.

Assessoren in der CLEF-Kampagne wurden aufgrund der zuvor geschilderten Probleme bei der Topicerstellung dazu angehalten „in dubio pro reo“ zu urteilen, soweit dies die Interpretationsfähigkeit des Topics zuließ. Trotzdem sind für sehr spezielle Fragestellungen (z. B. Topic 129 „Sexualität und Behinderung“) nur sehr wenige Dokumente im Relevanzbeurteilten Pool enthalten. Die Organisatoren versuchten dem durch eine allgemeinere Fragestellung entgegenzuwirken, um so den Assessoren größeren Spielraum zu geben. Erschwert wird diese Arbeitsweise durch die binären Bewertungen in GIRT, die keine partielle Relevanz in der Bewertung vorsehen: Ein Dokument ist entweder relevant oder nicht. Der iSearch-Korpus hingegen setzt auf eine vierstufige Bewertungsskala (*highly relevant*, *fairly relevant*, *marginally relevant* und *not relevant*).

In Tabelle 8.3 wird die Anzahl der Relevanzbewertungen und die durchschnittliche Anzahl an Bewertungen pro Topic für die beiden Korpora GIRT und iSearch zusammengefasst. Da die iSearch-Kollektion mit einer vierstufigen Relevanzskala arbeitet, wurden die drei Abstufungen *highly relevant*, *fairly relevant* und *marginally relevant* für diese Tabelle als relevant (rel) zusammengefasst. Die vier Unterkollektionen für iSearch sind Metakollektionen, die folgenden Regeln unterliegen: PDF sind alle vorhandene Metadaten, Abstracts und die PDF-Volltexte; ABS besteht nur aus den Metadaten deren Dokumente

Feldfilter	# alle Dokument	# bewertete Dokumente
kein Filter	135 214	47 198
AUTHOR	117 883	39 850
EDITOR	11 445	4 898
ISSN	48 512	15 581
ISBN	66 347	23 958
SUBJECT	135 214	47 198
CLASSIFICATION	135 214	47 198
METHOD	135 214	47 198
LOCATION	50 168	19 451
PUBYEAR	135 214	47 198
PUBLISHER	33 889	12 751
LANGUAGE	135 214	47 198
AUTHOR+ISSN	117 883	14 871

(a) Größe des GIRT4-DE-EXT-Korpus

Feldfilter	# alle Dokument	# bewertete Dokumente
kein Filter	453 239	9 480
AUTHOR	450 126	9 472
SOURCE	309 138	6 608
PUBYEAR	233 259	3 994
ABSTRACT	444 936	9 298

(b) Größe des iSearch-Korpus (alle Varianten)

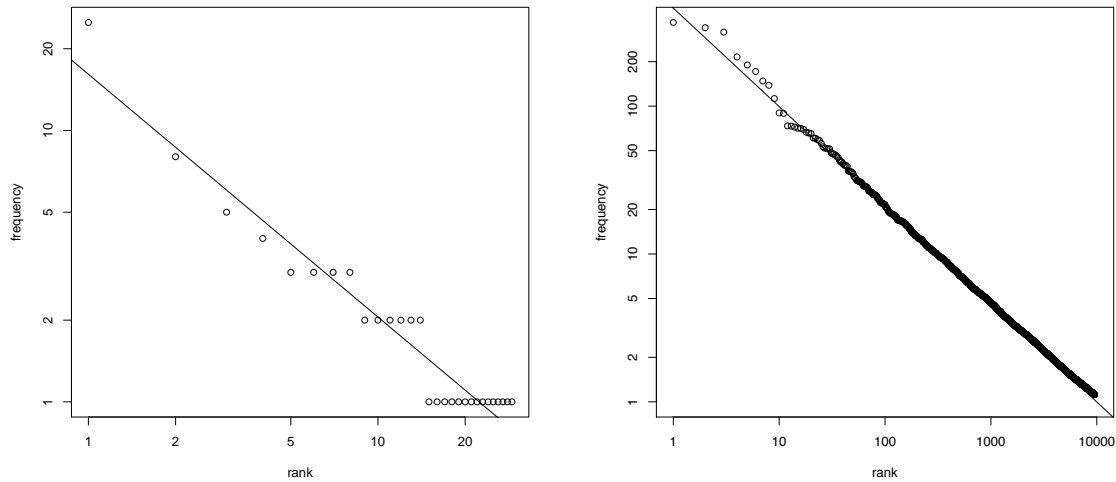
Tabelle 8.4: Die Anzahl aller und der bewerteten Dokumente der beiden Testkollektionen GIRT4-DE-EXT und iSearch (alle Varianten), jeweils gefiltert nach dem Vorhandensein eines bestimmten Feldes.

ein Abstract enthalten; Books sind die Metadaten des Bibliothekskatalogs BK; Meta umfasst alle Metadaten, allerdings ohne die Volltexte.

Die Unterschiede in den Bewertungsphilosophien (binary vs. graded) spiegeln sich auch in den durchschnittlichen Zahlen der Relevanzbewertungen wieder. Bedingt durch die sehr speziellen Suchanfragen der Assessoren sind die Relevanzurteile zum Teil sehr hart, was die sehr geringe Anzahl an hochrelevanten Urteilen wiedergibt. Insgesamt sind die iSearch-Assessoren allerdings etwas großzügiger in der Bewertung vorgegangen, sodass insgesamt der Anteil an relevanten Dokumenten bei 27 % liegt. Verglichen mit den 20 % relevanten Dokumenten des GIRT-Korpus ist diese Zahl allerdings noch recht hoch.

Die Anzahl der insgesamt bewerteten Dokumente ist bei GIRT im Vergleich zu iSearch wesentlich größer. In Tabelle 8.4 sieht man die Größe der Untermenge von GIRT und iSearch. Unterschieden wird zwischen der Gesamtanzahl der Dokumente und der Dokumente, die im Laufe von Kampagnen in einem Pool enthalten waren und daher





(a) Plot von freq1 aus Listing 8.3.

(b) Plot von freq2 aus Listing 8.3.

Abbildung 8.3: Zwei Plots für Power-Law-Verteilungen und deren Approximation mit R.

2. Danach wird aus den möglichen  $\hat{x}_{min}$  Werten derjenige ausgewählt, der den kleinsten  $D$ -Wert aller vorhandenen  $\hat{x}_{min}$  besitzt.

Die Autoren weisen darauf hin, dass dieses Verfahren keine Aussage über die Güte der Parameter oder die Validität der Schätzung macht. Dies ist allerdings systembedingt, da keine Power-Law-Verteilung sondern eine z. B. Log-Normal-Verteilung vorliegen könnte.

Neben der Ermittlung des Exponenten und der  $\hat{x}_{min}$  Parameter kann ebenfalls mittels des Kolmogorov-Smirnov-Tests (s. Abschnitt 6.1.9) die Güte der Schätzung (gof, *goodness-of-fit*) der Parameter errechnet werden. Bei der Berechnung des Tests werden sowohl die Werte gof als auch ein Wahrscheinlichkeitswert  $p$  ermittelt. Wenn  $p$  sehr groß ist (also nahe an 1), kann davon ausgegangen werden, dass ein Unterschied zwischen den empirischen Daten und dem errechneten Modell einer statistischen Schwankung (*statistical fluctuations*) zuzuschreiben ist. Sollte  $p$  sehr klein sein (nahe an 0), so kann davon ausgegangen werden, dass das Power-Law-Modell nicht zu den Daten passt. Dies ist zunächst kontra-intuitiv, da allgemein bei der Überprüfung einer Null-Hypothese kleine Werte von  $p$  als aussagekräftig angesehen werden. In diesem Fall wird allerdings eine Hypothese überprüft, weshalb hohe Werte von  $p$  gewünscht werden. Clauset u. a. (2009, S. 17) schlagen als Schwellenwert  $p \geq 0,1$  vor. Liegt der ermittelte Wert  $p$  über dieser Schranke, so kann von einer Power-Law-Verteilung für die ermittelten Parameter ausgegangen werden.

### 8.3 Zusammenfassung

Als Grundlage für die Evaluation der Verfahren aus Kapitel 7 wurden neben den grundlegenden Evaluationstechniken aus Kapitel 5 zwei konkrete Testkollektionen vorgestellt: Der GIRT-Korpus sozialwissenschaftlicher Dokumente, inklusive der Topics und Relevanzurteile des CLEF-Domain-Specific-Track, als auch die iSearch-Testkollektion, die sich aus Dokumenten des Open-Access- und Preprint-Repositories arXiv aus der Physik zusammensetzt. Beide Kollektionen unterscheiden sich klar in ihrem Aufbau, ihrer fachlich-inhaltlichen Abdeckung und ihrer Erschließungsqualität.





**Teil III**

**Auswertung und Diskussion**



## 9 Auswertung der Experimente zur Anfrageerweiterung

Im folgenden Kapitel werden die Ergebnisse der Evaluation der Systeme, wie sie in Abschnitt 7.1 beschrieben wurden, zusammengetragen. Sie teilen sich auf in die Experimente zum (1) Einfluss der Kombination unterschiedlicher Dokumentattribute, (2) Einfluss der Anfrageformulierung und (3) Einfluss der Dokumentfilter. Die Experimente werden jeweils auf den beiden Testkollektionen GIRT und iSearch durchgeführt. Das Ziel der Untersuchungen ist die Beantwortung der Forschungsfragen 1 und 2, wie sie in der Einleitung dieser Arbeit formuliert wurden.

### 9.1 Aufbau des Experimentes für die Anfrageerweiterung

Wie auch später in den Experimenten zum Reranking werden im folgenden Abschnitt die IR-Testkollektionen GIRT und iSearch verwendet. Es werden die Versuchsaufbauten aus Abschnitt 7.1 angewendet und besonderes Augenmerk auf (1) die Effekte der unterschiedlichen Entitäten bei der Anfrageerweiterung, (2) die Anwendung der Polyrepräsentation in Form von strukturierten und unstrukturierten Anfragen und (3) die Anwendung von Dokumentfiltern gelegt. In Abbildung 9.1 werden die einzelnen Komponenten der Evaluation und deren Permutationen dargestellt.

Für die Evaluation kamen drei verschiedene (Unter-)Korpora in Betracht: Die Gesamtmenge  $D$  aller Dokumente kann weiter unterteilt werden in die beiden Untermengen  $D_t$  und  $D_{tq}$ .  $D_t$  ist die Untermenge von  $D$ , in der alle Dokumente, die zu einem Topic gehören (die also in einem Evaluationspool enthalten waren und folglich bewertet wurden), enthalten sind. Theoretisch gibt es von  $D_t$  wiederum eine Untermenge  $D_{tq}$ , die nur die Dokumente enthält, die zum Topic gehören und durch die automatische Anfragegenerierung, die auf den Titeln der Topicbeschreibungen basiert, erfasst wurden. Es gilt daher:  $D \geq D_t \geq D_{tq}$ .

Für die Aufgabe der Anfrageerweiterung wurde auf die ungefilterte Variante  $D$  zurückgegriffen, da beim Szenario einer Anfrageerweiterung potentiell viele neue Dokumente zur Ergebnismenge hinzugefügt werden, was bei einer Einschränkung der Korpora keinen Mehrwert zeigen würde. Stattdessen wird als einschränkende Variante ein Filter eingebaut, der nur Dokumente, die über einen Autor und eine Quellenangabe bzw. einen ISSN-Code verfügen, für die Evaluation betrachtet. Dieser Subkorpus wird mit  $D_{AS}$  abgekürzt.

Die unterschiedlichen Verfahren (Runs, s. Abschnitt 5.1) werden mit einem Code versehen, der die Konfiguration des Runs beschreibt. Zunächst wird zwischen dem Baselinesystem  $S$  auf Grundlage der Suchmaschine Solr und der Anfrageerweiterung (E) unterschieden. Die Dokumentattribute, die für die Anfrageerweiterung verwendet

## 9 Auswertung der Experimente zur Anfrageerweiterung

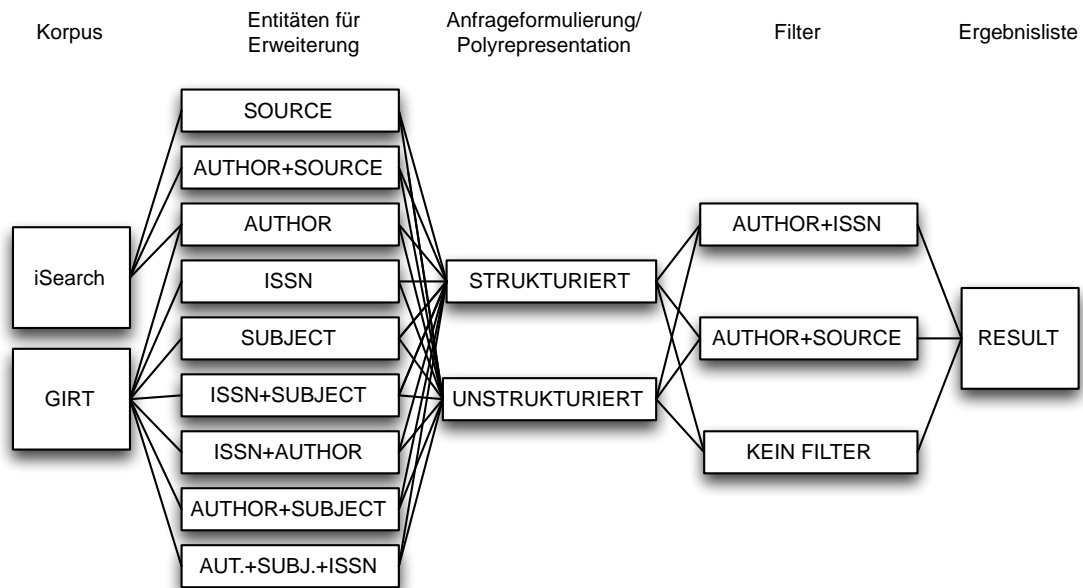


Abbildung 9.1: Eingesetzte Komponenten für die Evaluation der Anfrageerweiterungen: die zwei verschiedenen Korpora, die einzelnen Attribute bzw. Entitäten, die für die Erweiterung genutzt werden, die Art der Anfrageformulierung und die Filterung der Korpora.

werden, werden mit den Abkürzungen A (*author*), S (*source*, bzw. *issn*) und T (*subject*, bzw. *thesaurus*) bezeichnet. Der Code E-A+I beschreibt folglich eine Anfrageerweiterung sowohl mit Autorennamen als auch mit ISSN-Codes.

In den Ergebnistabellen werden die jeweils besten Ergebnisse **fett** markiert, die zweitbesten *kursiv*. Alle Veränderungen gegenüber der Baseline S wurden mittels eines T-Tests auf statistische Signifikanz geprüft und die positiven Veränderungen markiert. Es wurden die folgenden Signifikanzniveaus gekennzeichnet:

(\*) *schwach signifikante* Veränderungen (p-Wert  $\leq 0,1$ ),

\* *signifikante* Veränderungen (p-Wert  $\leq 0,05$ ),

\*\* *hoch signifikante* Veränderungen (p-Wert  $\leq 0,01$ ),

\*\*\* *höchst signifikante* Veränderungen (p-Wert  $\leq 0,001$ ).

Das Baselinesystem S verwendet immer eine unstrukturierte Freitextsuche, wie sie auch in gängigen Suchsystemen (s. Kap. 2) angeboten wird. Die Anfrageerweiterungen werden in zwei Varianten implementiert. In der ersten Variante werden die zur Anfrageerweiterung herangezogenen Attribute feldabhängig bzw. strukturiert genutzt. Es werden Autorennamen nur im Feld *author* und Publikationsquellen nur in den Feldern *source* bzw. *issn* gesucht. In der zweiten Variante wird eine feldunabhängige/unstrukturierte

Anfrageformulierung verwendet, wie sie in Abschnitt 7.1 beschrieben und vorgetestet wurde.

## 9.2 Effekte der Anfrageerweiterung

Die folgenden Unterabschnitte beschäftigen sich jeweils mit einer Gruppe dieser Fragestellungen. Alle Experimente verwenden die gleichen Evaluationsparameter: R, MAP, RP, bpref und die Cut-off-Precision.

### 9.2.1 Einfluss der Kombination unterschiedlicher Dokumentattributen

Zunächst wird für beide Korpora eine konservative Systemkonfiguration gewählt, in der auf der Dokumentmenge  $D$  und mit einer strukturierten Suchanfrage gearbeitet wird. Die Ergebnisse für beide Korpora werden einzeln besprochen.

#### Einfluss der Kombination unterschiedlicher Dokumentattribute für GIRT

Es sind Steigerungen in den Recallwerten für die Systeme E-A, E-T und E-T+A zu beobachten, wobei allerdings nur die Steigerung bei E-A statistisch signifikant ist (s. Tab. 9.1). Der Recallwert der Systeme E-I und E-A+I fällt eindeutig gegenüber der Baseline ab, die anderen Systeme zeigen nur marginale Veränderungen gegenüber der Baseline. Es fällt auf, dass leistungsschwache Systeme mit ISSN-Codes erweitern. Die Erweiterungen mit Autorennamen und Thesaurustermen sind stabiler bzw. zeigen in der Kombination bei E-T+A die besten Ergebnisse, knapp gefolgt von den Ergebnissen des Systems E-T.

Die restlichen Evaluationsparameter zeigen bis auf die bpref-Werte einen sehr starken Abfall gegenüber der Baseline S. Wie auch bei den Recallwerten ist dieses Abfallen besonders prägnant bei Systemen, die mit ISSN-Codes erweitern.

Die starken Einbrüche in den gemessenen Parametern (ausgenommen die bpref-Werte) könnten darauf zurückzuführen sein, dass nur ein Teil der Dokumente in GIRT Zeitschriftenartikel sind, aber alle Dokumente einen Autoreneintrag besitzen bzw. mit Thesaurusbegriffen verschlagwortet wurden.

run	R	MAP	RP	bpref	P@5	P@10	P@20	P@100
S	0,462	<b>0,223</b>	<b>0,271</b>	<b>0,310</b>	<b>0,600</b>	<b>0,535</b>	<b>0,460</b>	<b>0,245</b>
E-A	0,469**	0,107	0,133	0,290	0,307	0,277	0,219	0,109
E-I	0,149	0,044	0,061	0,112	0,325	0,235	0,165	0,054
E-T	<i>0,485</i>	<i>0,159</i>	<i>0,197</i>	0,291	<i>0,412</i>	<i>0,378</i>	<i>0,325</i>	<i>0,194</i>
E-A+I	0,185	0,045	0,081	0,138	0,253	0,213	0,180	0,087
E-T+A	<b>0,491</b>	0,148	0,194	<i>0,293</i>	0,375	0,340	0,294	0,194
E-T+I	0,447	0,131	0,172	0,276	0,363	0,319	0,273	0,175
E-A+T+I	0,454	0,128	0,172	0,279	0,365	0,317	0,276	0,172

Tabelle 9.1: Retrievalergebnisse für eine Anfrageerweiterung für die Dokumentmenge  $D$  aus der GIRT-Testkollektion mit einer strukturierten Anfrageformulierung.

	E-A	E-I	E-T	E-A+I	E-T+A	E-T+I	E-T+I+A	S
E-A	1	-0,221	-0,157	-0,200	0,076	-0,194	-0,144	0,845
E-I		1	-0,403	0,923	-0,550	-0,089	-0,131	-0,178
E-T			1	-0,184	0,813	0,378	0,279	-0,141
E-A+I				1	-0,290	-0,128	-0,231	-0,108
E-T+A					1	0,283	0,325	0,018
E-T+I						1	0,841	-0,173
E-T+I+A							1	-0,170
S								1

Tabelle 9.2: Vergleich der gerankten Ergebnislisten der Systeme E-A, E-I, E-T, E-A+I, E-T+A, E-T+I, E-T+I+A und S mittels Kendalls  $\tau$  für die Systeme aus Tabelle 9.1.

Bei der Gegenüberstellung der gerankten Listen in Tabelle 9.2 fallen die sehr starken Schwankungen zwischen den Systemen auf. Die  $\tau$ -Werte schlagen allgemein stark in den negativen Bereich aus. Im Vergleich zur Baseline zeigt einzig die Erweiterung mit Autorennamen einen hohen  $\tau$ -Wert von 0,845. Das Ranking der beiden Systeme ist sehr ähnlich und die Erweiterung mit Namen hat wenig Einfluss auf die Ergebnisliste. Die hinzugefügten Autorennamen haben in der Gesamtsortierung einen geringen Effekt, jedoch einen großen Einfluss auf MAP, RP und auf die Cut-off-Precision. Die hinzugefügten Autorennamen bringen zwar einige neue relevante Dokumente in die Ergebnisliste ein (signifikant höherer Recall), aber es werden zu viele nicht-bewertete (und daher für alle Evaluationsparameter außer bpref nicht-relevante) Dokumente an den Anfang sortiert.

Die fast gleiche Sortierung von E-I und E-A+I, E-T und E-T+A sowie von E-T+I und E-T+I+A zeigt außerdem den vergleichsweise geringen Einfluss der Autorennamen, da diese Rankings sich mit  $\tau$ -Werten von 0,923, 0,813 und 0,841 kaum unterscheiden.

Obwohl diese Evaluationskonfiguration am ehesten an ein wirkliches Retrievalsystem heranreicht, scheint es für die späteren Versuche sinnvoll, gewisse Einschränkungen für die Dokumentmenge (z. B. bzgl. Zeitschriftenartikel) vorzunehmen.

### Einfluss der Kombination unterschiedlicher Dokumentattribute für iSearch

Bedingt durch den Wegfall der Erweiterung mit kontrollierten Termen bei iSearch bleiben nur drei Permutationen zu betrachten: E-A, E-S und E-A+S. Wie auch in GIRT fallen die ermittelten Werte, die in Tabelle 9.3 aufgelistet sind, teils drastisch ab. So sind für die beiden Systeme E-S und E-A+S sowohl für den Recall als auch für die unterschiedlichen Precisionwerte Verluste gegenüber der Baseline S von bis zu 75% zu verzeichnen; so fällt der MAP-Wert vom Baselinewert 0,081 bis auf 0,016 für E-A+S ab. Einzig das System E-A, das die Anfrage mit Autorennamen erweitert, kann, gemessen an R und bpref, ähnliche Werte wie die Baseline erreichen. Hier liegen die Werte zwar unterhalb der Baseline, doch sind die Abstände von 0,368 zu 0,334 für R und 0,232 zu 0,225 für bpref nicht so groß wie für die beiden anderen Systeme. Der leichte Abfall gegenüber der Baseline von E-A für bpref ist darüber hinaus mit einem  $p$ -Wert von 0,15 statistisch

nicht signifikant.

run	R	MAP	RP	bpref	P@5	P@10	P@20	P@100
S	<b>0,368</b>	<b>0,081</b>	<b>0,121</b>	<b>0,232</b>	<b>0,182</b>	<b>0,154</b>	<b>0,128</b>	<b>0,068</b>
E-A	0,334	0,030	0,031	0,225	0,086	0,065	0,042	0,014
E-S	0,088	0,031	0,044	0,072	0,080	0,062	0,051	0,018
E-A+S	0,098	0,016	0,024	0,084	0,025	0,029	0,025	0,010

Tabelle 9.3: Retrievalergebnisse für eine Anfrageerweiterung für die Dokumentmenge  $D$  aus der iSearch-Testkollektion mit einer strukturierten Anfrageformulierung.

	E-A	E-S	E-A+S	S
E-A	1	-0,519	-0,279	0,456
E-S		1	0,850	-0,344
E-A+S			1	0,037
S				1

Tabelle 9.4: Vergleich der gerankten Ergebnislisten der Systeme E-A, E-S, E-A+S und S mittels Kendalls  $\tau$  für die Systeme aus Tabelle 9.3.

Im Vergleich zur Baseline zeigen alle Varianten der Anfrageerweiterung einen deutlichen Unterschied (s. Tab. 9.4), wobei der  $\tau$ -Wert zwischen S und E-S am niedrigsten ist. Der  $\tau$ -Wert von 0,456 zwischen S und E-A zeigt zwar deutlich die Unterschiedlichkeit in der Sortierung, doch ist dieser im Vergleich zu den anderen Werten moderat. Wie auch bei GIRT ist der Unterschied zwischen E-S und E-A+S mit 0,850 recht gering. Dies überrascht in einem gewissen Maße, da die Retrievalleistungen so stark auseinandergehen trotz einer sehr ähnlichen Sortierung. Die zur Suchanfrage hinzugefügten Quellenangaben bei E-S und E-A+S scheinen folglich einen sehr starken negativen Einfluss auf die Leistungsfähigkeit zu haben.

### 9.2.2 Einfluss der Art der Anfrageformulierung: Strukturierte vs. unstrukturierte Anfragen

Während im vorherigen Abschnitt die Anfrageerweiterung strukturiert war, z. B. Autorennamen in der Anfrageformulierung auch wirklich nur im Feld *author* auftreten durften, wird in den folgenden Experimenten der Einfluss einer gezielten Mischung aus strukturierten und unstrukturierten Anfrage analysiert. Die Baseline bleibt dabei gleich: S ist eine auf Solr basierte Freitextsuche über alle vorhandenen Metadaten bzw. Volltexte (bei iSearch). Die verschiedenen Anfrageerweiterungen suchen jeweils im passenden Feld als auch in anderen Feldern bzw. Volltexten (s. Abschnitt 7.1.1). Im folgenden Abschnitt werden zwei Verfahren gegeneinander gestellt: zum einen eine traditionelle Anfrageerwei-

terung, die die Entitäten der Anfrageerweiterung berücksichtigt und eine Variante, die diese mit einer Breitensuche über alle möglichen Felder kombiniert.

### Einfluss der Anfrageformulierungsart für GIRT

Im Vergleich zu den strukturierten Systemen aus Abschnitt 9.2.1 fallen bei den Ergebnissen aus Tabelle 9.5 die geringeren Veränderungen gegenüber der Baseline und der Systeme untereinander auf. Allerdings geht auch die gemessene Signifikanz in den positiven Veränderungen gegenüber der Baseline verloren. So sind die ehemals signifikanten Verbesserungen für das System E-A im Recall nicht mehr vorhanden. Auch sonst scheinen sich die Ergebnisse eher anzugleichen: Die ehemals besseren Ergebnisse werden schlechter und die schlechteren Ergebnisse werden besser. Die jeweiligen Tiefen und Spitzen der Systeme werden durch die zusätzliche Anfrage im Freitext ausgeglichen.

run	R	MAP	RP	bpref	P@5	P@10	P@20	P@100
S	0,462	<b>0,223</b>	<b>0,271</b>	<b>0,310</b>	<b>0,600</b>	<b>0,535</b>	<b>0,460</b>	<b>0,245</b>
E-A	0,467	0,102	0,129	0,287	0,273	0,253	0,198	0,105
E-I	0,206	0,053	0,076	0,149	0,348	0,265	0,185	0,064
E-T	<i>0,467</i>	<i>0,146</i>	0,182	<i>0,286</i>	<i>0,383</i>	<i>0,340</i>	<i>0,294</i>	0,179
E-A+I	0,235	0,047	0,079	0,163	0,243	0,213	0,176	0,081
E-T+A	<b>0,471</b>	0,141	0,184	0,287	0,376	0,334	0,291	<i>0,182</i>
E-T+I	0,458	0,138	0,177	0,281	0,371	0,331	0,286	0,174
E-A+T+I	0,462	0,136	0,179	0,284	0,373	0,331	0,290	0,178

Tabelle 9.5: Retrievalergebnisse für eine Anfrageerweiterung für die Dokumentmenge  $D$  aus der GIRT-Testkollektion mit einer unstrukturierten Anfrageformulierung.

Tendenziell sind immer noch die gleichen Ergebnisse wie in Abschnitt 9.2.1 zu beobachten. So sind die Systeme, die mit ISSN-Codes in einfacher oder doppelter Kombination arbeiten, in allen gemessenen Evaluationsmaßen die leistungsschwachen Systeme. Die Erweiterung mit Thesaurustermen und Autorennamen führt vereinzelt, wie auch in der Kombination zu leichten, aber nicht signifikanten Verbesserungen in der Baseline. Außer bpref fallen alle Precision-orientierten Werte klar gegenüber der Baseline ab. Die nicht signifikante Verschlechterung gegenüber der Baseline von bpref für diese Systeme zeigt aber auch, dass erneut nicht-bewertete Dokumente für diesen Abfall verantwortlich sein müssen.



	E-A	E-I	E-T	E-A+I	E-T+A	E-T+I	E-T+I+A	S
E-A	1	-0,213	-0,182	-0,221	0,053	-0,228	-0,164	0,799
E-I		1	-0,404	0,952	-0,535	-0,268	-0,288	-0,164
E-T			1	-0,161	0,807	0,713	0,558	-0,167
E-A+I				1	-0,313	-0,125	-0,282	-0,081
E-T+A					1	0,633	0,652	0,017
E-T+I						1	0,829	-0,201
E-T+I+A							1	-0,191
S								1

Tabelle 9.6: Vergleich der gerankten Ergebnislisten der Systeme E-A, E-I, E-T, E-A+I, E-T+A, E-T+I, E-T+I+A und S mittels Kendalls  $\tau$  für die Systeme aus Tabelle 9.5.

Auch bei den gemessenen  $\tau$ -Werten in Tabelle 9.6 zeigen sich keine großen Veränderungen gegenüber der strukturierten Anfrageformulierung, sodass die Ergebnisse der Analyse von Tabelle 9.2 nahezu unverändert übernommen werden können: Der Einfluss der Autorennamen ist gering im Vergleich zu den deutlichen Veränderungen im Ranking, das durch die Thesaurusbegriffe erwirkt wird.

### Einfluss der Anfrageformulierungsart für iSearch

Im Gegensatz zu den eher dämpfenden Eigenschaften der Erweiterung der Anfrage auf Freitextfeldern bei GIRT, zeigt sich bei den Werten für iSearch in Tabelle 9.7, dass die Werte für alle Systeme im Vergleich zur Baseline und Systemen aus Abschnitt 9.2.1 noch weiter abfallen. Weiterhin sind die geringen und nicht signifikanten Unterschiede für R und bpref für E-A aus Abschnitt 9.2.1 in diesem Versuchsaufbau mit  $p$ -Werten von 0,001 für R und 0,018 für bpref statistisch signifikant. Die Verschlechterung ist nicht mehr nur zufällig, sondern substantiell. Die weiteren Systeme E-S und E-A+S können sich hinsichtlich keinem der gemessenen Werte mit der Baseline vergleichen, sie liegen wie zuvor deutlich unter der Baseline, wobei der Wechsel der Anfrageformulierung diesen Trend noch weiter verstärkt hat.

run	R	MAP	RP	bpref	P@5	P@10	P@20	P@100
S	<b>0,368</b>	<b>0,081</b>	<b>0,121</b>	<b>0,232</b>	<b>0,182</b>	<b>0,154</b>	<b>0,128</b>	<b>0,128</b>
E-A	0,316	0,024	0,028	0,216	0,052	0,045	0,032	0,032
E-S	0,056	0,015	0,019	0,045	0,037	0,031	0,026	0,026
E-A+S	0,067	0,014	0,021	0,056	0,015	0,020	0,019	0,019

Tabelle 9.7: Retrievalergebnisse für eine Anfrageerweiterung für die Dokumentmenge  $D$  aus der iSearch-Testkollektion mit einer unstrukturierten Anfrageformulierung.

	E-A	E-S	E-A+S	S
E-A	1	-0,528	-0,280	0,299
E-S		1	0,785	-0,320
E-A+S			1	0,118
S				1

Tabelle 9.8: Vergleich der gerankten Ergebnislisten der Systeme E-A, E-S, E-A+A und S mittels Kendalls  $\tau$  für die Systeme aus Tabelle 9.7.

Die ermittelten Werte für Kendalls  $\tau$  in Tabelle 9.8 zeigen, wie auch die Retrieval-ergebnisse, keine großen Veränderungen im Vergleich zum vorherigen Experiment: Die  $\tau$ -Werte liegen generell unter den vorherigen Werten. Gemeinsam mit den schlechteren Retrievalergebnissen sind auch die Übereinstimmungen in den Rankings zurückgegangen. Ein weiteres Indiz für einen Query-Drift.

### 9.2.3 Einfluss der Dokumentfilter

Während in den beiden vorherigen Experimenten jeweils auf der ungefilterten gesamten Dokumentmenge  $D$  gearbeitet wurde, soll nun ein Dokumentfilter hinzukommen. In der Dokumentmenge  $D_{AS}$  sind nur die Dokumente enthalten, die sowohl einen Autor, als auch eine Publikationsquelle (*source* für iSearch und *issn* für GIRT), besitzen. Die Anfrageformulierung der Anfrageerweiterung wird auf Grundlage der teilweise unterschiedlichen Ergebnisse für die beiden Korpora GIRT und iSearch in Abschnitt 9.2.2 aufgeteilt. Während für GIRT die Anfrage sowohl auf Freitext- als auch auf Feldbasis durchgeführt, wird für iSearch die Formulierung strikt strukturiert durchgeführt. Die Baseline S für beide Korpora ist natürlich nach wie vor eine Freitextsuche über alle vorhandenen Metadaten und Volltexte.

#### Einfluss der Dokumentfilter für GIRT

Generell liegen die ermittelten Retrievalergebnisse in Tabelle 9.9 sowohl für Recall als auch die Precisionwerte unter denen der ungefilterten Baselinesysteme aus den Abschnitten 9.2.1 und 9.2.2. Dies ist durch den doppelten Dokumentfilter bei  $D_{AS}$  zu begründen, der die Dokumentmenge erheblich verkleinert und somit auch viele potentiell relevant Dokumente ausfiltert. Durch die Dokumentfilter kann allerdings ein Effekt der mehrfachen Anfrageerweiterung wiederholt werden, wie er z. B. auch schon in den Pretest aus Abschnitt 7.1.3 zu sehen war: Die Kombination mehrerer Dokumentattribute für die Anfrageerweiterung zeigt positive Effekte.

Wie in Tabelle 9.9 zu sehen, bleibt es zunächst bei den negativen Auswirkungen für die Ergebnisse von MAP, RP und den unterschiedlichen Cut-off-Precisionwerten. Diese fallen alle verglichen mit der Baseline S klar ab. Im Vergleich zu den Experimenten mit dem Dokumentenkörper  $D$  ist dieser Abfall für  $D_{AS}$  aber nicht so drastisch. Wie auch in den vorherigen Experimenten ist die Erweiterung mit ISSN-Codes in den beiden Systemen E-I und E-A+I wesentlich schlechter als die Baseline. Der Einfluss der Autorennamen ist bei E-A+I gering und kann den Query-Drift nicht verhindern. Die Autorennamen

run	R	MAP	RP	bpref	P@5	P@10	P@20	P@100
S	0,164	<b>0,078</b>	<b>0,116</b>	0,130	<b>0,463</b>	<b>0,393</b>	<b>0,295</b>	<b>0,115</b>
E-A	0,167***	0,040	0,088	0,131***	0,207	0,165	0,136	0,089
E-I	0,105	0,038	0,058	0,087	0,317	0,233	0,165	0,055
E-T	0,197**	0,061	0,103	0,143	0,303	0,265	0,213	0,100
E-A+I	0,111	0,026	0,055	0,091	0,200	0,153	0,117	0,058
E-T+A	0,200**	0,059	0,102	0,146	0,304	0,256	0,208	0,100
E-T+I	0,197**	0,057	0,096	0,144	0,295	0,256	0,201	0,094
E-A+T+I	<b>0,200**</b>	0,056	0,096	<b>0,146</b>	0,300	0,261	0,207	0,096

Tabelle 9.9: Retrievalergebnisse für eine Anfrageerweiterung für die Dokumentmenge  $D_{AS}$  aus der GIRT-Testkollektion mit einer unstrukturierten Anfrageformulierung.

	E-A	E-I	E-T	E-A+I	E-T+A	E-T+I	E-T+I+A	S
E-A	1	-0,153	-0,133	-0,202	0,037	-0,114	-0,127	0,905
E-I		1	-0,394	0,977	-0,514	-0,157	-0,169	-0,122
E-T			1	-0,180	0,927	0,439	0,390	-0,116
E-A+I				1	-0,231	-0,144	-0,199	-0,142
E-T+A					1	0,391	0,406	0,007
E-T+I						1	0,930	-0,125
E-T+I+A							1	-0,126
S								1

Tabelle 9.10: Vergleich der gerankten Ergebnislisten der Systeme E-A, E-I, E-T, E-A+I, E-T+A, E-T+I, E-T+I+A und S mittels Kendalls  $\tau$  für die Systeme aus Tabelle 9.9.

zeigen aber in ihrer alleinigen Kombination bei E-A einen zwar geringen, aber statistisch signifikanten Anstieg bei bpref gegenüber der Baseline S. Bis auf die Ausnahme der beiden ISSN-Systeme E-I und E-A+I liegen alle Systeme bei R und bpref über der Baseline S. Im Falle der Recallwerte sind die Anstiege alle statistisch signifikant. Die Steigerung der bpref-Rate liegt im Maximum bei einer Steigerung von 0,130 auf 0,146 für die Systeme E-T+A bzw. E-A+T+I. Diese beiden Systeme liefern die besten Ergebnisse und unterscheiden sich nur geringfügig.

Betrachtet man die  $\tau$ -Werte in Tabelle 9.10 sieht man, dass, obwohl die Retrievalwerte sehr nah beieinanderliegen, die Rankings jedoch unterschiedlich sind. Die Systeme E-T+A und E-T+I+A unterscheiden sich zwar in Recall und bpref kaum, allerdings sind die Rankings mit einem  $\tau$ -Wert von 0,406 essentiell unterschiedlich. Der vergleichsweise schwache Einfluss der Autorennamen auf die Rankings der Systeme zeigt sich auch für  $D_{AS}$ , so liegt nur der  $\tau$ -Wert für die Systeme S und E-A klar über 0 und zeigt mit einem Wert von 0,905 auch die faktisch sehr hohe Ähnlichkeit der beiden Rankings an.

**Einfluss der Dokumentfilter für iSearch**

Auch bei Experimenten mit dem iSearch-Korpus zeigen sich die Auswirkungen der Dokumentenfilter klar in Tabelle 9.11. Das System E-A liegt in den für die Anfrageerweiterungsexperimente wichtigen Evaluationskennzahlen R und bpref gleichauf mit der Baseline bzw. ganz knapp darüber. Statistisch signifikant sind die minimalen Unterschiede nicht, doch zeigen sie noch einmal klar, dass bei der Erweiterung viele nicht-bewertete Dokumente des iSearch-Korpus nach vorne sortiert werden. Die für diese Tatsache anfälligen Kennzahlen wie MAP, RP und die Cut-off-Precisionwerte liegen daher auch für alle evaluierten Systeme E-A, E-S und E-A+S deutlich unter der Baseline.

In Tabelle 9.12 sieht man die gleichen Muster, wie auch bei den vorherigen Experimenten mit iSearch: Die Unterschiede zur Baseline sind für E-S und E-A+S sehr groß. Zwar sind die Unterschiede auch bei E-A mit einem  $\tau$ -Wert von 0,551 klar vorhanden, aber kann hier in Verbindung mit den guten bpref- und R-Werten von einer substantiellen Veränderung gegenüber der Baseline ausgegangen werden. Die Erweiterung mit Autorennamen hat also auch für diesen Korpus positive Effekte.

run	R	MAP	RP	bpref	P@5	P@10	P@20	P@100
S	<b>0,179</b>	<b>0,025</b>	<b>0,052</b>	<i>0,106</i>	<b>0,129</b>	<b>0,100</b>	<b>0,074</b>	<b>0,040</b>
E-A	<i>0,175</i>	<i>0,006</i>	0,010	<b>0,106</b>	0,028	0,023	0,014	0,007
E-S	0,038	0,006	<i>0,016</i>	0,033	<i>0,046</i>	<i>0,038</i>	<i>0,028</i>	<i>0,011</i>
E-A+S	0,042	0,002	0,008	0,038	0,012	0,014	0,009	0,004

Tabelle 9.11: Retrievalergebnisse für eine Anfrageerweiterung für die Dokumentmenge  $D$  aus der iSearch-Testkollektion mit einer strukturierten Anfrageformulierung.

	E-A	E-S	E-A+S	S
E-A	1	-0,473	-0,320	0,551
E-S		1	0,920	-0,342
E-A+S			1	0,005
S				1

Tabelle 9.12: Vergleich der gerankten Ergebnislisten der Systeme E-A, E-S, E-A+S und S mittels Kendalls  $\tau$  für die Systeme aus Tabelle 9.11.

**9.3 Zusammenfassung der Ergebnisse**

In den Abschnitten 7.1.2 und 7.1.3 wurden zwei Pretests zu den evaluierten Anfrageerweiterungen beschrieben. In den Pretests zeigte sich, dass sowohl eine Erweiterung mit Thesaurustermen, als auch eine Kombination von Thesaurustermen und Autorennamen einen positiven Effekt auf die Recall- und die Precisionwerte haben konnte. In den nun

vorliegenden Evaluationen, die auf einer größeren Anzahl an Topics und zwei unterschiedlichen Testkollektionen basierten, zeigen sich im Detail andere Ergebnisse als noch in den Pretests zu vermuten war.

Die Leistung der jeweiligen Systeme ist in den einzeln untersuchten Gruppen weit gestreut, sodass die Einflüsse und Effekte der jeweiligen Systemparameter zum Vorschein kommen. Diese Einflüsse und Effekte wurden basierend auf den drei Aspekten (1) Kombination der Dokumentattribute, (2) Art der Anfrageformulierung und (3) eingesetzte Dokumentfilter untersucht.

Die Unterschiede in den Testkollektionen und die darauf basierenden Auswirkungen auf die Anfrageerweiterungen führten zu teils widersprüchlichen Ergebnissen. War in dem Pretest in Abschnitt 7.1.3 noch die einfache These zu vertreten, dass je mehr Dokumentattribute zur Anfrageerweiterung hinzugezogen werden, desto besser das Gesamtergebnis ausfalle, zeigt sich im vorliegenden Experiment eine andere Tendenz: Die Wahl der jeweiligen Erweiterungsattribute ist, neben der Abhängigkeit vom jeweiligen Dokumentkorpus, essentiell.

**1. Forschungsfrage** *Für welche Entitäten bzw. Kombinationen von Entitäten einer digitalen Bibliothek eignet sich die Kookkurrenzanalyse als Werkzeug für eine Anfrageerweiterung?*

Es zeigt sich, dass für beide Korpora die Erweiterung mit ISSN bzw. Publikationsquellen nicht die gewünschten positiven Ergebnisse erbringen konnte. Fast alle Systeme, die auf diese Dokumentattribute zurückgriffen, hatten mit klaren Einbußen gegenüber der Baseline zu kämpfen, die allerdings wieder durch die Kombination mit anderen Dokumentattributen wie Thesaurusbegriffen oder Autorennamen abgeschwächt werden konnte. Für GIRT zeigte sich vor allem der große Einfluss, den die Thesaurusbegriffe bei den Anfrageerweiterungen haben. Durch ihren Einsatz konnte eine nahezu vollständig andere Sortierung der Ergebnislisten, bei im Vergleich zur Baseline gleichbleibender Retrievalleistung, erreicht werden. Einen weiteren positiven Effekt hatte die Polyrepräsentation in Form von unstrukturierten Anfragen: Die Leistungsspitzen und -tiefen wurden durch die zusätzlichen Anfragen im Freitext ausgeglichen.

Deutlich wurden die positiven Effekte der Anfrageerweiterung allerdings erst, als die Testkollektionen künstlich verkleinert wurden und anstelle der gesamten Dokumentmengen  $D$  die kleineren Untermengen  $D_{AS}$  verwendet wurden. Für GIRT waren nun klar die positiven Einflüsse der Autorennamen und der Thesaurusterme zu erkennen. Für iSearch waren die Effekte nicht so ersichtlich und aufgrund der fehlenden Verschlagwortung auch nur mit den Autorennamen reproduzierbar. In beiden Fällen konnte sowohl Recall als auch bpref mindestens auf dem Niveau der Baseline bleiben oder diese klar übertreffen. Gleichzeitig zeigte die Analyse mit Kendalls  $\tau$  die Unterschiede in der Sortierung.

**2. Forschungsfrage** *Lassen sich positive Veränderungen in der Retrievalleistung gegenüber einer State-of-the-Art-Baseline im Rahmen einer Cranfield-Evaluation messen?*

Bei der Analyse der Anfrageerweiterungen wurde deutlich, dass die verwendeten Evaluationskennzahlen teilweise ungeeignet waren. Im Gegensatz zu den Pretests waren die

## 9 Auswertung der Experimente zur Anfrageerweiterung

Evaluationskennzahlen MAP, RP und die Cut-off-Precisionwerte nicht so stabil wie erwartet. Durch die Größe der zunächst ungefilterten Dokumentenkorpora war die Anzahl an nicht-bewerteten Dokumenten, die durch die Anfrageerweiterungen zu den Ergebnislisten hinzugefügt wurden, sehr hoch. Die verwendeten Evaluationskennzahlen sind für diese hohe Anzahl an nicht-bewerteten Dokumenten empfindlich, sodass diese Kennzahlen in nahezu allen Experimenten stark gegenüber der Baseline abfallen. Die sinnvollen und leistungsfähigen Kombinationen der unterschiedlichen Dokumentattribute zeigen sich erst beim Einsatz der Kennzahl *bpref*, da sich diese gegenüber nicht-bewerteter Dokumente stabil verhält. Erst durch ihren Einsatz konnten die Verbesserungen der erfolgreichsten Implementationen E-A+T+I für GIRT und E-A für iSearch sichtbar gemacht werden. Zusammen mit den signifikant gesteigerten Recallwerten für beide Systeme kann dies als ein erfolgreiches Ergebnis gewertet werden.

# 10 Auswertung der Experimente zum Reranking

Im folgenden Kapitel werden die Ergebnisse der Evaluation der Rerankingverfahren, wie sie in Abschnitt 7.2 beschrieben wurden, zusammengetragen. Sie teilen sich auf in Beschreibungen der Ergebnisse der Analyse der Power-Law-Verteilungen in GIRT und iSearch und in die darauf basierenden alternativen Rankingverfahren. Ziel der aufgeführten Untersuchungen ist die Bearbeitung der Forschungsfragen 3–6, wie sie in der Einleitung dieser Arbeit formuliert wurden.

## 10.1 Aufbau der Power-Law-Analyse und der Reranking-Experimente

In Abbildung 10.1 ist eine schematische Darstellung der Evaluationsparameter zu sehen. Es werden die beiden Korpora GIRT und iSearch verwendet, die jeweils mit acht bzw. zwei verschiedenen Dokumentattributen gerankt werden, wobei eine von vier verschiedenen Gewichtungen (Boost) und das Standardranking TF\*IDF zum Einsatz kommen. So ergeben sich insgesamt für GIRT  $8 \cdot 5 = 40$  und für iSearch  $2 \cdot 5 = 10$  verschiedene Kombinationen.

Im Einzelnen werden für GIRT die Entitäten *subject* (Thesaurus-Verschlagwortung), *author* (Autorennamen), *issn*, *classification* (Klassifikation Sozialwissenschaften), *pubyear* (Erscheinungsjahr), *publisher* (Herausgebername), *location* (Erscheinungsort) und *method* (Forschungsmethode) und für iSearch die Entitäten *author* und *source* (Zeitschriften oder Konferenzname) verwendet.

Wie in Abbildung 10.1 zu sehen, werden zusätzlich für jede Reranking-Entität vier verschiedene Gewichtsverfahren, wie in Abschnitt 7.2.2 beschrieben, angewendet:  $B_1$  (einfache Boostgewichtung),  $B_2$  (quadrierte Boostgewichtung, Mischzonen),  $B_4$  (klassisches Zonen-Reranking) und  $B_E$  (Boostgewichtung auf Grundlage von EF\*IDF).

Alle Experimente wurden analog zu den Experimenten in Kapitel 9 auf einer Untermenge  $D_t$  der Gesamtmenge  $D$  aller Dokumente ausgeführt.  $D_t$  ist die Untermenge von  $D$ , in der alle Dokumente, die zu einem Topic gehören (die also in einem Evaluationspool enthalten waren und folglich bewertet wurden), enthalten sind. Theoretisch gibt es von  $D_t$  wiederum eine Untermenge  $D_{tq}$ , die nur die Dokumente enthält, die sowohl zum Topic gehören als auch durch die automatische Generierung der Anfragen, die auf den Titeln der Topicbeschreibungen basiert, erfasst wurden. Es gilt daher:  $D \geq D_t \geq D_{tq}$ . Die unterschiedlichen Mengen stellen verschiedene Anwendungsgebiete und Filter dar. In den folgenden Abschnitten werden nur die Ergebnisse der Evaluation mit  $D_t$  im Einzelnen

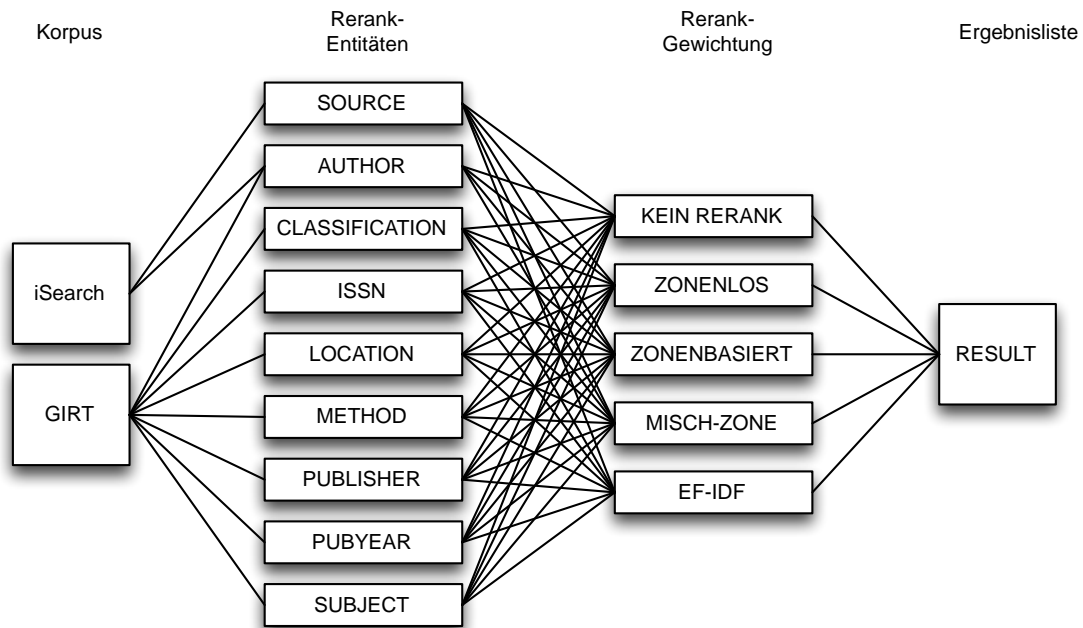


Abbildung 10.1: Eingesetzte Komponenten für die Reranking-Experimente: die zwei verschiedenen Korpora, die Attribute bzw. Entitäten, die zum Reranking genutzt werden sowie die Art der Reranking-Gewichtung.

besprochen, da für  $D$  die Anzahl der nicht-bewerteten Dokumente in der Ergebnismenge zu groß ist und für  $D_{tq}$  die Ergebnismenge wiederum zu klein ist, als dass mit ihr noch eine gezielte Power-Law-Analyse durchgeführt werden könnte.

In den Ergebnistabellen werden die jeweils besten Ergebnisse **fett** markiert, die zweitbesten *kursiv*. Alle Veränderungen gegenüber der Baseline S wurden mittels eines T-Tests auf statistische Signifikanz geprüft und die positiven Veränderungen markiert. Es werden die gleichen Signifikanzniveaus wie in Abschnitt 9.1 gekennzeichnet:

- (\*) *schwach signifikante* Veränderungen (p-Wert  $\leq 0,1$ ),
- \* *signifikante* Veränderungen (p-Wert  $\leq 0,05$ ),
- \*\* *hoch signifikante* Veränderungen (p-Wert  $\leq 0,01$ ),
- \*\*\* *höchst signifikante* Veränderungen (p-Wert  $\leq 0,001$ ).

## 10.2 Power-Law-Verteilungen in GIRT und iSearch

Die folgenden Tabellen 10.1 und 10.2 zeigen die Ergebnisse einer Analyse der in Abschnitt 7.2.2 beschriebenen Verfahren zur Bestimmung von Power-Laws in einer Häufigkeitsverteilung. Mittels der von Clauset u. a. (2009) bereitgestellten Verfahren wurde für unterschiedliche Dokumentmengen und Dokumentattribute eine Frequenzanalyse durchgeführt, um



zu testen, ob es sich bei den gemessenen Frequenzen um Power-Law-Verteilungen handelt oder nicht. Die Tabellen zeigen die ermittelten Power-Law-Exponenten für alle drei Dokumentmengen  $D$ ,  $D_t$  und  $D_{tq}$ . Angegeben werden der Durchschnitt des geschätzten Exponenten  $\hat{\alpha}$ , die Anzahl der tatsächlich ermittelten Exponenten, angegeben durch die Anzahl der errechneten  $\hat{x}_{min}$  und die prozentuale Abdeckung im Vergleich zur maximalen Anzahl der Topics. Weiterhin werden die durchschnittliche Größe von  $\hat{x}_{min}$  und der größte Wert für  $\hat{x}_{min}$  angegeben.

Auf Grundlage dieser Werte in Tabelle 10.1 sieht man, dass für den GIRT-Korpus und die Dokumentmenge  $D$  im Durchschnitt in 89 % der Fälle  $\hat{x}_{min}$  durch den Algorithmus geschätzt werden konnte. Gemittelt über alle acht Entitäten wurde in 89 % der Fälle eine potentielle Power-Law-Verteilung gefunden. Für  $D_t$  und  $D_{tq}$  liegen diese Werte bei 100 % und 86 %. Dies sagt noch nichts über die tatsächliche Existenz eines Power-Laws aus, da hierzu ein Kolmogorov-Smirnov-Test durchgeführt werden muss (s. Abschnitt 8.2). Für die weiteren Analysen wird allerdings, wie bereits beschrieben, nur die Untermenge  $D_t$  betrachtet. Die späteren Berechnungen und statistischen Abschätzungen der Wahrscheinlichkeiten einer Power-Law-Verteilung sind nur für diese Dokumentmenge operationalisierbar. Die anderen Dokumentmengen sind zu groß, sodass einzelne Entitäten direkt aus dem Raster fallen ( $D$  hat vier Entitäten, die erst ab der einhundertsten bzw. zweihundertsten Stelle ein  $\hat{x}_{min}$  besitzen).

Analog zum GIRT-Korpus wurden die gleichen Analysen mit dem iSearch-Korpus durchgeführt. Durch die unkontrollierte Erschließung der Datensätze (s. Analyse in Abschnitt 8.1.1) wurde die Power-Law-Analyse nur für zwei Entitäten durchgeführt (s. Tab. 10.2. Auch hier ist ein zu GIRT analoges Verhalten zu beobachten. Obwohl nicht für 100 % der Topics eine Power-Law-Verteilung ermittelt werden konnte, sondern im Schnitt nur für 96 %, ist für die Untermenge  $D_t$  die Anzahl der ermittelten  $\hat{x}_{min}$  hier am höchsten. Zwar ist für  $D$  die prozentuale Abdeckung der ermittelten  $\hat{x}_{min}$  mit 95,5 % fast genauso hoch, doch ist hier der durchschnittliche Wert von  $\hat{x}_{min}$  wesentlich höher. In  $D$  finden sich viele Power-Laws erst in der Mitte oder am Ende der Verteilung. Das Verhältnis ist für  $D_t$  ausgewogener. Wie auch bei GIRT ist für  $D_{tq}$  die ermittelte Dokumentmenge im Schnitt zu klein, um eine durchgängige Analyse durchzuführen, was sich in den vergleichsweise niedrigen Zahlen von  $\hat{x}_{min}$  zeigt.

Auf Grundlage der Auswertung der Parameter  $\hat{\alpha}$ ,  $|\hat{x}_{min}|$  sowie  $\max(\hat{x}_{min})$  wurde auf eine weitere Auswertung der beiden Dokumentmengen  $D$  und  $D_{tq}$ , sowohl für GIRT als auch für iSearch, verzichtet. Zunächst dauerte die Auswertung mit mindestens 1000 Iterationen, die für ein statistisch vertretbares Ergebnis nötig sind, für die Dokumentmenge  $D$  aufgrund ihrer Größe zu lange und überdies war für  $D_{tq}$  die Dokumentmenge zu klein, sodass für zu viele Topics und Entitäten keine Berechnung hätte durchgeführt werden können. Beide Varianten stehen in keinem vernünftigen Verhältnis zu Aufwand und zu erwartenden Nutzen. Da alle zu erwartenden Verteilungen auch in der Dokumentmenge  $D_t$  auftreten, kann auf eine weitere Betrachtung der beiden anderen Dokumentmengen an dieser Stelle verzichtet werden. In Anhang B ist eine Einzelzusammenstellung aller Häufigkeitsverteilungen und der auf ihrer Grundlage ermittelten Power-Law-Parameter sowie einem Plot zusammengefasst. Im Folgenden werden nur die Zusammenfassungen dieser Daten besprochen.

## 10 Auswertung der Experimente zum Reranking

Entität	$\varnothing \hat{\alpha}$	$ x $	$ \hat{x}_{min} $	%	$\varnothing \hat{x}_{min}$	$\max(\hat{x}_{min})$
author	2,99	150	125	83 %	2,34	7
classification	2,31	150	140	93 %	200,42	1 605
issn	2,44	150	126	84 %	14,39	94
location	1,87	150	127	85 %	4,31	31
method	2,04	150	141	94 %	277,87	2 344
publisher	1,95	150	128	85 %	3,71	36
pubyear	2,23	150	138	92 %	115,01	2 549
subject	2,49	150	142	95 %	111,48	883

(a) Übersicht über die ermittelten Parameter für die Untermenge  $D$

Entität	$\varnothing \hat{\alpha}$	$ x $	$ \hat{x}_{min} $	%	$\varnothing \hat{x}_{min}$	$\max(\hat{x}_{min})$
author	3,27	150	150	100 %	2,09	4
classification	1,98	150	150	100 %	6,00	25
issn	2,66	150	150	100 %	2,38	5
location	2,12	150	150	100 %	2,79	12
method	1,97	150	150	100 %	12,63	63
publisher	2,64	150	150	100 %	2,27	6
pubyear	2,65	150	150	100 %	20,76	59
subject	2,19	150	150	100 %	5,03	18

(b) Übersicht über die ermittelten Parameter für die Untermenge  $D_t$

Entität	$\varnothing \hat{\alpha}$	$ x $	$ \hat{x}_{min} $	%	$\varnothing \hat{x}_{min}$	$\max(\hat{x}_{min})$
author	3,28	150	117	78 %	2,12	4
classification	2,05	150	140	93 %	4,16	22
issn	2,62	150	120	80 %	2,46	9
location	2,21	150	123	82 %	2,84	13
method	2,06	150	141	94 %	8,58	67
publisher	2,62	150	113	75 %	2,18	4
pubyear	2,72	150	138	92 %	13,17	41
subject	2,30	150	142	95 %	3,64	18

(c) Übersicht über die ermittelten Parameter für die Untermenge  $D_{tq}$

Tabelle 10.1: Übersicht über die ermittelten Power-Law-Exponenten  $\hat{\alpha}$  für 150 Topic im GIRT-Datensatz und 8 verschiedene Entitäten jeweils für die Dokumentmengen  $D$ ,  $D_t$  und  $D_{tq}$ . Angegeben werden der Durchschnitt des Exponenten  $\hat{\alpha}$ , die Anzahl der tatsächlich ermittelten Exponenten, angegeben durch die Anzahl der errechneten  $\hat{x}_{min}$ -Werte und die prozentuale Abdeckung im Vergleich zur maximalen Anzahl der Topics. Weiterhin wird die durchschnittliche Größe von  $\hat{x}_{min}$  und der größte Wert für  $\hat{x}_{min}$  angegeben.

Entität	$\varnothing \hat{\alpha}$	$ x $	$ \hat{x}_{min} $	%	$\varnothing \hat{x}_{min}$	$\max(\hat{x}_{min})$
author	3,27	65	61	94 %	8,56	42
source	1,84	65	63	97 %	2,78	13

(a) Übersicht über die ermittelten Parameter für die Untermenge  $D$ 

Entität	$\varnothing \hat{\alpha}$	$ x $	$ \hat{x}_{min} $	%	$\varnothing \hat{x}_{min}$	$\max(\hat{x}_{min})$
author	3,43	65	61	94 %	2,13	3
source	2,03	65	64	98 %	2,58	14

(b) Übersicht über die ermittelten Parameter für die Untermenge  $D_t$ 

Entität	$\varnothing \hat{\alpha}$	$ x $	$ \hat{x}_{min} $	%	$\varnothing \hat{x}_{min}$	$\max(\hat{x}_{min})$
author	3,46	65	43	66 %	2,09	3
source	1,86	65	49	75 %	2,76	14

(c) Übersicht über die ermittelten Parameter für die Untermenge  $D_{tq}$ 

Tabelle 10.2: Übersicht über die ermittelten Power-Law-Exponenten  $\hat{\alpha}$  für 65 Topics in iSearch und zwei verschiedene Entitäten jeweils für die Dokumentmengen  $D$ ,  $D_t$  und  $D_{tq}$ . Angegeben werden der Durchschnitt des Exponent  $\hat{\alpha}$ , die Anzahl der tatsächlich ermittelten Exponenten, angegeben durch die Anzahl der errechneten  $\hat{x}_{min}$  Werte und die prozentuale Abdeckung im Vergleich zur maximalen Anzahl der Topics. Weiterhin wird die durchschnittliche Größe von  $\hat{x}_{min}$  und der größte Wert für  $\hat{x}_{min}$  angegeben.

Die Tabellen 10.3a und 10.3b zeigen die Ergebnisse der ermittelten *goodness-of-fit*-Werte (gof), die den Abstand zwischen den empirischen Daten und dem geschätzten Modell beschreiben. Der ebenfalls ermittelte Wert  $p$  beschreibt die Plausibilität der Power-Law-Hypothese und basiert auf dem Vergleich des geschätzten Modells mit den empirischen und mit synthetischen Daten (Details zur Berechnung sind der Publikation von Clauset u. a. (2009, S. 676) zu entnehmen). Eine Aufstellung aller Einzelwerte für die jeweiligen einzelnen Topics findet sich im Anhang B.

Anhand der durchschnittlichen  $p$ -Werte im GIRT-Korpus sieht man, dass sowohl *classification* als auch *pubyear* die unwahrscheinlichsten Power-Law-Verteilungen darstellen. Kombiniert man die  $p$ -Werte mit dem durchschnittlichen  $\hat{x}_{min}$ , sieht man, dass auch *method* und *subject* im Durchschnitt einen niedrigen  $p$ -Wert und einen hohen  $\hat{x}_{min}$  erzielen. Die Entitäten *author*, *classification*, *location* und *publisher* besitzen hingegen sowohl hohe  $p$ - (alle über 0,50) als auch niedrige  $\hat{x}_{min}$ -Werte (alle unter 2,79). Für diese Entitäten ist im Schnitt häufig eine tatsächliche Power-Law-Verteilung zu beobachten. Für iSearch konnte sowohl für *author* als auch für *source* eine hohe Anzahl an  $\hat{x}_{min}$  ermittelt werden, doch ist die Wahrscheinlichkeit einer tatsächlichen Power-Law-Verteilung bei *source* mit 0,66 gegenüber 0,34 für *author* fast doppelt so hoch. Im Schnitt scheint aber für beide

Entität	$\varnothing p$	$\varnothing \text{gof}$	$\varnothing \hat{\alpha}$	$\varnothing \hat{x}_{min}$	$F_p$	$F_{\hat{\alpha}}$	$F_{p\hat{\alpha}}$
author	0,50	0,06	3,27	2,09	132	64	62
classification	0,19	0,08	1,98	6,00	77	90	36
issn	0,59	0,09	2,66	2,38	143	81	80
location	0,55	0,10	2,11	2,79	147	94	92
method	0,43	0,12	1,98	12,63	126	85	63
publisher	0,59	0,09	2,66	2,27	145	88	86
pubyear	0,28	0,21	2,62	20,76	97	55	17
subject	0,32	0,03	2,19	5,03	104	79	41

(a) Ergebnisse für GIRT

Entität	$\varnothing p$	$\varnothing \text{gof}$	$\varnothing \hat{\alpha}$	$\varnothing \hat{x}_{min}$	$F_p$	$F_{\hat{\alpha}}$	$F_{p\hat{\alpha}}$
author	0,34	0,09	3,43	61	41	11	10
source	0,66	0,15	2,03	64	63	34	33

(b) Ergebnisse für iSearch

Tabelle 10.3: Übersicht über die ermittelten  $p$ - und  $\text{gof}$ -Werte für 150 Topics in GIRT und 65 Topics in iSearch und acht bzw. zwei verschiedene Entitäten für die Dokumentmengen  $D_t$ . Angegeben werden der Durchschnitt von  $p$  und  $\text{gof}$  sowie die bereits in den Tabellen 10.1 und 10.2 ermittelten durchschnittlichen Werte für  $\hat{\alpha}$  und  $\hat{x}_{min}$ . Zusätzlich werden die Anzahl der verbleibenden Topics ermittelt, wenn mit  $F_p = |p \geq 0,1|$ ,  $F_{\hat{\alpha}} = |\hat{\alpha}_i \leq \varnothing \hat{\alpha}|$  bzw. beiden Kriterien  $F_{p\hat{\alpha}} = |p \geq 0,1 \cup \hat{\alpha}_i \leq \varnothing \hat{\alpha}|$  gefiltert würde.

Entitäten oft eine tatsächliche Power-Law-Verteilung vorzuliegen. Filtert man für GIRT die vorhandenen 150 Topics hinsichtlich der Wahrscheinlichkeit einer tatsächlich ermittelten Power-Law-Verteilung (anhand von  $p \geq 0,1$ ), so bleiben sowohl für *classification* als auch für *pubyear* weniger als 100 Topics zurück, die ein potentielles Power-Law enthalten. Zieht man zusätzlich den Durchschnitt der ermittelten Werte für  $\hat{\alpha}$ , den Exponenten des geschätzten Power-Laws, der die negative Steigung (*slope*) der Verteilungen angibt, hinzu, so sieht man sehr unterschiedliche Werte. So ist für *author* mit einem  $\hat{\alpha}$ -Wert von 3,27 ein sehr viel stärker abfallendes Power-Law im Vergleich z. B. zu *method* oder *classification* mit einem  $\hat{\alpha}$ -Wert von 1,98 zu beobachten. Auch der durchschnittliche Wert von  $\hat{\alpha}$  kann als Filter verwendet werden, um im Vergleich zum Durchschnitt extreme Verteilungen bzw. Ausreißer zu unterdrücken. Kombiniert man beide Filterkriterien, so bleiben z. B. für *pubyear* nur 17 Topics von 150 übrig. Für iSearch verbleiben für *author* mit 41 von 65 knapp 2/3 der ermittelten Power-Law-Exponenten oberhalb der Schwelle von  $p \geq 0,1$ , bei *source* sind dies mit 63 von 65 fast alle. Filtert man anhand der durchschnittlichen Größe von  $\hat{\alpha}$ , so bleiben erheblich weniger Topics übrig. Diese Werte können im späteren Verlauf des Experiments verwendet werden, wie sich im folgenden Abschnitt zeigt.

## 10.3 Effekte des Rerankings

Im folgenden Abschnitt werden die Ergebnisse der Retrieval-Evaluation für zwei Rankingverfahren und vier Boostvarianten zusammengefasst. Das Rankingverfahren S ist das Baselinesystem auf Grundlage der Suchmaschine Solr, R bezeichnet das Rerankingverfahren und B die verschiedenen Boostvarianten. Zusätzlich sind an die beiden Rankingverfahren die Kürzel der jeweilig betrachteten Entität angehängt. Diese lauten wie folgt: A (*author*), C (*classification*), I (*issn*), L (*location*), M (*method*), P (*publisher*), Y (*pubyear*) und T (*subject*, bzw. *thesaurus*) für den GIRT-Korpus und A (*author*) und S (*source*) für iSearch. Für das Baselinesystem bedeutet die Entität natürlich kein entsprechendes Reranking nach dieser speziellen Entität, sondern, dass hier nach dieser Entität gefiltert wurde, um vergleichbare Dokumentmengen zu erhalten. So wurden für die Ergebnismenge von S-A nur die Dokumente für das Retrieval genutzt, die überhaupt einen Autor besitzen (Herausgeberbände fallen z. B. durch das Raster).

### 10.3.1 Reranking mit Autorennamen in GIRT

Im Vergleich zu den fünf Systemen in Tabelle 10.4 können die Rerankingverfahren nicht an die Baseline S-A heranreichen, die in allen dargestellten Verfahren bessere Retrievalwerte liefert. Innerhalb der Rerankingverfahren ist B<sub>4</sub> das beste Verfahren, liegt aber fast 25% unter dem vergleichbaren MAP-Wert der Baseline (B<sub>4</sub>R-A mit 0,249 gegen S-A mit 0,330). Auch die anderen Evaluationsparameter RP, bpref und die Cut-off-Precision zeigen ein ähnliches Verhalten: Alle alternativen Rankingverfahren liegen unter diesen Werten. Sortiert nach den erreichten MAP-Werten belegen die Verfahren B<sub>2</sub>, B<sub>1</sub> und B<sub>E</sub> die Plätze 3, 4 bzw. 5.

run	R	MAP	RP	bpref	P@5	P@10	P@20	P@100
S-A	0,749	<b>0,330</b>	<b>0,358</b>	<b>0,320</b>	<b>0,631</b>	<b>0,571</b>	<b>0,497</b>	<b>0,337</b>
B <sub>1</sub> R-A	0,748	0,222	0,271	0,217	0,251	0,283	0,291	0,284
B <sub>2</sub> R-A	0,748	0,237	0,280	0,228	0,341	0,346	0,336	0,296
B <sub>4</sub> R-A	0,748	0,249	0,283	0,235	0,412	0,385	0,361	0,302
B <sub>E</sub> R-A	0,748	0,196	0,241	0,183	0,227	0,249	0,249	0,248

Tabelle 10.4: Retrievalergebnisse für ein Reranking mit Autorennamen für  $D_t$ .

Vergleicht man die einzelnen Ergebnisse hinsichtlich ihrer Sortierung mittels Kendalls  $\tau$  (s. Abschnitt 5.3) in Tabelle 10.5, so sieht man eine geringe Abweichung der drei einfachen Boostverfahren B<sub>1</sub>, B<sub>2</sub> und B<sub>4</sub>. Die Werte liegen hier jeweils über 0,864, die Unterschiede sind nur gering. Größer werden die Unterschiede zwischen der Baseline S-A und dem vierten Boost B<sub>E</sub>. Hier sind die Unterschiede zwischen B<sub>4</sub> und S-A mit 0,383 am größten, wobei die Unterschiede zu den anderen Boostverfahren nur gering sind. Im Vergleich zur Baseline liefern alle vier Rerankingvarianten eine anders sortierte Ergebnisliste.

Wendet man nun die beschriebenen Filter nach  $p$ - und  $\hat{\alpha}$ -Werten an, so zeigen sich deutliche Veränderungen in den Retrievalleistungen der einzelnen Systeme (s. Tab. 10.6).

	B <sub>1</sub> R-A	B <sub>2</sub> R-A	B <sub>4</sub> R-A	B <sub>E</sub> R-A	S-A
B <sub>1</sub> R-A	1	0,922	0,864	0,609	0,388
B <sub>2</sub> R-A		1	0,943	0,530	0,384
B <sub>4</sub> R-A			1	0,473	0,383
B <sub>E</sub> R-A				1	0,393
S-A					1

Tabelle 10.5: Vergleich der gerankten Ergebnislisten der Systeme B<sub>1</sub>R-A, B<sub>2</sub>R-A, B<sub>4</sub>R-A, B<sub>E</sub>R-A und S-A mittels Kendalls  $\tau$ .

Durch die Filterung verbessern sich alle dargestellten Werte von MAP, RP, bpref und P@10 (auf die Darstellung der weiteren Werte wurde aus Platzgründen verzichtet). Die Filterung durch die  $\hat{\alpha}$ -Werte bzw. die Kombination beider Kriterien führt darüber hinaus zu einer weiteren Verbesserung der Werte, die gegenüber der Baseline signifikant bzw. schwach signifikant ist.

Im Detail können für das Verfahren B<sub>4</sub>R-A <sub>$p\hat{\alpha}$</sub>  Retrievalleistungen erreicht werden, die zwar nicht vollständig, aber doch sehr nah an die Baseline S-A heranreichen. Die anderen Boost-Implementationen B<sub>1</sub>, B<sub>2</sub> und B<sub>E</sub> sind alle deutlich schwächer als die Referenz S-A. Konkret ist so z. B. der MAP-Wert von 0,313 sehr nah an der Baseline mit 0,330, bzw. die RP- und bpref-Werte von 0,348 und 0,303 noch näher an den Baselinewerten von 0,358 bzw. 0,320. Gleichzeitig ist aber bei der Filterung nach  $p$ - und  $\hat{\alpha}$ -Werten eine Verringerung der genutzten Topics von ehemals 150 Topics auf nur noch 62 zu verzeichnen. 88 Topics (58 %) mussten durch den Filtermechanismus aus dem Verfahren ausgeschlossen werden. Dieses Muster ist bei allen vier Boost-Implementationen gleich: Eine schärfere Filterung mittels  $p$  und  $\hat{\alpha}$  führt bei allen zu einem besseren Ergebnis als eine ungefilterte Implementation.

### 10.3.2 Reranking mit Klassifikationscodes in GIRT

Bei einem Reranking mit Klassifikationscodes (s. Tab. 10.7) fällt zunächst der unterschiedliche Recallwert R zwischen der Referenzimplementation S-C und den vier Rerankings auf. Dies ist auf eine technische Beschränkung der Software `trec_eval` zurückzuführen, die maximal 1 000 Dokumente pro Topic verarbeitet. Da für einige Topics mehr als 1 000 Dokumente zurückgeliefert wurden, wurde durch das alternative Ranking neben den unterschiedlichen Precisionberechnungen auch der Recall beeinflusst. Alle Rerankingimplementationen liegen ca. 10 Prozentpunkte unter dem sehr hohen Recallwert von 0,884 für S-C.

Die Rerankingverfahren unterscheiden sich untereinander nur schwach, sowohl hinsichtlich des Recalls als auch der MAP-, RP-, bpref- und Cut-off-Precisionwerte. Einzig B<sub>4</sub>R-C ist etwas schlechter, als die anderen drei Rerankingverfahren. Insgesamt sind aber alle Rerankings klar unter der Baseline S-C.

Vergleicht man die einzelnen Ergebnisse hinsichtlich ihrer Sortierung mittels Kendalls  $\tau$  in Tabelle 10.8, so sieht man eine geringe Abweichung der drei einfachen Boostverfahren B<sub>1</sub>,

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>1</sub> R-A	-	-	150	0,222	0,272	0,217	0,283
B <sub>1</sub> R-A <sub>p</sub>	✓	-	132	0,234	0,286	0,231	0,301
B <sub>1</sub> R-A <sub><math>\hat{\alpha}</math></sub>	-	✓	64	0,278 <sup>*</sup>	0,329 <sup>(*)</sup>	0,279 <sup>*</sup>	0,347
B <sub>1</sub> R-A <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	62	<b>0,284<sup>*</sup></b>	<b>0,337<sup>*</sup></b>	<b>0,287<sup>*</sup></b>	<b>0,357<sup>(*)</sup></b>

(a) Ergebnisse der Filterung für das Boostverfahren B<sub>1</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>2</sub> R-A	-	-	150	0,237	0,280	0,228	0,346
B <sub>2</sub> R-A <sub>p</sub>	✓	-	132	0,251	0,296	0,243	0,367
B <sub>2</sub> R-A <sub><math>\hat{\alpha}</math></sub>	-	✓	64	0,296 <sup>*</sup>	0,339 <sup>(*)</sup>	0,290 <sup>(*)</sup>	0,397
B <sub>2</sub> R-A <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	62	<b>0,302<sup>*</sup></b>	<b>0,346<sup>*</sup></b>	<b>0,297<sup>*</sup></b>	<b>0,408</b>

(b) Ergebnisse der Filterung für das Boostverfahren B<sub>2</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>4</sub> R-A	-	-	150	0,249	0,283	0,235	0,385
B <sub>4</sub> R-A <sub>p</sub>	✓	-	132	0,263	0,298	0,250	0,404
B <sub>4</sub> R-A <sub><math>\hat{\alpha}</math></sub>	-	✓	64	0,306 <sup>(*)</sup>	0,341 <sup>(*)</sup>	0,296 <sup>(*)</sup>	0,425
B <sub>4</sub> R-A <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	62	<b>0,313<sup>*</sup></b>	<b>0,348<sup>*</sup></b>	<b>0,303<sup>*</sup></b>	<b>0,437</b>

(c) Ergebnisse der Filterung für das Boostverfahren B<sub>4</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>E</sub> R-A	-	-	150	0,196	0,241	0,183	0,249
B <sub>E</sub> R-A <sub>p</sub>	✓	-	132	0,206	0,253	0,194	0,265
B <sub>E</sub> R-A <sub><math>\hat{\alpha}</math></sub>	-	✓	64	0,240 <sup>(*)</sup>	0,291 <sup>(*)</sup>	0,232 <sup>(*)</sup>	0,300
B <sub>E</sub> R-A <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	62	<b>0,245<sup>(*)</sup></b>	<b>0,298<sup>(*)</sup></b>	<b>0,238<sup>*</sup></b>	<b>0,308</b>

(d) Ergebnisse der Filterung für das Boostverfahren B<sub>E</sub>Tabelle 10.6: Retrievalergebnisse für ein Reranking mit Autorennamen nach einer Filterung nach  $p \geq 0,1$ ,  $\hat{\alpha}_i \leq \varnothing \hat{\alpha}$  bzw. deren Verknüpfung.

run	R	MAP	RP	bpref	P@5	P@10	P@20	P@100
S-C	0,884	<b>0,389</b>	<b>0,387</b>	<b>0,358</b>	<b>0,643</b>	<b>0,594</b>	<b>0,525</b>	<b>0,356</b>
B <sub>1</sub> R-C	0,788	0,277	0,307	0,264	0,389	0,386	0,378	0,314
B <sub>2</sub> R-C	0,788	0,276	0,303	0,261	0,392	0,384	0,372	0,310
B <sub>4</sub> R-C	0,788	0,264	0,292	0,250	0,359	0,353	0,346	0,303
B <sub>E</sub> R-C	0,778	0,274	0,309	0,261	0,408	0,405	0,368	0,312

Tabelle 10.7: Retrievalergebnisse für ein Reranking mit Klassifikationscodes für  $D_t$ .

	B <sub>1</sub> R-C	B <sub>2</sub> R-C	B <sub>4</sub> R-C	B <sub>E</sub> R-C	S-C
B <sub>1</sub> R-C	1	0,926	0,832	0,795	0,080
B <sub>2</sub> R-C		1	0,905	0,745	0,069
B <sub>4</sub> R-C			1	0,668	0,019
B <sub>E</sub> R-C				1	0,092
S-C					1

Tabelle 10.8: Vergleich der gerankten Ergebnislisten der Systeme B<sub>1</sub>R-C, B<sub>2</sub>R-C, B<sub>4</sub>R-C, B<sub>E</sub>R-C und S-C mittels Kendalls  $\tau$ .

B<sub>2</sub> und B<sub>4</sub>. Die Werte sind hier jeweils über 0,832, die Unterschiede sind nur gering. Etwas größer werden die Unterschiede zwischen drei einfachen und dem vierten Boostverfahren B<sub>E</sub>; diese liegen zwischen 0,668 und 0,795. Am größten sind die Unterschiede zwischen der Baseline S-C und den vier Rerankings. Hier sind die Unterschiede zwischen B<sub>4</sub> und S-C mit 0,019 am größten, wobei die Abweichungen zu den anderen Boostverfahren nur gering ist. Auch hier liegt im Vergleich zu S-C kein Wert über 0,092. Diese Werte sind sehr nah am Kendallwert 0, der aussagt, dass die Sortierung der Systeme zueinander nahezu zufällig ist. Im Vergleich zur Baseline liefern alle vier Rerankvarianten eine nahezu vollständig anders sortierte Ergebnisliste.

Wendet man die  $p$ - und  $\hat{\alpha}$ -Filter an (s. Tab. 10.9), so erhält man eine erheblich verkleinerte Menge an verbleibenden Topics. Bleiben beim  $p$ -Filter 77 und beim  $\hat{\alpha}$ -Filter noch 90 Topics, so sind in der kombinierten Variante von ehemals 150 Topics nur noch 33 vorhanden. Dies entspricht nur noch 22% der ursprünglichen Topicanzahl, es wurden nahezu 4/5 aller Topics ausgefiltert. Bis auf einen einzigen Fall (RP für B<sub>E</sub>R-C <sub>$p\hat{\alpha}$</sub> ) führt diese strenge Filterung aber nicht zu den besten Ergebnissen, vielmehr ist für nahezu alle Fälle eine Filterung nach dem  $p$ -Wert für die besten Retrievalergebnisse verantwortlich.

Während für MAP, RP und bpref die unterschiedlichen Boostverfahren keine allzu große Auswirkung haben, sind die Unterschiede für P@10 vergleichsweise hoch. So kann B<sub>4</sub>R-C <sub>$p$</sub>  nur einen P@10-Wert von 0,373 erreichen, die beste Implementation B<sub>E</sub>R-C <sub>$p$</sub>  hingegen ist hier mit einem P@10-Wert von 0,456 fast um 8 Prozentpunkte besser. Insgesamt sind aber alle gemessenen Systemkonfigurationen nicht als statistisch signifikant unterschiedlich einzustufen. Vergleicht man die besten gefilterten Runs mit der Baseline S-C so zeigt sich, dass die Baseline nicht erreicht bzw. übertroffen werden kann. Das im Vergleich



	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>1</sub> R-C	-	-	150	0,277	0,307	0,264	0,386
B <sub>1</sub> R-C <sub>p</sub>	✓	-	77	<b>0,314</b>	<b>0,351</b>	<b>0,306</b>	<b>0,416</b>
B <sub>1</sub> R-C <sub><math>\hat{\alpha}</math></sub>	-	✓	90	0,275	0,307	0,259	0,393
B <sub>1</sub> R-C <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	36	0,308	0,351	0,294	0,406

(a) Ergebnisse der Filterung für das Boostverfahren B<sub>1</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>2</sub> R-C	-	-	150	0,276	0,303	0,261	0,384
B <sub>2</sub> R-C <sub>p</sub>	✓	-	77	<b>0,311</b>	<b>0,344</b>	<b>0,303</b>	<b>0,409</b>
B <sub>2</sub> R-C <sub><math>\hat{\alpha}</math></sub>	-	✓	90	0,273	0,298	0,253	0,393
B <sub>2</sub> R-C <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	36	0,304	0,339	0,288	0,406

(b) Ergebnisse der Filterung für das Boostverfahren B<sub>2</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>4</sub> R-C	-	-	150	0,264	0,292	0,250	0,353
B <sub>4</sub> R-C <sub>p</sub>	✓	-	77	<b>0,295</b>	<b>0,327</b>	<b>0,287</b>	<b>0,373</b>
B <sub>4</sub> R-C <sub><math>\hat{\alpha}</math></sub>	-	✓	90	0,257	0,287	0,239	0,330
B <sub>4</sub> R-C <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	36	0,279	0,317	0,265	0,311

(c) Ergebnisse der Filterung für das Boostverfahren B<sub>4</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>E</sub> R-C	-	-	150	0,274	0,309	0,261	0,405
B <sub>E</sub> R-C <sub>p</sub>	✓	-	77	<b>0,311</b>	0,356	<b>0,308</b>	<b>0,456</b>
B <sub>E</sub> R-C <sub><math>\hat{\alpha}</math></sub>	-	✓	90	0,271	0,308	0,251	0,402
B <sub>E</sub> R-C <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	36	0,300	<b>0,357</b>	0,291	0,439

(d) Ergebnisse der Filterung für das Boostverfahren B<sub>E</sub>Tabelle 10.9: Retrievalergebnisse für ein Reranking mit Klassifikationscodes nach einer Filterung nach  $p \geq 0,1$ ,  $\hat{\alpha}_i \leq \varnothing \hat{\alpha}$  bzw. deren Verknüpfung.

der Boostverfahren beste Verfahren  $B_4R-C_p$  bleibt sowohl bei MAP ( $0,311 < 0,389$ ), RP ( $0,356 < 0,387$ ), bpref ( $0,308 < 0,387$ ) als auch bei P@10 ( $0,456 < 0,594$ ) klar unter den Baselinewerten.

### 10.3.3 Reranking mit ISSN-Codes in GIRT

Die Leistung der vier ungefilterten Rerankingverfahren, die mit den ISSN-Codes arbeiten, liegt klar unterhalb der Leistung der Baseline S-I (s. Tab. 10.10). So kann selbst die beste Implementation  $B_{ER-I}$  nur einen MAP-Wert von 0,112 erreichen, was klar unterhalb des Wertes des Baselinesystems von 0,140 liegt. Auch für die anderen Leistungszahlen RP, bpref und Cut-off-Precision sind die Unterschiede zwischen den Rerankingverfahren und der Baseline teilweise sehr deutlich, so liegt z. B. der beste P@5-Wert eines Rerankingverfahrens bei 0,385, der der Baseline jedoch bei 0,533. Die Rerankingverfahren unterscheiden sich untereinander nur sehr wenig. Das jeweilige Boostverfahren scheint nur eine geringe Auswirkung auf die Retrievalleistung zu haben.

Der Recallwert ist bei allen vier Rerankingverfahren und auch bei der Baseline S-I mit 0,290 sehr niedrig. Es wurden durch die Einschränkung auf Zeitschriftenartikel viele relevante Dokumente aus den Ergebnismengen entfernt. Dies hat auch direkte Auswirkungen auf die im Vergleich zu den anderen Verfahren niedrigen MAP-, RP-, bpref- und Cut-off-Precisionwerte.

run	R	MAP	RP	bpref	P@5	P@10	P@20	P@100
S-I	0,290	<b>0,140</b>	<b>0,211</b>	<b>0,175</b>	<b>0,533</b>	<b>0,468</b>	<b>0,399</b>	<b>0,244</b>
$B_1R-I$	0,290	<i>0,111</i>	<i>0,190</i>	<i>0,149</i>	0,385	<i>0,342</i>	<i>0,313</i>	<i>0,237</i>
$B_2R-I$	0,290	0,106	0,186	0,144	<i>0,363</i>	0,320	0,300	0,234
$B_4R-I$	0,290	0,104	0,185	0,142	0,355	0,313	0,294	0,233
$B_{ER-I}$	0,290	0,098	0,190	0,145	0,283	0,296	0,292	0,234

Tabelle 10.10: Retrievalergebnisse für ein Reranking mit ISSN-Codes für  $D_t$ .

	$B_1R-I$	$B_2R-I$	$B_4R-I$	$B_{ER-I}$	S-I
$B_1R-I$	1	0,986	0,981	0,879	0,836
$B_2R-I$		1	0,995	0,866	0,830
$B_4R-I$			1	0,861	0,828
$B_{ER-I}$				1	0,837
S-I					1

Tabelle 10.11: Vergleich der gerankten Ergebnislisten der Systeme  $B_1R-I$ ,  $B_2R-I$ ,  $B_4R-I$ ,  $B_{ER-I}$  und S-I mittels Kendalls  $\tau$ .

Die Analyse der  $\tau$ -Werte in Tabelle 10.11 deutet ebenfalls auf einen geringen Einfluss der Rerankingverfahren hin. Die einfachen Boostverfahren  $B_1$ ,  $B_2$  und  $B_4$  unterscheiden sich nahezu kaum voneinander und die Rankings sind mit  $\tau$ -Werten zwischen 0,981 und 0,995 als fast identisch anzusehen. Auch ein Vergleich zum Boostverfahren  $B_E$  zeigt zwar eine leichte Veränderung des Rankings, diese ist aber sehr moderat. Gleiches gilt für den Vergleich zum Baselinesystem S-I. Hier ist der größte Unterschied zwischen S-I und  $B_4R-I$  mit einem  $\tau$ -Wert von 0,828 sehr gering.

Die Filterung nach  $p$  hat beim Reranking mit ISSN-Codes nahezu keinerlei Auswirkungen auf die Retrievalleistung (s. Tab. 10.12). Weder werden die Werte messbar besser oder schlechter, noch wird eine hohe Zahl an Topics gefiltert. So sind die ermittelten Power-Laws in nur 7 von 150 Topics keine wirklichen Power-Laws. Wendet man allerdings den  $\hat{\alpha}$ -Filter an, so bleiben nur noch 81 Topics übrig. Immerhin sind dies noch 54 % der ursprünglichen Topicanzahl. Hier kann für alle Systeme eine statistisch signifikante Steigerung der MAP-Werte festgestellt werden.

Kombiniert man die beiden Filter miteinander, so sieht man, dass sie sich nur in einem Topic unterscheiden. Der Unterschied der beiden Filterungen  $\hat{\alpha}$  und  $p\hat{\alpha}$  ist folglich auch sehr gering. In allen Fällen liegen die Werte der mit  $p\hat{\alpha}$  gefilterten Systeme nur sehr knapp unter den Systemen, die mit  $\hat{\alpha}$  gefiltert wurden. Die Unterschiede zur ungefilterten Implementation sind zusätzlich für nahezu alle Varianten (ausgenommen  $B_{ER-I_{p\hat{\alpha}}}$ ) statistisch signifikant. Für  $B_2R-I_{\hat{\alpha}}$  sind auch die Werte RP und bpref signifikant unterschiedlich zur Baseline, dies bleibt aber die Ausnahme.

Die besten Ergebnisse erzielt das Boostverfahren  $B_1$ , wobei die Unterschiede zu  $B_2$  und  $B_4$  gering sind. Das Boostverfahren  $B_E$  liefert das tendenziell schlechteste Ergebnis. Vergleicht man die Systemleistung des besten gefilterten Systems  $B_1R-I_{\hat{\alpha}}$  mit dem der Baseline S-I, so übertrifft die gefilterte Variante die Baseline für RP und bpref. Die MAP-Werte sind mit 0,139 nur sehr knapp unter der Baseline von 0,140. Einzig bei P@10 sind die Unterschiede mit 0,398 und 0,468 zu Ungunsten des gefilterten Rerankingverfahrens.

10 Auswertung der Experimente zum Reranking

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>1</sub> R-I	-	-	150	0,111	0,190	0,149	0,342
B <sub>1</sub> R-I <sub>p</sub>	✓	-	143	0,112	0,190	0,148	0,346
B <sub>1</sub> R-I $\hat{\alpha}$	-	✓	81	<b>0,139<sup>(*)</sup></b>	<b>0,221</b>	<b>0,176</b>	<b>0,398</b>
B <sub>1</sub> R-I <sub>p</sub> $\hat{\alpha}$	✓	✓	80	<i>0,137<sup>(*)</sup></i>	<i>0,218</i>	<i>0,173</i>	<i>0,393</i>

(a) Ergebnisse der Filterung für das Boostverfahren B<sub>1</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>2</sub> R-I	-	-	150	0,106	0,186	0,144	0,320
B <sub>2</sub> R-I <sub>p</sub>	✓	-	143	0,106	0,186	0,143	0,323
B <sub>2</sub> R-I $\hat{\alpha}$	-	✓	81	<b>0,135<sup>(*)</sup></b>	<b>0,218<sup>(*)</sup></b>	<b>0,172<sup>(*)</sup></b>	<b>0,377</b>
B <sub>2</sub> R-I <sub>p</sub> $\hat{\alpha}$	✓	✓	80	<i>0,133<sup>(*)</sup></i>	<i>0,216</i>	<i>0,169</i>	<i>0,371</i>

(b) Ergebnisse der Filterung für das Boostverfahren B<sub>2</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>4</sub> R-I	-	-	150	0,104	0,185	0,142	0,313
B <sub>4</sub> R-I <sub>p</sub>	✓	-	143	0,104	0,184	0,141	0,315
B <sub>4</sub> R-I $\hat{\alpha}$	-	✓	81	<b>0,132<sup>(*)</sup></b>	<b>0,216</b>	<b>0,169</b>	<b>0,365</b>
B <sub>4</sub> R-I <sub>p</sub> $\hat{\alpha}$	✓	✓	80	<i>0,130<sup>(*)</sup></i>	<i>0,213</i>	<i>0,167</i>	<i>0,360</i>

(c) Ergebnisse der Filterung für das Boostverfahren B<sub>4</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>E</sub> R-I	-	-	150	0,098	0,190	0,145	0,296
B <sub>E</sub> R-I <sub>p</sub>	✓	-	143	0,098	0,189	0,144	0,297
B <sub>E</sub> R-I $\hat{\alpha}$	-	✓	81	<b>0,123<sup>(*)</sup></b>	<b>0,219</b>	<b>0,170</b>	<b>0,337</b>
B <sub>E</sub> R-I <sub>p</sub> $\hat{\alpha}$	✓	✓	80	<i>0,121</i>	<i>0,216</i>	<i>0,167</i>	<i>0,334</i>

(d) Ergebnisse der Filterung für das Boostverfahren B<sub>E</sub>

Tabelle 10.12: Retrievalergebnisse für ein Reranking mit ISSN-Codes nach einer Filterung nach  $p \geq 0,1$ ,  $\hat{\alpha}_i \leq \varnothing \hat{\alpha}$  bzw. deren Verknüpfung.

### 10.3.4 Reranking mit Ortsangaben in GIRT

Ein Reranking unter Zuhilfenahme der Ortsangaben führt zu durchgängig schlechteren Retrievalleistungen, unabhängig davon welches Boostverfahren eingesetzt wird (s. Tab. 10.13). Da nur bei wenigen Dokumenten eine Ortsangaben hinterlegt ist, ist der Recallwert mit 0,375 recht gering. Auch der MAP-Wert von 0,164 für die Baseline S-L ist relativ niedrig. Die Boostverfahren liegen sehr dicht beieinander, einzig das Verfahren  $B_{ER-L}$  liegt noch unter den geringen Werten von  $B_1$ ,  $B_2$  und  $B_4$ .

run	R	MAP	RP	bpref	P@5	P@10	P@20	P@100
S-L	0,375	<b>0,164</b>	<b>0,236</b>	<b>0,194</b>	<b>0,577</b>	<b>0,509</b>	<b>0,427</b>	<b>0,251</b>
$B_1R-L$	0,375	0,107	0,187	0,137	0,397	0,316	0,277	0,211
$B_2R-L$	0,375	0,105	0,185	0,134	0,408	0,321	0,267	0,205
$B_4R-L$	0,375	0,100	0,183	0,132	0,363	0,287	0,243	0,200
$B_{ER-L}$	0,375	0,091	0,184	0,130	0,229	0,223	0,230	0,214

Tabelle 10.13: Retrievalergebnisse für ein Reranking mit Ortsangaben für  $D_t$ .

	$B_1R-L$	$B_2R-L$	$B_4R-L$	$B_{ER-L}$	S-L
$B_1R-L$	1	0,966	0,951	0,679	0,749
$B_2R-L$		1	0,985	0,647	0,738
$B_4R-L$			1	0,633	0,731
$B_{ER-L}$				1	0,754
S-L					1

Tabelle 10.14: Vergleich der gerankten Ergebnislisten der Systeme  $B_1R-L$ ,  $B_2R-L$ ,  $B_4R-L$ ,  $B_{ER-L}$  und S-L mittels Kendalls  $\tau$ .

Die Kendallanalyse zeigt nahezu keinerlei Unterschiede zwischen den Boostverfahren  $B_1$ ,  $B_2$  und  $B_4$ . Die Werte liegen hier alle über 0,951. Das Verfahren  $B_{ER-L}$  liefert das zu den anderen Boostverfahren unterschiedlichste Ranking mit  $\tau$ -Werten zwischen 0,633 und 0,679. Im Vergleich mit der Baseline unterscheiden sich die Rerankings wiederum schwach mit  $\tau$ -Werten zwischen 0,731 und 0,754. Das Reranking nach der Ortsangabe führt zu einer nur leicht anders sortierten Liste im Vergleich zur Baseline.

Die Filterung nach  $p$  und  $\hat{\alpha}$  führt zu kaum messbaren Unterschieden. Obwohl für  $\hat{\alpha}$  und die Kombinationsfilterung von  $p$  und  $\hat{\alpha}$  von ehemals 150 Topics nur 94 bzw. 92 Topics übrig bleiben, variieren die Retrievalwerte kaum und werden für das Boostverfahren  $B_E$  auch schlechter als die ungefilterte Baseline. Zwar sind auch sehr geringe Steigerungen für MAP, RP, bpref und die Cut-off-Precision zu sehen, aber diese sind so gering, dass hier keinerlei statistische Signifikanz ermittelt werden konnte.

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>1</sub> R-L	-	-	150	0,107	0,187	0,137	0,316
B <sub>1</sub> R-L <sub>p</sub>	✓	-	147	<b>0,108</b>	<b>0,188</b>	<b>0,138</b>	<b>0,318</b>
B <sub>1</sub> R-L <sub><math>\hat{\alpha}</math></sub>	-	✓	94	0,105	0,186	0,133	0,303
B <sub>1</sub> R-L <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	92	0,106	0,188	0,135	0,309

(a) Ergebnisse der Filterung für das Boostverfahren B<sub>1</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>2</sub> R-L	-	-	150	0,105	0,185	0,134	0,321
B <sub>2</sub> R-L <sub>p</sub>	✓	-	147	0,105	0,185	<b>0,135</b>	0,322
B <sub>2</sub> R-L <sub><math>\hat{\alpha}</math></sub>	-	✓	94	0,106	0,185	0,133	0,319
B <sub>2</sub> R-L <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	92	<b>0,106</b>	<b>0,187</b>	0,134	<b>0,323</b>

(b) Ergebnisse der Filterung für das Boostverfahren B<sub>2</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>4</sub> R-L	-	-	150	0,100	0,183	0,132	0,287
B <sub>4</sub> R-L <sub>p</sub>	✓	-	147	<b>0,101</b>	0,183	<b>0,132</b>	<b>0,287</b>
B <sub>4</sub> R-L <sub><math>\hat{\alpha}</math></sub>	-	✓	94	0,100	0,182	0,129	0,280
B <sub>4</sub> R-L <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	92	0,100	<b>0,184</b>	0,130	0,283

(c) Ergebnisse der Filterung für das Boostverfahren B<sub>4</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>E</sub> R-L	-	-	150	0,091	0,184	0,130	0,223
B <sub>E</sub> R-L <sub>p</sub>	✓	-	147	<b>0,092</b>	<b>0,184</b>	<b>0,130</b>	<b>0,223</b>
B <sub>E</sub> R-L <sub><math>\hat{\alpha}</math></sub>	-	✓	94	0,088	0,177	0,122	0,214
B <sub>E</sub> R-L <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	92	0,089	0,179	0,124	0,215

(d) Ergebnisse der Filterung für das Boostverfahren B<sub>E</sub>Tabelle 10.15: Retrievalergebnisse für ein Reranking mit Ortsangaben nach einer Filterung nach  $p \geq 0,1$ ,  $\hat{\alpha}_i \leq \varnothing \hat{\alpha}$  bzw. deren Verknüpfung.

### 10.3.5 Reranking mit Forschungsmethoden in GIRT

Das Reranking mittels der jeweiligen Forschungsmethode kann im Vergleich zur Baseline S-M keine besseren Ergebnisse erzielen (s. Tab. 10.16), wie bereits bei den Ortsangaben sind die Ergebnisse schlechter und unterscheiden sich untereinander kaum. So liegen alle Rerankingvarianten bei MAP-Werten um 0,230. Ein Wert, der deutlich unter der Baseline von 0,389 für S-M liegt.

run	R	MAP	RP	bpref	P@5	P@10	P@20	P@100
S-M	0,884	<b>0,389</b>	<b>0,387</b>	<b>0,358</b>	<b>0,643</b>	<b>0,594</b>	<b>0,525</b>	<b>0,356</b>
B <sub>1</sub> R-M	0,868	0,227	0,239	0,188	0,305	0,281	0,262	0,245
B <sub>2</sub> R-M	0,868	0,230	0,243	0,191	0,292	0,289	0,262	0,249
B <sub>4</sub> R-M	0,868	0,230	<i>0,246</i>	<i>0,193</i>	0,297	0,283	0,262	0,249
B <sub>E</sub> R-M	0,865	<i>0,230</i>	0,244	0,192	<i>0,319</i>	<i>0,311</i>	<i>0,294</i>	<i>0,254</i>

Tabelle 10.16: Retrievalergebnisse für ein Reranking mit Forschungsmethoden für  $D_t$ .

	B <sub>1</sub> R-M	B <sub>2</sub> R-M	B <sub>4</sub> R-M	B <sub>E</sub> R-M	S-M
B <sub>1</sub> R-M	1	0,848	0,736	0,679	0,070
B <sub>2</sub> R-M		1	0,883	0,543	0,066
B <sub>4</sub> R-M			1	0,444	0,042
B <sub>E</sub> R-M				1	0,088
S-M					1

Tabelle 10.17: Vergleich der gerankten Ergebnislisten der Systeme B<sub>1</sub>R-M, B<sub>2</sub>R-M, B<sub>4</sub>R-M, B<sub>E</sub>R-M und S-M mittels Kendalls  $\tau$ .

Die Analyseergebnisse von Kendalls  $\tau$  in Tabelle 10.17 zeigen, dass sich die drei Boostverfahren B<sub>1</sub>, B<sub>2</sub> und B<sub>4</sub> nur moderat voneinander unterscheiden. Der Unterschied zwischen B<sub>1</sub> und B<sub>4</sub> ist mit 0,736 am größten. Die Unterschiede steigen an, wenn B<sub>E</sub> als Boostverfahren eingesetzt wird, hier liegen die  $\tau$ -Werte zwischen 0,444 und 0,679. Vergleicht man die Boostverfahren mit der Baseline S-M so sieht man nahezu eine zufällige Sortierung der Systeme zueinander. Keiner der Vergleichswerte zwischen Baseline und den Boostverfahren liegt über 0,088.

Die in Tabelle 10.18 dokumentierte Filterung mittels  $p$  und  $\hat{\alpha}$  führt zu stark verkleinerten Dokumentmengen, so bleiben bei der kombinierten Filterung nur noch 63 Dokumente übrig. Allerdings bringt diese Filterung keine Steigerung der Retrievalleistung mit sich. In allen Boostvarianten führt sowohl die Filterung mittels  $p$ ,  $\hat{\alpha}$  als auch  $p\hat{\alpha}$  zu schlechteren Ergebnissen. Zwar sind die Ergebnisse nicht signifikant, doch ist die Verschlechterung eindeutig.

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>1</sub> R-M	-	-	150	<b>0,227</b>	<b>0,239</b>	<b>0,188</b>	<b>0,281</b>
B <sub>1</sub> R-M <sub>p</sub>	✓	-	126	0,219	0,231	0,181	0,275
B <sub>1</sub> R-M <sub><math>\hat{\alpha}</math></sub>	-	✓	85	0,224	0,238	0,187	0,274
B <sub>1</sub> R-M <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	63	0,212	0,226	0,178	0,265

(a) Ergebnisse der Filterung für das Boostverfahren B<sub>1</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>2</sub> R-M	-	-	150	<b>0,230</b>	<b>0,243</b>	<b>0,191</b>	<b>0,289</b>
B <sub>2</sub> R-M <sub>p</sub>	✓	-	126	0,221	0,233	0,184	0,284
B <sub>2</sub> R-M <sub><math>\hat{\alpha}</math></sub>	-	✓	85	0,226	0,242	0,190	0,286
B <sub>2</sub> R-M <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	63	0,214	0,230	0,181	0,284

(b) Ergebnisse der Filterung für das Boostverfahren B<sub>2</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>4</sub> R-M	-	-	150	<b>0,230</b>	<b>0,246</b>	<b>0,193</b>	<b>0,283</b>
B <sub>4</sub> R-M <sub>p</sub>	✓	-	126	0,221	0,237	0,186	0,279
B <sub>4</sub> R-M <sub><math>\hat{\alpha}</math></sub>	-	✓	85	0,225	0,241	0,189	0,275
B <sub>4</sub> R-M <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	63	0,213	0,230	0,180	0,273

(c) Ergebnisse der Filterung für das Boostverfahren B<sub>4</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>E</sub> R-M	-	-	150	<b>0,230</b>	<b>0,244</b>	<b>0,192</b>	<b>0,311</b>
B <sub>E</sub> R-M <sub>p</sub>	✓	-	126	0,222	0,235	0,185	0,310
B <sub>E</sub> R-M <sub><math>\hat{\alpha}</math></sub>	-	✓	85	0,227	0,241	0,191	0,311
B <sub>E</sub> R-M <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	63	0,215	0,230	0,182	0,311

(d) Ergebnisse der Filterung für das Boostverfahren B<sub>E</sub>Tabelle 10.18: Retrievalergebnisse für ein Reranking mit Forschungsmethoden nach einer Filterung nach  $p \geq 0,1$ ,  $\hat{\alpha}_i \leq \varnothing \hat{\alpha}$  bzw. deren Verknüpfung.



### 10.3.6 Reranking mit Verlagsnamen in GIRT

Das Reranking mittels des jeweiligen Verlags kann im Vergleich zur Baseline S-P keine besseren Ergebnisse erzielen (s. Tab. 10.19), wie bereits bei den Ortsangaben und Forschungsmethoden sind die Ergebnisse schlechter und unterscheiden sich untereinander kaum. So liegen alle Rerankingvarianten bei MAP-Werten von 0,137 bis 0,149 klar unter der Baseline von 0,172 für S-P. Auffällig ist ebenfalls ein niedriger Recallwert für alle Verfahren von 0,238. Verlagsinformationen sind nur bei wenigen Dokumenten vorhanden.

run	R	MAP	RP	bpref	P@5	P@10	P@20	P@100
S-P	0,238	<b>0,109</b>	<b>0,172</b>	<b>0,143</b>	<b>0,513</b>	<b>0,438</b>	<b>0,367</b>	<b>0,196</b>
B <sub>1</sub> R-P	0,238	0,073	0,149	0,110	0,329	0,285	0,252	0,183
B <sub>2</sub> R-P	0,238	0,066	0,144	0,104	0,299	0,256	0,237	0,180
B <sub>4</sub> R-P	0,238	0,066	0,141	0,103	0,289	0,247	0,231	0,179
B <sub>E</sub> R-P	0,238	0,060	0,137	0,099	0,227	0,225	0,223	0,179

Tabelle 10.19: Retrievalergebnisse für ein Reranking mit Verlagsnamen für  $D_t$ .

	B <sub>1</sub> R-P	B <sub>2</sub> R-P	B <sub>4</sub> R-P	B <sub>E</sub> R-P	S-P
B <sub>1</sub> R-P	1	0,975	0,970	0,821	0,853
B <sub>2</sub> R-P		1	0,995	0,800	0,846
B <sub>4</sub> R-P			1	0,795	0,843
B <sub>E</sub> R-P				1	0,844
S-P					1

Tabelle 10.20: Vergleich der gerankten Ergebnislisten der Systeme B<sub>1</sub>R-P, B<sub>2</sub>R-P, B<sub>4</sub>R-P, B<sub>E</sub>R-P und S-P mittels Kendalls  $\tau$ .

Die Analyse von Kendalls  $\tau$  in Tabelle 10.20 zeigt nahezu identische Sortierungen für die Boostverfahren B<sub>1</sub>, B<sub>2</sub> und B<sub>4</sub>. Hier liegen alle Werte über 0,970. Etwas größer sind die Unterschiede zu B<sub>E</sub> und der Baseline S-P, aber auch hier sind die  $\tau$ -Werte durchweg größer als 0,795. Es sind kaum Unterschiede in den Sortierungen festzustellen.

Die in Tabelle 10.21 dokumentierte Filterung mittels  $p$  und  $\hat{\alpha}$  führt zu moderat verkleinerten Dokumentmengen. So bleiben bei der kombinierten Filterung nur noch 86 Dokumente übrig, eine Filterung nach  $p$  entfernt sogar nur 5 Dokumente. Allerdings bringt diese Filterung keine Verbesserung oder Verschlechterung der Retrievalleistung mit sich. Die Werte sind teilweise bis auf die dritte Nachkommastelle gleich. Die Veränderungen sind daher auch nicht statistisch signifikant. Insgesamt kann kein Verfahren an die ursprüngliche Baseline S-P heranreichen.

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>1</sub> R-P	-	-	150	<b>0,073</b>	<b>0,149</b>	<b>0,110</b>	<b>0,285</b>
B <sub>1</sub> R-P <sub>p</sub>	✓	-	145	0,072	0,149	0,109	0,279
B <sub>1</sub> R-P <sub><math>\hat{\alpha}</math></sub>	-	✓	88	0,073	0,143	0,105	0,264
B <sub>1</sub> R-P <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	86	0,072	0,143	0,105	0,260

(a) Ergebnisse der Filterung für das Boostverfahren B<sub>1</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>2</sub> R-P	-	-	150	0,066	<b>0,144</b>	<b>0,104</b>	<b>0,256</b>
B <sub>2</sub> R-P <sub>p</sub>	✓	-	145	0,066	0,144	0,103	0,250
B <sub>2</sub> R-P <sub><math>\hat{\alpha}</math></sub>	-	✓	88	<b>0,068</b>	0,138	0,100	0,236
B <sub>2</sub> R-P <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	86	0,067	0,138	0,099	0,233

(b) Ergebnisse der Filterung für das Boostverfahren B<sub>2</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>4</sub> R-P	-	-	150	0,066	<b>0,141</b>	<b>0,103</b>	<b>0,247</b>
B <sub>4</sub> R-P <sub>p</sub>	✓	-	145	0,065	0,141	0,102	0,241
B <sub>4</sub> R-P <sub><math>\hat{\alpha}</math></sub>	-	✓	88	<b>0,067</b>	0,137	0,099	0,235
B <sub>4</sub> R-P <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	86	0,066	0,136	0,098	0,231

(c) Ergebnisse der Filterung für das Boostverfahren B<sub>4</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>E</sub> R-P	-	-	150	0,060	<b>0,137</b>	<b>0,099</b>	<b>0,225</b>
B <sub>E</sub> R-P <sub>p</sub>	✓	-	145	0,060	0,137	0,098	0,220
B <sub>E</sub> R-P <sub><math>\hat{\alpha}</math></sub>	-	✓	88	<b>0,061</b>	0,135	0,096	0,214
B <sub>E</sub> R-P <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	86	0,060	0,134	0,099	0,225

(d) Ergebnisse der Filterung für das Boostverfahren B<sub>E</sub>Tabelle 10.21: Retrievalergebnisse für ein Reranking mit Verlagsnamen nach einer Filterung nach  $p \geq 0,1$ ,  $\hat{\alpha}_i \leq \varnothing \hat{\alpha}$  bzw. deren Verknüpfung.

### 10.3.7 Reranking nach Erscheinungsjahr in GIRT

Beim Reranking mittels des Erscheinungsjahres fallen zunächst die hohen Recallwerte von 0,884 auf. Dies ist auf die komplette Abdeckung der Dokumente mit dieser Information zurückzuführen. Die verschiedenen Rerankingverfahren unterscheiden sich teilweise deutlich voneinander, auch wenn sie in ihrer Effektivität nicht an die Baseline S-Y heranreichen können. Die Baseline ist durchgehend besser als alle anderen Verfahren. Unter den Rerankingverfahren ist mit einem MAP-Wert von 0,300 das Verfahren B<sub>E</sub>R-Y das leistungsfähigste, bei den Cut-off-Precisionwerten ist dies das Verfahren B<sub>1</sub>R-Y.

run	R	MAP	RP	bpref	P@5	P@10	P@20	P@100
S-Y	0,884	<b>0,389</b>	<b>0,387</b>	<b>0,358</b>	<b>0,643</b>	<b>0,594</b>	<b>0,525</b>	<b>0,356</b>
B <sub>1</sub> R-Y	0,884	0,279	0,287	0,240	0,499	0,425	0,372	0,282
B <sub>2</sub> R-Y	0,884	0,261	0,263	0,218	0,471	0,397	0,354	0,270
B <sub>4</sub> R-Y	0,884	0,239	0,249	0,200	0,336	0,313	0,286	0,252
B <sub>E</sub> R-Y	0,884	0,300	0,328	0,281	0,304	0,341	0,360	0,308

Tabelle 10.22: Retrievalergebnisse für ein Reranking mit dem Erscheinungsjahr für  $D_t$ .

	B <sub>1</sub> R-Y	B <sub>2</sub> R-Y	B <sub>4</sub> R-Y	B <sub>E</sub> R-Y	S-Y
B <sub>1</sub> R-Y	1	0,900	0,822	0,304	0,199
B <sub>2</sub> R-Y		1	0,922	0,208	0,138
B <sub>4</sub> R-Y			1	0,132	0,081
B <sub>E</sub> R-Y				1	0,409
S-Y					1

Tabelle 10.23: Vergleich der gerankten Ergebnislisten der Systeme B<sub>1</sub>R-Y, B<sub>2</sub>R-Y, B<sub>4</sub>R-Y, B<sub>E</sub>R-Y und S-Y mittels Kendalls  $\tau$ .

Die Kendallanalyse zeigt nur geringe Unterschiede zwischen den Boostverfahren B<sub>1</sub>, B<sub>2</sub> und B<sub>4</sub>. Die Unterschiede zu B<sub>E</sub> und der Baseline S-Y sind allerdings größer. Das nach der Baseline S-Y zweitplatzierte System B<sub>E</sub>R-Y hat noch einen recht moderaten  $\tau$ -Wert von 0,409, die anderen Verfahren liegen mit 0,081 bis 0,199 klar darunter.

Die Filterung mit  $p$  und  $\hat{\alpha}$  für zu deutlich kleineren Dokumentmengen. Bei einer Filterung durch  $p$  bleiben 97, mit  $\hat{\alpha}$  nur 55 und bei einer Kombination gar nur 17 übrig. In einigen Fällen wie bei B<sub>2</sub>R-Y <sub>$p\hat{\alpha}$</sub>  oder B<sub>4</sub>R-Y <sub>$p$</sub>  kann diese Filterung zwar zu besseren Ergebnissen, als bei der Baseline führen, doch sind diese Steigerungen der Retrievalleistung nicht statistisch signifikant. Die Veränderungen sind auch in absoluten Zahlen nur sehr gering und sind betreffen meist nur die zweite oder dritte Nachkommastelle. War im Vergleich zur Gesamtbaseline S-Y das System B<sub>E</sub>R-Y noch relativ gut positioniert werden durch eine Filterung hier die Werte nur noch schlechter und unterschreiten teilweise sogar die anderen Boostverfahren, sowohl in gefilterter, wie ungefilterten Varianten.

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>1</sub> R-Y	-	-	150	<b>0,279</b>	<b>0,287</b>	<b>0,240</b>	<b>0,425</b>
B <sub>1</sub> R-Y <sub>p</sub>	✓	-	97	0,272	0,277	0,230	0,405
B <sub>1</sub> R-Y <sub><math>\hat{\alpha}</math></sub>	-	✓	55	0,272	0,273	0,238	0,449
B <sub>1</sub> R-Y <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	17	0,271	0,258	0,233	0,429

(a) Ergebnisse der Filterung für das Boostverfahren B<sub>1</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>2</sub> R-Y	-	-	150	0,261	<b>0,263</b>	0,218	0,397
B <sub>2</sub> R-Y <sub>p</sub>	✓	-	97	0,258	0,260	0,213	0,382
B <sub>2</sub> R-Y <sub><math>\hat{\alpha}</math></sub>	-	✓	55	0,252	0,247	0,212	0,420
B <sub>2</sub> R-Y <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	17	<b>0,264</b>	0,259	<b>0,226</b>	<b>0,435</b>

(b) Ergebnisse der Filterung für das Boostverfahren B<sub>2</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>4</sub> R-Y	-	-	150	0,239	0,249	0,200	<b>0,313</b>
B <sub>4</sub> R-Y <sub>p</sub>	✓	-	97	<b>0,239</b>	<b>0,251</b>	<b>0,201</b>	0,305
B <sub>4</sub> R-Y <sub><math>\hat{\alpha}</math></sub>	-	✓	55	0,222	0,224	0,185	0,302
B <sub>4</sub> R-Y <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	17	0,226	0,233	0,197	0,241

(c) Ergebnisse der Filterung für das Boostverfahren B<sub>4</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>E</sub> R-Y	-	-	150	<b>0,300</b>	<b>0,328</b>	<b>0,281</b>	0,341
B <sub>E</sub> R-Y <sub>p</sub>	✓	-	97	0,299	0,326	0,280	<b>0,359</b>
B <sub>E</sub> R-Y <sub><math>\hat{\alpha}</math></sub>	-	✓	55	0,267	0,299	0,251	0,284
B <sub>E</sub> R-Y <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	17	0,261	0,302	0,252	0,276

(d) Ergebnisse der Filterung für das Boostverfahren B<sub>E</sub>Tabelle 10.24: Retrievalergebnisse für ein Reranking mit dem Erscheinungsjahr nach einer Filterung nach  $p \geq 0,1$ ,  $\hat{\alpha}_i \leq \varnothing \hat{\alpha}$  bzw. deren Verknüpfung.

### 10.3.8 Reranking mit Thesaurustermen in GIRT

Zunächst fällt bei der Betrachtung der Rerankingverfahren mittels Thesaurustermen auf, dass ein einziger Parameter im ungefilterten Ansatz besser als die Baseline S-T ist (s. Tab. 10.25). Bei P@100 kann das Verfahren B<sub>E</sub>R-T den Wert von 0,356 mit 0,378 übertreffen. Auch im Gesamtvergleich ist das Verfahren B<sub>E</sub>R-T sehr nah an der Baseline und unterscheidet sich teilweise nur minimal in der Retrievalleistung von S-T. Die anderen drei Boostvarianten B<sub>1</sub>, B<sub>2</sub> und B<sub>4</sub> sind deutlich unterhalb der beiden Verfahren S-T und B<sub>E</sub>R-T. B<sub>1</sub> und B<sub>2</sub> sind dabei nahezu identisch und B<sub>4</sub> ist das schlechteste System. Alle Systeme haben relativ hohe Recallwerte von über 0,844.

run	R	MAP	RP	bpref	P@5	P@10	P@20	P@100
S-T	0,884	<b>0,389</b>	<b>0,387</b>	<b>0,358</b>	<b>0,643</b>	<b>0,594</b>	<b>0,525</b>	0,356
B <sub>1</sub> R-T	0,854	0,350	0,363	0,329	0,489	0,473	0,437	0,358
B <sub>2</sub> R-T	0,854	0,350	0,360	0,327	0,497	0,472	0,433	0,356
B <sub>4</sub> R-T	0,854	0,337	0,349	0,315	0,467	0,448	0,420	0,353
B <sub>E</sub> R-T	0,844	0,370	0,384	0,353	0,556	0,515	0,464	<b>0,378</b>

Tabelle 10.25: Retrievalergebnisse für ein Reranking mit Thesaurustermen für  $D_t$ .

	B <sub>1</sub> R-T	B <sub>2</sub> R-T	B <sub>4</sub> R-T	B <sub>E</sub> R-T	S-T
B <sub>1</sub> R-T	1	0,119	0,119	0,088	0,133
B <sub>2</sub> R-T		1	0,915	0,793	0,746
B <sub>4</sub> R-T			1	0,873	0,707
B <sub>E</sub> R-T				1	0,616
S-T					1

Tabelle 10.26: Vergleich der gerankten Ergebnislisten der Systeme B<sub>1</sub>R-T, B<sub>2</sub>R-T, B<sub>4</sub>R-T, B<sub>E</sub>R-T und S-T mittels Kendalls  $\tau$ .

Die Beobachtungen der Retrievalleistung lassen sich auch auf die Kendallanalyse in Tabelle 10.26 übertragen. Die Boostvariante B<sub>1</sub> liefert ein zu allen anderen Verfahren unterschiedliches Ranking mit  $\tau$ -Werten, die unter 0,133 liegen. Die anderen Verfahren liegen näher beieinander mit  $\tau$ -Werten über 0,616, wobei die Übereinstimmung der Rankings zwischen B<sub>2</sub> und B<sub>4</sub> mit 0,915 am größten und zwischen S-T und B<sub>E</sub> mit 0,616 am niedrigsten ist. Die Filterung mit  $p$  und  $\hat{\alpha}$  führt zu deutlichen Effekten bei der Leistungsfähigkeit der Rankingverfahren (s. Tab. 10.27). In allen Varianten ist die kombinierte Filterung  $p\hat{\alpha}$  die effektivste Variante, die deutlich vor den ungefilterten und den zweiplatzierten Varianten liegt. Die Steigerung ist für fast alle Varianten statistisch signifikant und kann für MAP, RP und bpref immer die Leistung des Baselinesystems S-T übertreffen. Die Filterung ist unterschiedlich streng. Bei der Filterung mit  $p$  werden mit 46 Topics ca. 1/3 aller Topics herausgefiltert, bei der kombinierten Variante bleibt mit 41 Topics nur weniger als 1/3 aller Topics übrig.

10 Auswertung der Experimente zum Reranking

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>1</sub> R-T	-	-	150	0,350	0,363	0,329	0,473
B <sub>1</sub> R-T <sub>p</sub>	✓	-	104	0,366	0,378	0,348	0,488
B <sub>1</sub> R-T <sub><math>\hat{\alpha}</math></sub>	-	✓	79	0,370	0,380	0,344	0,473
B <sub>1</sub> R-T <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	41	<b>0,435<sup>(*)</sup></b>	<b>0,439</b>	<b>0,412</b>	<b>0,539</b>

(a) Ergebnisse der Filterung für das Boostverfahren B<sub>1</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>2</sub> R-T	-	-	150	0,350	0,360	0,327	0,472
B <sub>2</sub> R-T <sub>p</sub>	✓	-	104	0,366	0,374	0,345	0,488
B <sub>2</sub> R-T <sub><math>\hat{\alpha}</math></sub>	-	✓	79	0,371	0,377	0,343	0,468
B <sub>2</sub> R-T <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	41	<b>0,437<sup>(*)</sup></b>	<b>0,437</b>	<b>0,412<sup>(*)</sup></b>	<b>0,529</b>

(b) Ergebnisse der Filterung für das Boostverfahren B<sub>2</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>4</sub> R-T	-	-	150	0,337	0,349	0,315	0,448
B <sub>4</sub> R-T <sub>p</sub>	✓	-	104	0,356	0,366	0,336	0,465
B <sub>4</sub> R-T <sub><math>\hat{\alpha}</math></sub>	-	✓	79	0,363	0,369	0,336	0,452
B <sub>4</sub> R-T <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	41	<b>0,435<sup>*</sup></b>	<b>0,438<sup>(*)</sup></b>	<b>0,411<sup>(*)</sup></b>	<b>0,522</b>

(c) Ergebnisse der Filterung für das Boostverfahren B<sub>4</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>E</sub> R-T	-	-	150	0,370	0,384	0,353	0,515
B <sub>E</sub> R-T <sub>p</sub>	✓	-	104	0,390	0,407	0,376	0,530
B <sub>E</sub> R-T <sub><math>\hat{\alpha}</math></sub>	-	✓	79	0,379	0,387	0,355	0,518
B <sub>E</sub> R-T <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	41	<b>0,445</b>	<b>0,451</b>	<b>0,422</b>	<b>0,568</b>

(d) Ergebnisse der Filterung für das Boostverfahren B<sub>E</sub>

Tabelle 10.27: Retrievalergebnisse für ein Reranking mit Thesaurustermen nach einer Filterung nach  $p \geq 0,1$ ,  $\hat{\alpha}_i \leq \varnothing \hat{\alpha}$  bzw. deren Verknüpfung.

### 10.3.9 Reranking mit Autorennamen in iSearch

Wie bei den bisherigen Beispielen aus dem GIRT-Korpus zeigt sich auch bei den iSearch-Beispielen, dass ein einfaches Reranking keine direkten Vorteile für MAP, RP, bpref oder die Cut-off-Precision bringt. Die Werte in Tabelle 10.28 liegen alle unter der Baseline S-A. Das zweitbeste System B<sub>4</sub>R-A liegt zwar ebenfalls unter der Baseline, führt aber deutlich vor den anderen Rerankingvarianten. Die Recallwerte sind sehr hoch und liegen für die Baseline fast bei 100 %.

run	R	MAP	RP	bpref	P@5	P@10	P@20	P@100
S-A	0,999	<b>0,412</b>	<b>0,372</b>	<b>0,314</b>	<b>0,535</b>	<b>0,472</b>	<b>0,397</b>	<b>0,277</b>
B <sub>1</sub> R-A	0,936	0,279	0,247	0,188	0,249	0,257	0,255	0,240
B <sub>2</sub> R-A	0,936	0,294	0,254	0,202	0,295	0,283	0,282	0,244
B <sub>4</sub> R-A	0,936	<i>0,310</i>	<i>0,280</i>	<i>0,221</i>	<i>0,323</i>	<i>0,317</i>	<i>0,307</i>	<i>0,248</i>
B <sub>E</sub> R-A	0,936	0,269	0,233	0,176	0,225	0,232	0,245	0,232

Tabelle 10.28: Retrievalergebnisse für ein Reranking mit dem Autorennamen für  $D_t$ .

	B <sub>1</sub> R-A	B <sub>2</sub> R-A	B <sub>4</sub> R-A	B <sub>E</sub> R-A	S-A
B <sub>1</sub> R-A	1	0,865	0,722	0,822	0,135
B <sub>2</sub> R-A		1	0,856	0,688	0,128
B <sub>4</sub> R-A			1	0,546	0,119
B <sub>E</sub> R-A				1	0,144
S-A					1

Tabelle 10.29: Vergleich der gerankten Ergebnislisten der Systeme B<sub>1</sub>R-A, B<sub>2</sub>R-A, B<sub>4</sub>R-A, B<sub>E</sub>R-A und S-A mittels Kendalls  $\tau$ .

Analog zu dem Autorenreranking im GIRT-Korpus unterscheiden sich die Rerankings untereinander nur moderat mit  $\tau$ -Werten zwischen 0,546 und 0,865 (s. Tab. 10.29). Innerhalb der Boostverfahren B<sub>1</sub>, B<sub>2</sub>, B<sub>4</sub> und B<sub>E</sub> sind die Systemleistungen von B<sub>E</sub> und B<sub>4</sub> am deutlichsten voneinander entfernt. Alle anderen Verfahren liegen recht nah beieinander. Vergleicht man die Rerankingansätze allerdings mit dem Baselinesystem S-A, so sieht man eine sehr starke Veränderung. Hier liegen die Werte alle unter 0,144, die Rankings sind nahezu vollständig unabhängig voneinander.

Die Filterung der Topics nach  $p$  und  $\hat{\alpha}$ , wie sie in Tabelle 10.30 zusammengefasst sind, führen zwar in absoluten Zahlen zu Verbesserungen, doch sind diese zu keiner Zeit statistisch signifikant. Über alle Boostverfahren hinweg zeigt sich die kombinierte Filterung nach  $p\hat{\alpha}$  als die effektivste Variante, mit Steigerungen der MAP-Werte auf bis zu 0,322 für das Boostverfahren B<sub>4</sub>. Allerdings ist die doppelte Filterung sehr streng und lässt von ehemals 65 Topics nur 10 übrig, sodass nur 15 % der Topics überhaupt verbleiben. Gleiches gilt für das zweitbeste Filterverfahren nach  $\hat{\alpha}$ , hier verbleiben nur 11 von 65 Topics. Filterung nach  $p$  zeigt kaum Effekte, lässt aber 41 Topics übrig. Insgesamt

## 10 Auswertung der Experimente zum Reranking

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>1</sub> R-A	-	-	65	0,279	0,247	0,188	<b>0,257</b>
B <sub>1</sub> R-A <sub>p</sub>	✓	-	41	0,279	0,245	0,188	0,246
B <sub>1</sub> R-A <sub><math>\hat{\alpha}</math></sub>	-	✓	11	0,288	0,266	0,200	0,236
B <sub>1</sub> R-A <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	10	<b>0,304</b>	<b>0,277</b>	<b>0,207</b>	0,240

(a) Ergebnisse der Filterung für das Boostverfahren B<sub>1</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>2</sub> R-A	-	-	65	0,294	0,254	0,202	0,283
B <sub>2</sub> R-A <sub>p</sub>	✓	-	41	0,299	0,257	0,205	<b>0,293</b>
B <sub>2</sub> R-A <sub><math>\hat{\alpha}</math></sub>	-	✓	11	0,296	0,248	0,200	0,236
B <sub>2</sub> R-A <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	10	<b>0,316</b>	<b>0,272</b>	<b>0,220</b>	0,260

(b) Ergebnisse der Filterung für das Boostverfahren B<sub>2</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>4</sub> R-A	-	-	65	0,310	0,280	0,221	0,317
B <sub>4</sub> R-A <sub>p</sub>	✓	-	41	0,311	<b>0,287</b>	0,221	<b>0,317</b>
B <sub>4</sub> R-A <sub><math>\hat{\alpha}</math></sub>	-	✓	11	0,300	0,260	0,211	0,264
B <sub>4</sub> R-A <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	10	<b>0,322</b>	0,286	<b>0,232</b>	0,290

(c) Ergebnisse der Filterung für das Boostverfahren B<sub>4</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>E</sub> R-A	-	-	65	0,269	0,233	0,176	0,232
B <sub>E</sub> R-A <sub>p</sub>	✓	-	41	0,267	0,228	0,172	0,220
B <sub>E</sub> R-A <sub><math>\hat{\alpha}</math></sub>	-	✓	11	0,287	0,254	0,196	0,264
B <sub>E</sub> R-A <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	10	<b>0,304</b>	<b>0,264</b>	<b>0,203</b>	<b>0,270</b>

(d) Ergebnisse der Filterung für das Boostverfahren B<sub>E</sub>

Tabelle 10.30: Retrievalergebnisse für ein Reranking mit dem Autorennamen nach einer Filterung nach  $p \geq 0,1$ ,  $\hat{\alpha}_i \leq \varnothing \hat{\alpha}$  bzw. deren Verknüpfung.



kann mit keinem Verfahren an die ursprüngliche Baseline S-A und deren vergleichsweise hohen MAP-Wert von 0,412 herangereicht werden.

### 10.3.10 Reranking mit Publikationsquellen in iSearch

Im Vergleich zum Reranking mit Autorennamen im iSearch-Korpus fällt beim Reranking nach den Publikationsquellen (*source*) in Tabelle 10.31 zunächst der niedrige Recallwert von 0,366 auf. Es liegt innerhalb der bewerteten Dokumente von iSearch nur in 1/3 der Fälle eine solche Quellenangabe vor.

run	R	MAP	RP	bpref	P@5	P@10	P@20	P@100
S-S	0,366	<b>0,152</b>	<b>0,204</b>	<b>0,161</b>	<b>0,382</b>	<b>0,337</b>	<b>0,289</b>	<b>0,168</b>
B <sub>1</sub> R-S	0,366	0,131	0,178	0,140	0,295	0,278	0,258	0,165
B <sub>2</sub> R-S	0,366	0,118	0,168	0,127	0,280	0,255	0,250	0,165
B <sub>4</sub> R-S	0,366	0,114	0,160	0,122	0,262	0,238	0,235	0,165
B <sub>E</sub> R-S	0,366	0,114	0,158	0,120	0,262	0,240	0,233	0,165

Tabelle 10.31: Retrievalergebnisse für ein Reranking mit der Publikationsquelle für  $D_t$ .

	B <sub>1</sub> R-S	B <sub>2</sub> R-S	B <sub>4</sub> R-S	B <sub>E</sub> R-S	S-S
B <sub>1</sub> R-S	1	0,948	0,982	0,825	0,762
B <sub>2</sub> R-S		1	0,930	0,833	0,750
B <sub>4</sub> R-S			1	0,834	0,742
B <sub>E</sub> R-S				1	0,757
S-S					1

Tabelle 10.32: Vergleich der gerankten Ergebnislisten der Systeme B<sub>1</sub>R-S, B<sub>2</sub>R-S, B<sub>4</sub>R-S, B<sub>E</sub>R-S und S-S mittels Kendalls  $\tau$ .

Des Weiteren sind die Unterschiede zwischen Baseline S-S und den nächstbesten Verfahren nicht ganz so groß, wie beim Reranking nach Autorennamen, so beträgt der Unterschied im MAP-Wert zwischen dem besten und zweitbesten System knapp 2 Prozentpunkte (0,152 für S-S und 0,131 für B<sub>1</sub>R-S).

Im Gegensatz zum Autorenreranking sind die Unterschiede zwischen den verschiedenen Boostverfahren und der Baseline gering (s. Tab. 10.32) und liegen zwischen 0,742 und 0,762. Die Unterschiede der Verfahren B<sub>1</sub>, B<sub>2</sub>, B<sub>4</sub> und B<sub>E</sub> sind noch geringer, wobei die ersten drei Verfahren fast gleiche Rankings liefern mit  $\tau$ -Werten über 0,930. Die Rankings unterscheiden sich untereinander kaum, was auch die moderaten Wertänderungen der Retrievalleistungen erklären kann.

Die Effekte der Filterungen nach  $p$  und  $\hat{\alpha}$  sind wesentlich moderater (s. Tab. 10.33), da von 65 Topics für die Filterung nach  $p$  noch 63, nach der Filterung nach  $\hat{\alpha}$  34 und für die kombinierte Filterung nach  $p\hat{\alpha}$  noch 33 Topics verbleiben. Trotzdem kann eine

## 10 Auswertung der Experimente zum Reranking

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>1</sub> R-S	-	-	65	0,131	0,178	0,140	0,278
B <sub>1</sub> R-S <sub>p</sub>	✓	-	63	0,134	0,184	0,145	0,287
B <sub>1</sub> R-S <sub><math>\hat{\alpha}</math></sub>	-	✓	34	<i>0,146</i>	<i>0,184</i>	<i>0,148</i>	<i>0,329</i>
B <sub>1</sub> R-S <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	33	<b>0,148</b>	<b>0,190</b>	<b>0,152</b>	<b>0,339</b>

(a) Ergebnisse der Filterung für das Boostverfahren B<sub>1</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>2</sub> R-S	-	-	65	0,118	0,168	0,127	0,255
B <sub>2</sub> R-S <sub>p</sub>	✓	-	63	0,121	0,174	0,131	0,263
B <sub>2</sub> R-S <sub><math>\hat{\alpha}</math></sub>	-	✓	30	<i>0,139</i>	<i>0,180</i>	<i>0,142</i>	<i>0,306</i>
B <sub>2</sub> R-S <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	33	<b>0,141</b>	<b>0,185</b>	<b>0,147</b>	<b>0,315</b>

(b) Ergebnisse der Filterung für das Boostverfahren B<sub>2</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>4</sub> R-S	-	-	65	0,114	0,160	0,122	0,238
B <sub>4</sub> R-S <sub>p</sub>	✓	-	63	0,117	0,165	0,126	0,246
B <sub>4</sub> R-S <sub><math>\hat{\alpha}</math></sub>	-	✓	30	<i>0,134</i>	<i>0,175</i>	<i>0,137</i>	<i>0,288</i>
B <sub>4</sub> R-S <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	33	<b>0,136</b>	<b>0,180</b>	<b>0,141</b>	<b>0,297</b>

(c) Ergebnisse der Filterung für das Boostverfahren B<sub>4</sub>

	$p \geq 0,1$	$\hat{\alpha}_i \leq \varnothing \hat{\alpha}$	Topics	MAP	RP	bpref	P@10
B <sub>E</sub> R-S	-	-	65	0,114	0,158	0,120	0,240
B <sub>E</sub> R-S <sub>p</sub>	✓	-	63	0,117	0,162	0,124	0,248
B <sub>E</sub> R-S <sub><math>\hat{\alpha}</math></sub>	-	✓	30	<i>0,135</i>	<i>0,178</i>	<i>0,138</i>	<i>0,294</i>
B <sub>E</sub> R-S <sub>p<math>\hat{\alpha}</math></sub>	✓	✓	33	<b>0,137</b>	<b>0,181</b>	<b>0,142</b>	<b>0,303</b>

(d) Ergebnisse der Filterung für das Boostverfahren B<sub>E</sub>

Tabelle 10.33: Retrievalergebnisse für ein Reranking mit der Publikationsquelle nach einer Filterung nach  $p \geq 0,1$ ,  $\hat{\alpha}_i \leq \varnothing \hat{\alpha}$  bzw. deren Verknüpfung.

deutliche Steigerung der Retrievalleistungen für alle Ansätze verzeichnet werden. In der Spitze kann das kombinierte Filterverfahren fast bis an die Baseline S-S heranreichen. So liegen die MAP-Werte von  $B_1R-S_{p\hat{\alpha}}$  mit 0,148 nur knapp unter der Baseline von 0,158. Auch die anderen Boostverfahren profitieren ausnahmslos von der Filterung, allerdings sind die Steigerungen nicht statistisch signifikant.

## 10.4 Zusammenfassung der Ergebnisse

Entlang der Forschungsfragen 3–6 sollen in diesem Abschnitt die Ergebnisse des vorangegangenen Kapitels zusammengefasst werden.

**3. Forschungsfrage** *Kann durch gezielten Einsatz von informetrischen Kennzahlen wie der Bestimmung einer Power-Law-Verteilung das Retrievalergebnis beeinflusst werden? Was für Auswirkungen hat der Einsatz auf die Ergebnismenge? Kann der Einfluss durch signifikante Veränderungen in Form einer Cranfield-Evaluation gemessen werden?*

Sowohl für GIRT mit seinen 150 als auch iSearch mit 65 Topics zeigten sich deutliche Veränderungen durch das alternative Ranking, wie in Abbildung 10.3 zu sehen. In der Abbildung sind jeweils die Retrievalergebnisse für die Baseline S und die vier alternativen Boostverfahren  $B_1$ ,  $B_2$ ,  $B_4$  und  $B_E$  zusammengestellt, jeweils eingeteilt nach Vorhandensein der jeweiligen Dokumentenattribute. In allen Varianten waren die Rerankingverfahren dem Baselinesystem, das auf dem Ranking der Suchmaschine Solr basierte, unterlegen. Dies zeigte sich in nahezu allen gemessenen Parametern der Cranfield-Evaluation; in Abbildung 10.3 sind aus Gründen der Übersichtlichkeit nur die MAP-Werte der jeweiligen Systeme dargestellt.

Eine substantielle Einschränkung war allerdings durch den Versuchsaufbau bei der Größe der Ergebnismengen zu beobachten. Je nach eingesetzten Filterverfahren waren die Ergebnismengen wesentlich kleiner als in der ungefilterten Variante, was Auswirkungen auf die Recall- und damit auch auf die dargestellten MAP-Werte hatte.

Dass bei den vorliegenden Experimenten ein tatsächlich alternatives Ranking der Ergebnismengen vorliegt, zeigt die Analyse der verschiedenen Ergebnislisten mittels Kendalls  $\tau$ . In Abbildung 10.2 sieht man auf einem Boxplot die zusammengefassten  $\tau$ -Werte der jeweiligen Rerankingverfahren  $B_1$ ,  $B_2$ ,  $B_4$  und  $B_E$  im Vergleich zur Baseline S für die acht Entitäten des GIRT-Korpus und der zwei Entitäten des iSearch-Korpus (diese sind mit dem Präfix  $i\_$  markiert). Der Boxplot zeigt den Median als dicken Strich innerhalb der Box, die 50 % der Daten umfasst und das obere und untere Quartil darstellt. Die beiden Whisker zeigen jeweils das Minimum und das Maximum der Verteilung an, Kreise markieren Ausreißer. Diese Art der Visualisierung erlaubt eine explorative Analyse der unterschiedlichen Ergebnisse der Kendallanalyse. Man sieht für die sechs Systeme *author*, *classification*, *method*, *pubyear*, *subject* und *i\_author* sehr niedrige Mediane, wenn man diese mit den vergleichsweise hohen Medianwerten der Systeme *issn*, *location*, *publisher* und *i\_source* vergleicht. Für die letztgenannten Systeme zeigt sich keine große Veränderung im Ranking, verglichen mit der Baseline, anders als bei den ersten sechs Systemen.

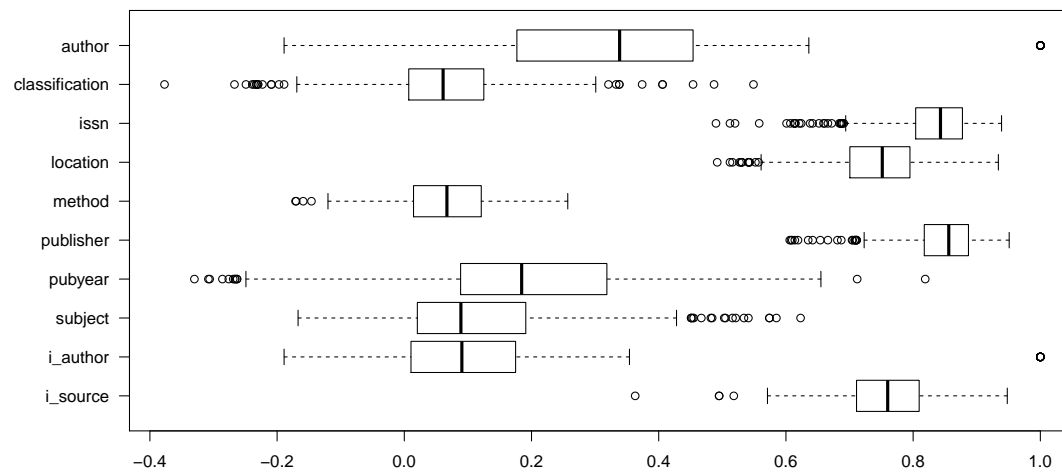


Abbildung 10.2: Boxplot der  $\tau$ -Werte der Rerankingexperimente für alle acht Entitäten des GIRT-Korpus und der zwei Entitäten des iSearch-Korpus (mit Präfix i\_).

**4. Forschungsfrage** *Gibt es einen signifikanten Unterschied zwischen dem Zonen-basierten Ranking des Bradfordizing und der Vermischung von TF\*IDF und lotkaisch-informetrischen Komponenten?*

Die unterschiedlichen Reranking- und Boostvarianten zeigten kein eindeutiges Verhalten. Für die stark eingeschränkten Subtests für ISSN-Codes, Ortsangaben, Verlagsnamen und Publikationsquellen war die Leistungsfähigkeit eindeutig geordnet:  $B_1 \geq B_2 \geq B_4 \geq B_E$ . Für die anderen Subtests war keine solch eindeutige Ordnung zu erkennen. Es zeigt sich aber z. B., dass für das Reranking nach Autorennamen sowohl bei GIRT als auch bei iSearch die klassische zonenbasierte, an das Bradfordizing angelehnte Variante  $B_4$  am leistungsfähigsten war. Bei den Publikationsjahren und den Thesaurustermen war allerdings immer die Variante  $B_E$  am stärksten.

**5. Forschungsfrage** *Ist die Art der untersuchten Entitäten und die jeweils zugrundeliegenden wissenschaftlichen Modelle von Bedeutung? Wie ist die Verbindung zu bekannten informetrischen Verteilungen wie den Gesetzmäßigkeiten nach Lotka oder Bradford?*

Es zeigt sich bei der Analyse der vorliegenden Daten, beispielsweise der MAP-Werte, wie sie in Abbildung 10.3 zusammengefasst sind, dass alleine schon durch das Vorhandensein bestimmter Dokumenteigenschaften (z. B. einen Autorennamen) einen großen Einfluss auf die Leistungsfähigkeit der getesteten Verfahren hat. Dies wird vor allem durch

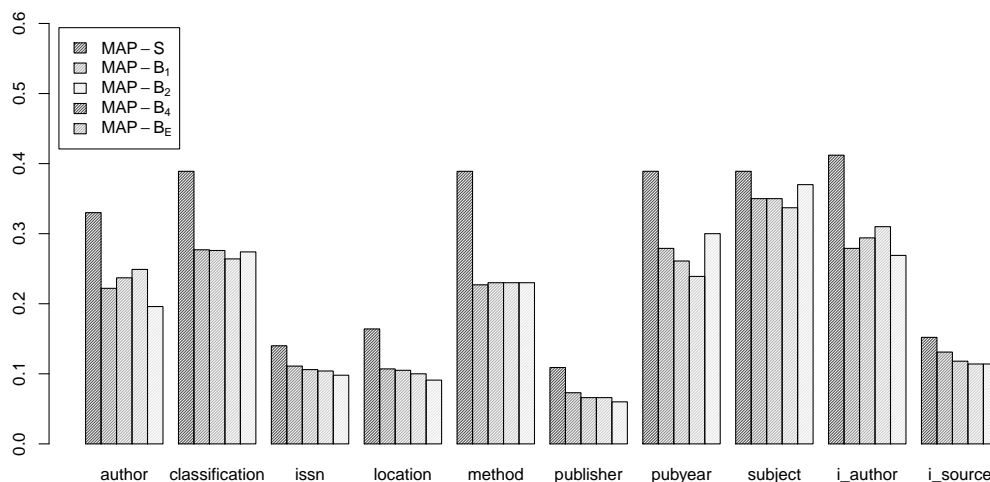


Abbildung 10.3: Filterung nach Entitäten und Vergleich der unbearbeiteten MAP-Werte der Systeme S, B<sub>1</sub>, B<sub>2</sub>, B<sub>4</sub> und B<sub>E</sub> für alle acht Entitäten des GIRT-Korpus und den zwei Entitäten des iSearch-Korpus (mit Präfix i\_).

die stark eingeschränkten Recallwerte deutlich. Während bei der Einschränkung auf Dokumente mit Autorennamen, Klassifikationscode, Forschungsmethode, Erscheinungsjahr und Thesaurustermen keine deutliche Verschlechterung der Recallwerte festzustellen war, war dies bei den Dokumentenfiltern nach ISSN-Codes, Ortsangaben, Verlagsnamen und der Publikationsquelle nicht der Fall. Für die letztgenannten Subtests wurden erhebliche Einbußen in der Retrievalleistung gemessen, wie auch in der Gegenüberstellung der Systeme gut zu sehen ist. Die Unterschiede waren allerdings nicht nur für die Rerankingverfahren, sondern auch für die Solr-Baseline zu sehen, auch wenn diese nicht so stark ausgeprägt waren. Dies war allerdings auch der Grund dafür, die einzelnen Entitäten getrennt voneinander zu evaluieren. Nur so waren die verschiedenen Boostverfahren zu vergleichen.

Einen Rückschluss auf die bestgeeignete Entität für ein Reranking lässt sich auf Grundlage der nicht weiterverarbeiteten Boostverfahren nicht ziehen, da kein System die Baseline übertreffen konnte. Auch zeigt sich in der Gegenüberstellung der MAP-Werte in Abbildung 10.3 kein genereller Trend hin zu bekannten bibliometrischen Verteilungen nach Lotka, Bradford etc., wie es zuvor zu vermuten war. Bei der Beantwortung der Forschungsfrage 6 wird sich allerdings zeigen, dass hier ein anderes Muster erkennbar ist.

**6. Forschungsfrage** *Sind Power-Law-Verteilungen direkt für gutes Ranking verantwortlich? Ist ein „sauberes“ Power-Law verantwortlich für eine bessere Retrievalleistung? Kann man durch Ermittlung der Exponenten oder einer anderen Kennzahl mit Bestimmtheit sagen, ob eine Power-Law-Verteilung geeignet ist, um damit zu ranken?*

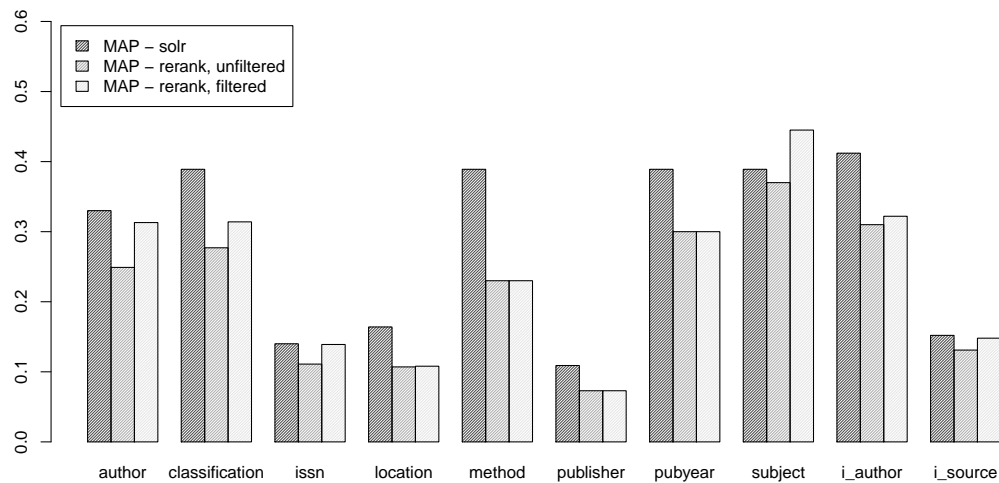


Abbildung 10.4: Vergleich der jeweils ungefilterten und gefilterten MAP-Werte für alle acht Entitäten des GIRT-Korpus und den zwei Entitäten des iSearch-Korpus (mit Präfix i\_).

*Kann man im Vorfeld berechnen, ob sich ein Reranking lohnt?*

Konnte in den ungefilterten Varianten kaum ein System für GIRT die Baseline in einem der Evaluationsparameter MAP, RP, bpref und Cut-off-Precision schlagen, so war dies mit den gefilterten Varianten möglich, wie in Abbildung 10.4 zu sehen ist. Ein Beispiel hierfür war das Reranking anhand der Thesaurusterme. Zwar konnten die ursprünglichen Baselinewerte nicht immer übertroffen werden, doch konnte in einigen Fällen, so z. B. für die Autorennamen und die ISSN-Codes zumindest sehr nah an die Baseline herangereicht werden. Alle drei Verfahren konnten dabei signifikante Verbesserungen im Vergleich zu den ungefilterten Boostverfahren aufweisen. Die anderen fünf Attribute Klassifikationscode, Erscheinungsort, Forschungsmethode, Verlagsname und Erscheinungsjahr eignen sich offensichtlich nicht für ein Reranking, da hier weder eine wirkliche Verbesserung, noch eine Annäherung an die jeweilige Baseline erreicht werden konnte. Für iSearch konnte das Reranking mit der Publikationsquelle an die Baseline heranreichen, das Reranking mit Autorennamen blieb hinter den Erwartungen zurück.

Ein Beispiel für eine erfolgreiche Anwendung des Filters ist in Abbildung 10.5 zu sehen, die zwei Plots für das GIRT-Topic 85 gegenüberstellt. Unterabbildung 10.5a zeigt den doppelt-logarithmierten Plot für die Verteilung der Thesaurusterme, Unterabbildung 10.5b zeigt Verteilung der Erscheinungsjahre. Für die Thesaurusterme ist auch klar das Power-Law zu erkennen, das mit den ermittelten Parameter  $\hat{\alpha} = 2,16$  und  $x_{min} = 3$  dargestellt ist. Für das Erscheinungsjahr konnte nur ein Power-Law mit  $x_{min} = 25$  gemessen werden. Wie auf dem Plot zu erkennen, ist dies die letzte mögliche Position im

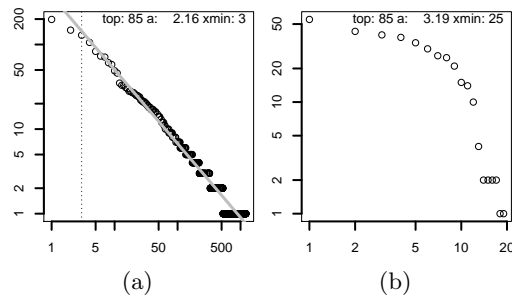


Abbildung 10.5: Gegenüberstellung zweier Plots der gemessenen Power-Law-Verteilungen für das GIRT-Topic 85: (a) Plot der Verteilung für Thesaurusterme, (b) Plot der Verteilung für Erscheinungsjahre.

Long-Tail. Wir können also von einem Messfehler ausgehen, wenn nach den in Abschnitt beschriebenen Messmethoden ein Power-Law-Exponent von  $\hat{\alpha} = 3,19$  ermittelt wurde. Zwar ist das dargestellte Beispiel nur eines von vielen, doch wird der Mehrwert durch den beschriebenen Filter deutlich: Während das System auf Basis der Thesaurusterme den MAP-Wert von 0,377 auf 0,429 ansteigen ließ, konnte das System auf Grundlage der Erscheinungsjahre nur eine Verschlechterung auf den MAP-Wert 0,313 erreichen. Eine Übersicht über alle ermittelten Power-Law-Verteilungen ist im Anhang B ab Seite 257 zusammengefasst.

Die Kendallanalyse zeigte für die vier erfolgreichen Systeme und Verfahren moderate Veränderungen, aber keine extremen Ausschläge, wie z. B. bei dem Reranking nach Klassifikationscodes, wo die  $\tau$ -Werte nahe bei 0 lagen. Die Systeme liefern ein anderes, aber kein vollständig umgestelltes Ergebnis im Vergleich zur Baseline.





# 11 Diskussion

Ausgangspunkt der vorliegenden Arbeit sind Erkenntnisse der IR-Forschung der letzten Jahre (u.a. bei Armstrong u. a., 2009; Cummins u. a., 2011), wonach im klassischen Ad-hoc-Retrieval die Leistung der Systeme stagniert. Obwohl der Vorwurf aufkommen mag, dass dies ein Problem der IR-Laborevaluation sei, zeigen Nutzerstudien, dass auch die Probleme bei der Nutzung von DL-Systemen seit Jahren unverändert sind (s. Abschnitt 2.5). Der grundsätzliche Gedanke dieser Arbeit ist daher, Alternativen zu klassischen Retrievalverfahren zu entwickeln, die über den Abgleich von Anfrage- und Dokumenttermen hinausgehen. Innerhalb des IR-Modells gibt es hierfür eine Reihe von Ansatzpunkten und Verfahren, die weiter optimiert werden können (s. Abb. 1.1 auf Seite 9). In dieser Arbeit wurden die beiden Ansatzpunkte der Anfrageerweiterung und des Dokumentrankings bearbeitet. Aufbauend auf den Ergebnissen, die in den vorangegangenen Kapitel beschrieben wurden, sollen im folgenden Kapitel die Hauptkenntnisse dieser Arbeit mit Hinweisen zu Literatur oder bereits besprochenen Abschnitten diskutiert werden.

## 11.1 Produktivität, Kookkurrenz und Relevanz

Zunächst ist festzuhalten, dass die in dieser Arbeit ermittelten Werte der Produktivitäts- und Kookkurrenzanalyse nicht mit Relevanz gleichzusetzen sind. Wie in Abschnitt 4.2 ausführlich beschrieben, ist Relevanz als ein multidimensionales, dynamisches und komplexes Konzept zu verstehen. Die ermittelten Produktivitätszahlen innerhalb eines IPP (s. Abschnitt 6.1.5) beziffern ausschließlich die Anzahl der von einer Informationsquelle (*Sources*, Zeitschriften, Autoren etc.) produzierten Informationseinheiten (*Items*, welche von den Sources produziert werden, z. B. Artikel). Gleiches gilt für die Kookkurrenzanalyse: Mit ihrer Hilfe kann streng genommen nur die Häufigkeit eines gemeinsamen Auftretens zweier Informationseinheiten (Worte in einem Text, Autorennamen, etc.) innerhalb eines Dokuments gemessen werden. Zwar wird die Annahme zugrunde gelegt, dass eine strukturelle bzw. semantische Abhängigkeit zwischen den Informationseinheiten besteht, doch ist dies nicht immer zwangsläufig korrekt. Beide Verfahren sind per se nicht in der Lage, *relevante* Dokumente zu liefern.

Eine ähnliche Diskussion ist in einer anderen informetrischen Disziplin, der Zitationsanalyse, zu beobachten. Wie von Ingwersen (2012b) beschrieben, ist auch die Anzahl an Zitationen kein Relevanzmerkmal:

„Citations do not signify relevance! But the number of citations signify utility in a particular work/context.“ (Ingwersen, 2012b, Minute 8)

Eine hohe Anzahl an Zitationen ist kein Hinweis auf (*topical*) Relevanz, aber auf die Nützlichkeit in einer jeweiligen Situation. So kann z. B. durch das Zitieren einer Arbeit auch ein negatives Beispiel angeführt werden („... wie von Quelle X gezeigt ist der gewählte Ansatz A falsch, daher wird der alternative Ansatz B verfolgt ...“). In diesem Fall wäre das gewählte Zitat ein Hinweis auf eine offensichtlich falsche Quelle, die objektiv betrachtet keine relevanten Informationen enthält. Mit Hilfe eines Zitationsindex kann laut Ingwersen folglich nur etwas über die „*social utility*“ oder die „*academic (re)cognition*“ ausgesagt werden. Er grenzt dies klar von Relevanz ab.

Diese konzeptionellen Schwierigkeiten zeigen sich auch in Arbeiten von Larsen (2004) sowie Skov u. a. (2008), die Referenzen und Zitationen für den Retrievalprozess nutzbar machen wollten und keine Verbesserungen im Vergleich zu einem regulären Best-Match-Verfahren feststellen konnten. Larsen sieht einen konkreten Schwachpunkt:

„References and citations are fundamentally different from conventional term-based representations, e.g., they retrieve different documents for the same request, and the overlap between the two is typically small.“ (Larsen, 2004, S. 234)

Zum Missverhältnis von Zitationen und Relevanz kommt der Aspekt der Nicht-Zitation („*uncitedness*“) vieler Werke (Egghe u. a., 2011), die den Einsatz von Verfahren der Zitationsanalyse im IR erschweren. Diese Probleme bestehen bei der Produktivitäts- und Kookkurrenzanalyse nicht, da hier mit absoluten Häufigkeiten von vorhandenen Daten gearbeitet wird, die nicht von der direkten Verknüpfung mit anderen Daten abhängig sind.

Der Vergleich der gewählten Verfahren mit der Zitationsanalyse ist an dieser Stelle legitim, da diese mit den gewählten informetrischen Verfahren einige grundlegende Schwächen gemein hat, so seien hier die starke Abhängigkeit von der betrachteten Disziplin und der Qualität bzw. fachlichen Abdeckung der Datengrundlage genannt. Diese Probleme sind allgegenwärtig bei informetrischen Analysen. In dieser Arbeit wurde diesem Umstand Rechnung getragen, indem zwei Datenkollektionen mit unterschiedlicher Erschließungsqualität und Disziplinarität für die Evaluation ausgewählt wurden.

Der Grund warum in dieser Arbeit Produktivitätsauszählung bzw. Kookkurrenzen verwendet wurden, ist darüber hinaus ein sehr pragmatischer: Es gibt wenige IR-Testkollektionen, die Zitationsdaten enthalten. Zahlen zur Produktivität bzw. zur Kookkurrenz lassen sich aus nahezu jedem Datenbestand extrahieren, wobei auch hier Einschränkungen zu machen sind, wie der nächste Abschnitt zeigt.

### 11.2 Bekannte Einschränkungen der evaluierten Verfahren

Die verwendeten Verfahren sind abhängig von bestimmten Einschränkungen, die nachfolgend diskutiert werden: Disziplinarität, fachliche und zeitliche Abdeckung der Datenbanken, Abhängigkeit der Verfahren von der Erschließungsqualität sowie weitere Gründe für Produktivitätsschwankungen.

Die Motivation hinter der Einschränkung dieser Arbeit auf das Anwendungsfeld der digitalen Bibliotheken und die Fokussierung auf die Datenbestände zweier spezieller

Systeme (Sowiport und arXiv) als Evaluationsumfeld sei an dieser Stelle erneut betont. Diese diene zur bewusste Ausklammerung von Websuchmaschinen und der Einbeziehung von strukturierten Nachweissystemen sowie Volltexten um so die Verfahren fokussierter evaluieren zu können. Das Evaluationsumfeld der digitalen Bibliotheken bietet dabei einen ausgewogenen Angelpunkt zwischen Websuchmaschinen und reinen Fachdatenbanken (s. Abschnitt 2.1)

### **Disziplinarität**

Eine klare Einschränkung der vorliegenden Experimente und beim Einsatz informetrischer Verfahren generell ist die starke Abhängigkeit von der untersuchten Forschungsdisziplin. Wie in Kapitel 6 mehrfach beschrieben, sind alle Kennzahlen, die mit Hilfe der Informetrie gewonnen werden, nicht zwangsläufig miteinander vergleichbar. So weisen Iribarren-Maestro u. a. (2009, S. 27 ff.) explizit auf die Besonderheiten und die Schwierigkeiten einer Evaluation innerhalb der Sozialwissenschaften hin. Sie listen z. B. die klare Fokussierung der Sozialwissenschaften auf nationale Themen und Publikationsorgane, aber auch einen starken Hang zur Verwendung der Muttersprache auf. Ein weiterer Punkt ist die Präferenz von Monografien, die im Vergleich zur Physik in den Sozialwissenschaften einen höheren Stellenwert besitzen. Ein klarer Indikator ist der deutlich gesunkene Recallwert in den Experimenten, die sich auf Zeitschriftenartikel beziehen (Einschränkung der Dokumentmenge durch eine vorhandene ISSN). Wie in Tabelle 8.4 zu sehen, sind z. B. für GIRT von ursprünglich 135 214 Dokumenten nur 48 512 Zeitschriftenartikel, allerdings 66 347 Monografien (identifizierbar durch eine ISBN). Für iSearch lässt sich diese Unterscheidung nicht eindeutig treffen, da hierfür die notwendigen Metadaten fehlen, doch ist allgemein davon auszugehen, dass der Anteil an Konferenz-, Workshop- und Zeitschriftenbeiträgen wesentlich größer ist.

Dies hat klare Auswirkungen auf die in dieser Arbeit verwendeten Produktivitätszahlen, die zwischen den Disziplinen nicht vergleichbar sind. Gleiches würde auch für Zitationszahlen gelten, die in einem internationalen Feld, wie der Physik sicherlich anders sind, als in den national-geprägten Sozialwissenschaften.

Nach Hjørland u. Nicolaisen ist neben den Unterschieden in den Disziplinen selbst die Zuordnung zu Disziplinen in den jeweiligen Datenbanken ein grundlegendes Problem. So wird bspw. die Herangehensweise von Bradford in dieser Frage als pragmatisch, aber nicht unumstritten beschrieben:

„Bradford’s conception of ‘subject’ reflects a view close to naïve realism.“  
(Hjørland u. Nicolaisen, 2005, S. 101)

Gerade Papiere, die nicht eindeutig einer einzigen Disziplin zuzuordnen sind, seien ein Problem für die informetrische Analyse. Dies ist besonders in so einem breiten wissenschaftlichen Gebiet wie den Sozialwissenschaften der Fall, die sich aus so verschiedenen Disziplinen wie der Soziologie oder der Politikwissenschaft zusammensetzen. So wird bspw. eine Suche nach Papieren, die sich mit Arbeitslosigkeit auseinandersetzen, in unterschiedlichen Disziplinen andere Aspekte und Akteure hervorbringen. In dieser Arbeit wurde bei den verwendeten Verfahren der von Hjørland u. Nicolaisen kritisierte Pragmatismus umgesetzt. Allerdings haben Untersuchungen von Lüke u. a. (2012) gezeigt,

dass eine fachliche Vorsortierung und Filterung positive Effekte bei der Berechnung von Kookkurrenzen bewirken kann.

### **Fachliche und zeitliche Abdeckungen der Datenbanken**

Ein weiterer Grund für nicht zwangsläufig vergleichbare Produktivitätszahlen sind unterschiedliche Abdeckungen der verwendeten Datenbanken. Während für SOLIS und damit auch GIRT gewisse Abdeckungsregeln definiert sind (s. Abschnitt 8.1.1), sind für den Open-Access-Preprint-Server arXiv, auf dem iSearch basiert, keine solchen Regeln bekannt. Diese könnten durch Art der Akquise bei arXiv, die auf der Selbstmeldung durch Autoren basiert, auch nicht durchgesetzt werden. Es ist zwar in beiden Korpora nicht garantiert, dass z. B. alle Arbeiten eines Autors oder eines Verlags auftauchen, allerdings ist die Abdeckung im Zweifel in GIRT für die vorhandenen Publikationsjahre 1990–2000 höher als in iSearch. Die Abdeckung und damit auch das Alter der Publikationen ist für die verwendeten bibliometrischen Verfahren wichtig. So wird z. B. für Lotkas Gesetz eine Mindestabdeckung von 10 Jahren als Untergrenze angegeben (s. Abschnitt 6.1.2). Diese Mindestabdeckung kann im Prinzip in beiden Korpora eingehalten werden, ob sie allerdings im Einzelfall, d. h. für einzelne Autoren vorliegt, bleibt ungewiss.

### **Abhängigkeit der Verfahren von der Erschließungsqualität**

Die in dieser Arbeit verwendeten Verfahren sind nicht nur abhängig von einer umfangreichen Abdeckung der Datenbanken, sondern auch von einer gewissen Erschließungsqualität der genutzten Metadaten. Wie in Abschnitt 2.4 beschrieben, gelten z. B. für ISSN-Codes klare Erfassungs- und Darstellungsregeln. Diese Regeln existieren für die Erfassung von Autorennamen oder von Publikationsquellen nicht zwangsläufig. Während in der Datenbank SOLIS und damit auch in GIRT klare Erfassungsregeln gelten, die sich auf die Qualität der Metadaten auswirken, sind diese klaren Regeln für iSearch nicht existent (s. Abschnitt 8.1.1 und Tabelle 8.1). Während bei GIRT die Ansetzung der Namen standardisiert ist, gibt es keine einheitlichen Regeln für iSearch, sodass der Einsatz von eindeutigen Autorenidentifikatoren (s. Abschnitt 2.4.3) von Vorteil wäre.

Im Sinne des Schalenmodells nach Krause (1996, 2006) stellen die beiden Testkollektionen GIRT und iSearch zwei unterschiedliche Erschließungstiefen und -qualitätsstufen dar, bzw. zwei unterschiedliche Schalen. SOLIS/GIRT wäre demnach der innersten Schale, dem Kernbereich mit vorselektierter, relevanter Literatur zuzurechnen. iSearch, mit seiner unkontrollierten Erfassung und dem Hintergrund als Open-Access-Plattform, wäre einer weiter außerhalb liegenden Schale zuzuordnen. Krause (2004) beschreibt diese Heterogenität als Zwangsläufigkeit, die nicht vermeidbar sei. Als Lösung für dieses Heterogenitätsproblem schlägt er sogenannte „bilaterale Transfermodule“ vor, die fehlende oder unzureichende Erschließungsinformationen der äußeren Schalen auf die Erschließungswerkzeuge der nächst höheren Schalen abbilden. Umgesetzt wurden solche Transfermodule z. B. in Form der sogenannten Heterogenitätskomponenten wie sie auch im Portal Sowiport eingesetzt werden (Mayr u. Petras, 2008).

Nicht nur die verwendeten Rerankingverfahren, auch die Kookkurrenzanalyse ist auf eine konsistente Erschließung angewiesen. Wie die Anwendung der in dieser Arbeit evaluierten Verfahren im Bereich der Europeana zeigte, ist die Datenqualität maßgeblich

für den Erfolg des Verfahrens verantwortlich (Schaer u. a., 2012a). In dem sehr heterogenen Datenbestand der europäischen digitalen Bibliothek Europeana mit mehr als 23 Millionen Datensätzen war keines der Kookkurrenzverfahren in der Lage, einen positiven Effekt auf den Retrievalprozess auszulösen. Die Datenqualität ließ keine „quantitativ-statistischen“ Ansätze, wie sie von Krause (2004, S. 80) neben Crosskondordanzen und qualitativ-deduktiven Verfahren zur Behandlung der Heterogenitätsproblematik vorgeschlagen werden, zu. In den Arbeiten im Europeana-Kontext wurde bisher nur mit Vokabularen und Thesauri gearbeitet, die als maßgeblicher Zugang zu einem IR-System gesehen wurden. Dass auch andere Dokumentattribute, wie Autoren-, Zeitschriften-, Verlagsnamen und sonstige Bezeichnungen von der Heterogenitätsproblematik betroffen sind, wird durch die Experimenten der vorliegenden Arbeit klar.

#### **Weitere Gründe für Produktivitätsschwankungen**

Außer Acht gelassen wurden in dieser Arbeit unterschiedliche Produktionsintervalle einzelner Informationsproduzenten. So erscheinen z. B. manche Zeitschriften in monatlichen, andere nur in jährlichen Abständen, sodass die Produktivitätszahlen im Detail nicht direkt miteinander verglichen werden können. Ähnliches gilt z. B. auch für unterschiedliche Autorengenerationen, wie von Fronczak u. a. (2007) gezeigt wurde. Für einen wirklich fairen Vergleich müssten das Alter und die Generation eines Autors mit in den Vergleich einbezogen werden. So tendieren jüngere Forschergenerationen zu einer raschen Publikation von Forschungsergebnissen, was im Vergleich zu älteren Generationen die Produktionszahlen in die Höhe treibt:

„Authors have to decide when and how to write up their work. The ideal time is when a piece of research is finished and can carry a convincing message, but in reality it is often submitted at the earliest possible moment (two papers count for twice as much as one, never mind if the second paper mainly corrects errors in the first). Findings are sliced as thin as salami and submitted to different journals to produce more papers.“ (Lawrence, 2003, S. 259)

Diese Politik des *publish or perish* hat natürlich auch Auswirkungen auf die Produktivitätsraten des gesamten Informationsmarktes, wie z. B. die von Verlagen oder Forschungseinrichtungen. Andere Ursachen für eine schlechte Vergleichbarkeit von Produktivitätszahlen sind in versteckten Gründen, wie z. B. wirtschaftlichen Problemen, Krankheit, Austritt aus dem Wissenschaftsbetrieb etc. zu finden. Auch wenn diese Entwicklungen in der Wissenschaftslandschaft nachzuvollziehen und zu beobachten sind, können sie nicht in den Indikatoren, die in dieser Arbeit verwendet wurden, berücksichtigt werden.

### **11.3 Die Konsequenz von informetrischer Analyse für das IR**

Trotz der System- und Datenbasis-bedingten Schwächen der Verfahren hat die vorliegende Arbeit klare Erkenntnisse hervorgebracht, die im folgenden Abschnitt zusammenfassend diskutiert werden.

### **Nicht-textuelle Informationen als Mehrwert für das Information Retrieval**

Der grundlegende Gedanke hinter der vorliegenden Arbeit ist der Einsatz von Methoden der Informatik zur Generierung von Mehrwerten für das IR. Die durch die informatischen Analysen gewonnenen Daten sollen die üblicherweise zum Retrieval verwendeten textuellen Daten um nicht-textuelle (Meta-)Daten erweitern, um so ein umfangreicheres und nutzerzentrierteres Retrieval zu ermöglichen.

Die Experimente zum Dokumentreranking haben die besondere Stärke der Varianten mit Thesaurustermen gezeigt. Auch in den vorliegenden Experimenten waren an allen erfolgreichen Systemen die Thesaurusterme beteiligt. Ein ähnliches Ergebnis haben Arbeiten von Kamps (2004) gezeigt:

„This paper introduced a new feedback technique that reranks the set of initially retrieved documents based on the controlled vocabulary terms assigned to the documents. Our reranking strategy significantly improved retrieval effectiveness in domain-specific collections, above and beyond the use of standard Rocchio blind feedback.“ (Kamps, 2004)

Kamps vergleicht in seiner Arbeit seinen Ansatz mit einem Pseudo-Relevance-Feedback-Verfahren und arbeitet dabei mit der GIRT-Testkollektion. Auch er kann den starken Einfluss der kontrollierten Vokabulare bestätigen. In seinen Experimenten zeigt sich auch der Vorteil der lokalen Analyse gegenüber der globalen (s. Abschnitt 3.3), wobei tendenziell entgegengesetzte Ergebnisse in eigenen Untersuchungen (Hienert u. a., 2011a; Schaer u. a., 2012c) gezeigt wurden.

Die in Abschnitt 2.5 vorgestellten Nutzerstudien in digitalen Bibliotheken und zum Suchverhalten haben deutlich gezeigt, dass Nutzer zur Relevanzeinschätzung auf nicht-textuelle Attribute wie „Prestige“ oder „Einschlägigkeit“ zurückgreifen. Auch zeigte sich, dass Nutzer des Portals Sowiport, das als Prototyp der in dieser Arbeit betrachteten digitalen Bibliotheken steht, in 1/3 der Suchanfragen gezielt nach Personen und nur in 1/5 der Fälle nach Fachbegriffen oder Schlagwörtern suchen.

Wertet man nun die gemessenen Unterschiede im Ranking zwischen dem gewählten Baseline-System auf Basis von Solr und den jeweiligen Mehrwertdiensten als einen Gradmesser für den gewonnenen Mehrwert, dann waren alle implementierten Dienste erfolgreich. Gemessen an den  $\tau$ -Werten nach Kendall konnten die Mehrwertdienste eine signifikant andere Ergebnisliste im Vergleich zur Baseline generieren. In Analogie zu vorherigen Arbeiten im Bereich der Suchunterstützung und der Mehrwertdienste für das Retrieval kann dieses Ergebnis als eine neue, alternative Sicht auf die Dokumentmenge gewertet werden.

„The user gets a new result cutout containing other relevant documents which are not listed in the first section of the original list. Additionally, the re-ranking via structure-oriented science models offer an opportunity to switch between term-based search and structure-oriented browsing of document sets [...]“ (Mutschke u. a., 2011, S. 362)

In vorherigen Arbeiten zu dieser Art des Mehrwertes wurde noch ein sehr einfaches Maß zur Bestimmung der Unterschiedlichkeit zweier Ergebnismengen gewählt, indem nur die Überschneidungen der jeweiligen Top 10 Dokumente gezählt wurden.

„A comparison of the intersection of the relevant top 10 document result sets between each pair of retrieval service shows that the result sets are nearly disjoint. 400 documents [...] only had 36 intersections in total [...]. Thus, there is no or very little overlap between the sets of relevant top documents obtained from different rankings.“ (Mutschke u. a., 2011, S. 360 f.)

Obwohl diese Art der Bestimmung von Differenz zwar in frühen TREC-Studien (bis TREC 3) verwendet wurde, wurde doch in Studien zur Relevanzbewertung gezeigt, dass bei einem Anstieg der Assessoren bzw. der zu bewertenden Kategorien diese Art der Bestimmung keine statistisch stabilen Ergebnisse liefert.

„In early years [...] measures like Jaccard coefficient or the intersection method were – and still today [...] are – used. These measures are getting unstable and unreliable as the number of categories [...] increases.“ (Schaer, 2012, S. 125)

Kendalls  $\tau$  wird allgemein zwar als Indikator für die Unterschiedlichkeit im Ranking in der IR-Community verwendet (Voorhees, 2000), allerdings selten, um damit eine neue Sicht auf die Dokumentmenge zu begründen. Um den Neuigkeitswert für den Benutzer und damit einen Mehrwert zu quantifizieren, ist Kendalls  $\tau$  aussagekräftiger als die Bestimmung der Größe der Schnittmenge der gemeinsamen Dokumente der unterschiedlichen Systeme. Natürlich besteht die Gefahr, dass die Mehrwertdienste zwar ein anderes, aber nicht zwangsläufig relevanteres Ergebnis liefern. Dies hat sich in der Analyse der Dienste gezeigt: Nur wenige Systemkonfigurationen waren in der Lage, die Baseline zu übertreffen oder zumindest eine vergleichbare Retrievalleistung zu liefern. Kendalls  $\tau$  sollte folglich nur zusammen mit anderen akzeptierten IR-Kennzahlen zur Messung der Retrievalleistung, wie MAP, b-pref etc. verwendet werden (Sanderson u. Soboroff, 2007).

#### **Die Messung informatrischer Gesetzmäßigkeiten als Prüfkriterium für IR-Filter**

In dieser Arbeit wurde zum ersten Mal ein empirischer Zusammenhang zwischen den quantitativen Ergebnissen einer informatrischen Analyse und deren Auswirkung auf das Information Retrieval hergestellt. Die Maße zur Bewertung der Güte der jeweiligen Analyse waren (1) die Prüfung der mathematischen Definition und der Stärke des Gefalles einer Power-Law-Verteilung (2) sowie die Stärke der Assoziativität zwischen textuellen und nicht-textuellen Dokumentattributen durch Kookkurrenzanalyse (s. Abschnitte 9.3 und 10.4).

Die Experimente, sowohl für die Rerankingverfahren als auch für die Anfrageerweiterung, haben klar die Grenzen der Verfahren gegenüber einer State-of-the-Art-Retrievalimplementierung gezeigt. Ohne besondere Versuchsaufbauten oder Filter konnte keines der Verfahren aus dem Stand mit den Ergebnissen, die das Solr-System mit seiner Freitextsuche als Baseline für alle Experimente lieferte, konkurrieren. Durch die gezielte Analyse der Häufigkeitsverteilungen für kookkurrierende Dokumentattribute

und deren Untersuchung hinsichtlich des Vorliegens einer Power-Law-Verteilung und deren Anwendung als Filter für die entwickelten Systeme konnte die Retrievalleistung der betrachteten Systeme hingegen die Baseline schlagen oder zumindest an sie heranreichen. Dies war sowohl für die Dokumentattribute Autorennamen, Zeitschriftenzugehörigkeit (ISSN-Codes) und kontrollierte Verschlagwortung mit Thesaurustermen für GIRT als auch für die Publikationsquelle für iSearch, zu beobachten. Gleichzeitig konnte die Analyse der  $\tau$ -Werte zeigen, dass das Reranking für diese Systeme eine substantiell veränderte Ergebnisliste generiert. Gleiches gilt für die Anfrageerweiterungen, wobei hier die aufgetretenen Effekte für die Sortierung noch deutlicher sind.

Dies ist zunächst erstaunlich, da die gewählten Verfahren und die eingesetzten Filter- und Gütekriterien keinen direkten Bezug zur Relevanz haben (s. Abschnitt 11.1). Die Interpretation der Power-Law-Exponenten als Filterkriterium für das Retrieval ist in dieser Anwendung ungewöhnlich, da dieser in der Literatur nicht für praktische Schlussfolgerungen genutzt wird. Der Grad des Gefälles (*slope*) der Power-Law-Funktion wird allgemein als die Stärke der Unterschiede in einer Verteilung interpretiert. Eine akzeptierte Interpretation ist das Gefälle bspw. als einen Gradmesser der „Fairness“ in einem Netzwerk zu sehen (Kunegis u. Preusse, 2012). Eine exakte Quantifizierung der Power-Law-Verteilungen hat aber meist nur einen beschreibenden Charakter oder wird als nicht notwendig eingestuft:

„I also hope that we as a community start asking ourselves what exactly we mean when we say that such-and-such a quantity is 'power-law distributed', and whether our meaning would be better served at times by using less precise terms such as 'heavy-tailed' or simply 'heterogeneous'“ (Clauset, 2007)

Die von Aaron Clauset (auf dessen Berechnungsverfahren die Abschätzung der Power-Law in dieser Arbeit beruht) vertretene Meinung deckt sich mit der von Howard D. White, der in persönlichen Gesprächen beschrieb, er selbst prüfe nicht exakt auf die Existenz eines Power-Laws, sondern werfe nur einen Blick auf den Plot der Verteilung. Dies reiche ihm als Einschätzung, um zu erkennen, ob er eine „power-law-like“ Verteilung betrachte.

Auch Umstätter spricht sich gegen eine exakte Auszählung (konkret beim *Bradford's Law of Scattering*, BLS) aus:

„Die Grundaussage im BLS war, dass sich eine immer gleiche Menge an Literatur auf verschiedene, thematisch geordnete Zeitschriften wie  $n_0:n_1:n_2:n_3 \dots$  verteilt. Schon an dieser Schreibweise ist erkennbar, dass es sich lediglich um eine Präzisierung der 80:20 Regel handelt. [...] Bei jeder natürlichen Beobachtung in der Fachliteratur würde ein Streit darüber, ob man beispielsweise mit 1:30:900:27000 oder 1:31:931:27.931 rechnen müsste, schon allein durch die auftretenden Streuungen ad absurdum geführt.“ (Umstätter, 2005a)

Für Bradford's Law of Scattering heißt dies, dass es nicht „die zentrale“ Zeitschrift gibt, sondern dass immer ein Kern von zentralen Zeitschriften betrachtet werden muss. Dieser kann je nach Größe der betrachteten Zeitschriftenmenge sehr viele Zeitschriften enthalten. Die Menge der zu einem Thema relevanten Zeitschriften ist ausschlaggebend. Auch in den



Arbeiten von Mayr (2009b) zum Bradfordizing stand nur die Unterscheidung zwischen *Core* und *Scatter*, als dem Kern und der Peripherie, im Vordergrund.

Die Analyse und Gegenüberstellung der verschiedenen Boostverfahren in dieser Arbeit (s. Abschnitte 7.2.2 und 10.4) zeigen allerdings, dass es durchaus einen Unterschied machen kann, wenn auch Entitäten untereinander nicht nur in Zonen eingeteilt, sondern absolut gerankt werden. Im Gegensatz zur klassischen BLS-Forschung haben sich die vorliegenden Untersuchungen dabei nicht nur auf Zeitschriften beschränkt.

#### **Empirisch umfangreiche Evaluation nach dem Cranfield-Paradigma**

Eine Stärke der vorliegenden Experimente ist die konsequente Anwendung des Cranfield-Paradigmas der IR-Laborevaluation wie es in Kapitel 5 ausführlich vorgestellt wurde. Wie bereits in diesem Abschnitt angedeutet, ist die Laborevaluation innerhalb der „Cranfield Cave“ (Ingwersen u. Järvelin, 2007) nur eine mögliche Ausprägung, wie IR-Systeme evaluiert werden können. Ein klarer Vorteil ist dabei die starke empirische Grundlage der Untersuchungen, die eine ständige Wiederholbarkeit und Nachvollziehbarkeit der Experimente erlaubt, da alle Systemparameter bekannt sind.

Vergleicht man die Ergebnisse der vorliegenden Experimente mit vorangegangenen Experimenten z. B. im Bereich des Bradfordizing von Mayr (2009b), so fällt auf, dass dort Verbesserungen in der Precision von bis zu 56 % gemessen wurden. Die von Mayr vorgelegten Precisionwerte dürfen allerdings nicht mit den in dieser Arbeit ermittelten Werten verglichen werden. Zunächst ist der Versuchsaufbau gänzlich verschieden, so werden bei Mayr zunächst nur die drei Zonen Core, Zone 2 und Zone 3 mit der Baseline verglichen. Die Baseline ist die gemittelte Precision aller Dokumente im Pool. Es wurde kein Dokumentranking aufgestellt, sondern nur die drei Zonen als grobe Einteilung herangezogen. Die gemittelte Precision über alle Dokumente hinweg impliziert, dass ein Baseline-System die Dokumente gleichverteilt mit der Durchschnittsprecision anordnen würde. Dass dies nicht der Fall bei einem Relevanzranking ist, ist direkt ersichtlich. Mayr hat im Grunde ein ungeranktes Retrievalsystem (gemittelte Precision) mit einem gerankten System (Bradfordizing) verglichen. In dieser Arbeit wurde stattdessen ein tatsächliches IR-System mit einem durchgängigen Ranking auf Dokumentbasis implementiert und evaluiert. Die gewählte Baseline auf Grundlage der Suchmaschine Solr ist daher eine realistischere Baseline.

Weiterhin wurden bei Mayr jeweils die einzelnen Jahrgänge von CLEF als Evaluationsgrundlage gewählt, die jeweils 25 Topics umfassten. Die Analyse basierte dabei jeweils auf den Jahrgängen, was allerdings den Nachteil hat, dass die Größe von 25 Topics allgemein als zu niedrig für ein statistisch aussagekräftiges Evaluationsergebnis betrachtet wird. Meist wird eine Topicanzahl von mindestens 50 als untere Grenze angesehen, wenn mit Evaluationsmaßen wie der Cut-off-Precision gearbeitet wird (Voorhees, 2009). Da in dieser Arbeit für die beiden Testkollektionen mit 150 bzw. 65 Topics gearbeitet wurde, können die hier präsentierten Ergebnisse als aussagekräftiger angesehen werden.

Natürlich kann die Wahl von Solr als Baseline in dieser Arbeit in Frage gestellt werden, da eine andere IR-Engine oder ein anderes Retrievalverfahren, z. B. Sprachmodelle oder andere auf dem probabilistischen Retrievalmodell (s. Abschnitt 4.3.4) basierende Verfahren, hätten eingesetzt werden können. Die gewählte Baseline entspricht allerdings

dem Stand der Technik in heutigen Systemen der wissenschaftlichen Informationssuche in Portalen wie Sowiport. Dieses System verwendet den gleichen Softwarestack wie er in den vorliegenden Experimenten verwendet wurde (s. Abschnitte 2.6 und 4.3.5). In vielen aktuellen Systemen, wie arXiv auf dessen Datenbasis die iSearch-Kollektion basiert, ist erst gar kein Relevanzranking implementiert. Gleiche Argumente gelten für die Implementierung der Anfrageerweiterung. Auch hier wurde der Stand der Technik in vorhandenen Systemen als Baseline gewählt. Das Ziel war es nicht eine optimale Gewichtungsfunktion oder ein möglichst effizientes Werkzeug zur Kookkurrenzanalyse zu finden. Vielmehr wurde das Hauptaugenmerk auf die Analyse der unterschiedlichen Entitäten für die Anfrageerweiterung gelegt.

### **Möglichkeiten der Evaluation im interaktiven IR**

Es konnte für das Rerankingverfahren ein eindeutiger Zusammenhang zwischen der Güte der Power-Law-Verteilung und der Retrievalleistung gezeigt werden. Diese Information könnte für das Design eines interaktiven und proaktiven Dienstes zur Suchunterstützung genutzt werden (s. Abschnitt 4.5), indem ein solches Reranking dem Benutzer erst angeboten wird, wenn dieses sich zumindest rechnerisch lohnt. Dass diese Art der Suchunterstützung klare Verbesserungen im Retrievalprozess leisten kann, zeigen Kriewel u. Fuhr (2010):

„Users with assistance were able to search more successfully, as measured by the number of relevant documents saved – at least for some tasks. It seems that not all tasks benefit from the suggestions in their current form. If this results from a lack of suitable cases, is due to the specific search task, or points to a more general problem of the approach needs further examination. For those tasks where users were helped by the suggestions, the differences were very significant ( $p = 0.011$  and  $p = 0.022$ ).“ (Kriewel u. Fuhr, 2010, S. 553 f.)

Ein solches System kann allerdings mit den Mitteln der IR-Laborevaluation nicht umfassend untersucht werden. Vielmehr müssten hierfür Evaluationsmethoden des IIR zum Einsatz kommen, wie sie z. B. von Kelly (2009) vorgestellt werden.

### **Information Retrieval als Lackmus-Test für die Informatik**

Wie bereits zuvor beschrieben, haben die vorgestellten Verfahren und die informatische Analyse als Qualitäts- und Filterkriterium eine Auswirkung auf die Leistungsfähigkeit des verwendeten Retrievalsystems. Aber auch die Interpretation in die andere Richtung ist möglich, die nicht den Mehrwert der Informatik für das IR betrachtet, sondern den des IR für die Informatik. Die empirische Überprüfung von informatischen Modellen oder von „science models“ (Scharnhorst, 2012) stellt ein Problem dar, dem mit dem rigiden Evaluationsframework des Cranfield-Paradigmas begegnet werden soll:

„The science models studied are therefore verified as expressive models of science, as an evaluation of retrieval quality is seen as a litmus test of the adequacy of the models investigated.“ (Mutschke u. a., 2011, S. 362)

Dieser Lackmustest (*litmus test*) zielt darauf ab, die Vorteile der IR-Laborevaluation für die Informatrie zu nutzen. Wie auch in der Chemie wird hier der Begriff des Lackmustest nur für eine grobe Abschätzung verwendet (z. B. ob eine saure oder basische Lösung vorliegt, bzw. ob die Gültigkeit eines informatrischen Modells plausibel erscheint oder nicht).

Die Gültigkeit der Modelle soll durch die Anwendung im IR (zumindest teilweise) belegt werden. Ein auf diesen Modellen basiertes IR-System kann bei erfolgreicher Evaluation mit einer IR-Testkollektion als ein Indikator für die Gültigkeit bzw. Plausibilität des betrachteten Modells gesehen werden. Die Verknüpfung von IR-Relevanzurteilen und den Ergebnissen einer informatrischen Analyse ist allerdings nur ein Indikator, der genutzt werden könnte, um einen Hinweis auf Korrektheit eines zugrundeliegenden Modells zu erlangen kann.

Dass dieser Gedanke prinzipiell machbar erscheint und auch andere Szenarien denkbar sind, zeigen Schaer u. Heck (2013) in einem Pretest mit Hilfe der iSearch-Testkollektion. In dieser Vorstudie wurden zehn Informationsbedürfnisse von Wissenschaftlern des Forschungszentrums Jülich untersucht. Die Wissenschaftler wurden gebeten ihr Forschungsinteresse in eigenen Worten zu beschreiben und für sie einschlägige Autoren zu nennen. Es zeigte sich, dass sowohl eine Vielzahl an Dokumenten passend zum Forschungsinteresse als auch die Autoren in der iSearch-Testkollektion vorhanden waren. Von den jeweils genannten Autoren waren ca. 70 % in iSearch enthalten, wobei durchschnittlich 19 Artikel pro Autor vorhanden waren. Die iSearch-Testkollektion könnte als Testbett für beliebige Modelle, die auf Autoren oder Forschungsinteressen basieren, angewendet werden. Die jeweils dokumentierten tatsächlichen Informationsbedürfnisse und Interessen der Forscher könnten als Gold-Standard genutzt werden. Gleichzeitig zeigte ein IR-Retrievaltest auf Grundlage von Solr, dass mit einem herkömmlichen Relevanzranking unter den ersten 50 Treffern im Schnitt nur 4,8 einschlägige Autoren und 4,8 Papiere dieser Autoren gefunden wurden.



## 12 Ausblick

Ziel eines jeden Verfahrens, das im Bereich digitaler Bibliotheken entworfen wird, sollte es sein, einen tatsächlichen Mehrwert für Benutzer zu erbringen. Hierzu können die durchgeführten Analysen und Evaluationen nur indirekt eine Aussage treffen. In dieser Arbeit wurde stark auf die Evaluation der Verfahren mittels der klassischen IR-Laborevaluation gesetzt, allerdings wurden keine Benutzertests oder anderen Methoden des interaktiven IR (IIR, s. Abschnitt 5.1) eingesetzt. Für eine umfassendere, holistische Bewertung der vorliegenden Verfahren, müsste die Evaluation außerhalb der Cranfield Cave stattfinden. Die in dieser Arbeit formulierten Ergebnisse und Diskussionen geben einen starken Hinweis auf die Wirksamkeit der vorgeschlagenen Verfahren. Um die Erkenntnisse und die Wirksamkeit allerdings endgültig empirisch abzusichern, bedarf es weitergehender Untersuchungen, die Benutzer und deren konkreten Informationsbedürfnisse in den Evaluationsprozess mit einbeziehen.

Neben qualitativer Benutzerstudien müssten die Verfahren in heterogenen Datenbeständen von digitalen Bibliotheken getestet werden. Zwar geben die beiden Beispieldatensätze GIRT und iSearch einen Hinweis auf die Wirksamkeit in verschiedenen Domänen und Datenqualitätsstufen, doch müsste für eine stärkere Aussagekraft der Experimente mit größeren und heterogeneren Datenbeständen, wie sie z. B. in Sowiport vorliegen, gearbeitet werden. Auch das Zusammenwirken der Verfahren untereinander und im Zusammenspiel mit anderen Mehrwertdiensten in DL-Systemen, wie z. B. den Heterogenitätskomponenten, ist bisher nur unzureichend untersucht worden.

Einen Hinweis auf die Ausgestaltung dieser Experimente haben die vorliegenden Untersuchungen bereits gegeben. Es liegt nahe, die informatrische Analyse als Kriterium für das Reranking der Ergebnismenge proaktiv bzw. kontextsensitiv anzubieten und darüber hinaus als interaktiven Dienst einzubinden, anstelle eines automatischen Rerankings (s. Abschnitt 4.5 und Ingwersen, 2012b). Es konnte für das Rerankingverfahren ein eindeutiger Zusammenhang zwischen der Güte der Power-Law-Verteilung und der Retrievalleistung gezeigt werden. Das Wissen des Systems über das Vorliegen einer tatsächlichen Power-Law-Verteilung könnte so als ein Bestandteil des „Interactive Information Retrieval Probability Ranking Principle“ (IIR-PRP), wie es von Fuhr (2008) vorgestellt wurde, einfließen. Hierbei werden interaktive Prozesse in einem Retrievalprozess durch die Übergangswahrscheinlichkeiten zwischen verschiedenen Stadien und Interaktionsmöglichkeiten in einem IR-System modelliert (Tran u. Fuhr, 2012). Im Rahmen des beantragten DFG-Projektes AMUR von GESIS und der AG Information Engineering der Universität Duisburg-Essen unter Leitung von Norbert Fuhr sollen vergleichbare Experimente vor dem Hintergrund des sozialwissenschaftlichen Portals Sowiport durchgeführt werden.

Ein weiterer Projektantrag, der Ergebnisse dieser Arbeit aufgreift, ist das Projekt Author Map, das parallel bei der DFG und der amerikanischen NEH eingereicht wurde.

Zusammen mit Howard D. White und Xia Lin des College of Information Science and Technology der Drexel University soll darin u. a. an der interaktiven Visualisierung sogenannter Pennant-Diagramme (s. Abb. 4.1) und an Pathfinder-Netzwerken gearbeitet werden. Die in dieser Arbeit verwendeten Kookkurrenzanalysen eignen sich sehr für die Darstellung in solchen Netzwerken. So könnten bspw. die für einen Suchbegriff stark assoziierten Autoren oder Zeitschriften visualisiert und dem Benutzer als interaktiver Suchunterstützungsdienst angeboten werden. Wie bereits in Systemen wie Daffodil oder ezDL (Kriewel u. a., 2004) demonstriert, könnte so mit Hilfe von Techniken des HCIR (s. Abschnitt 4.5) das klassische Retrieval um interaktive Elemente erweitert werden. Bereits in diesen Systemen hat sich gezeigt, dass sich durch die Einbeziehung von Suchtaktiken und -strategien (Bates, 1990) ein erheblicher Mehrwert bei der Suche für den Nutzer schaffen lässt.

Neben den beiden durch Projektanträge skizzierten nächsten Schritten könnte die Übertragung der informetrischen Analysen auf andere Verfahren vielversprechend sein. Wie von Egghe u. Rousseau (2006) gezeigt wurde, kann der h-Index in einem IPP wie folgt definiert werden:

„This definition is extended here to the general framework of Information Production Processes (IPPs), using a source-item terminology. It is further shown that in each practical situation an IPP always has a unique h-index. In Lotkaian systems  $h = T^{1/\alpha}$ , where T is the total number of sources and  $\alpha$  is the Lotka exponent.“ (Egghe u. Rousseau, 2006, S. 121)

Diese Verbindung zwischen h-Index und den hier untersuchten Power-Law-Verteilungen könnte ausgenutzt werden, um die vorhandenen Schwächen von zitationsbasierten Dokumentrankingverfahren (Larsen u. Ingwersen, 2006) zu kompensieren.

Neben den traditionellen informetrischen Analyseverfahren ließe sich eine Verbindung zum Altmetrics-Gedanken (Priem u. a., 2011) herstellen.

„Citation counting measures are useful, but not sufficient. [...] Citation measures are narrow; influential work may remain uncited. These metrics are narrow; they neglect impact outside the academy, and also ignore the context and reasons for citation. The JIF, which measures journals' average citations per article, is often incorrectly used to assess the impact of individual articles. [...] altmetrics are themselves diverse, they're great for measuring impact in this diverse scholarly ecosystem. In fact, altmetrics will be essential to sift these new forms, since they're outside the scope of traditional filters. (Priem u. a., 2011)

Ausgangspunkt der Diskussion ist, dass im Web klassische informetrische bzw. webometrische Verfahren (Björneborn u. Ingwersen, 2004) nicht mit der Schnellebigkeit des Webs Schritt halten können und viele Aspekte, die für eine ganzheitliche Bewertung des Impacts eines Wissenschaftlers wichtig wären, nicht berücksichtigt werden können. Die Autoren nennen hier beispielsweise:

„The sharing of ‚raw science‘ like datasets, code, and experimental designs Semantic publishing or ‚nanopublication‘, where the citeable unit is an argument

or passage rather than entire article. Widespread self-publishing via blogging, microblogging, and comments or annotations on existing work.“ (Priem u. a., 2011)

Möchte man Teile der vorliegenden Arbeit, wie die Verbindung von Informetrie und IR in das Web-Retrieval überführen, so müssten neben den bekannten webometrischen Indikatoren wie dem *Web Impact Factor*, auch Faktoren der Altmetrics betrachtet werden. In eine ähnliche Richtung gehen auch neuere Überlegungen des sogenannten *Expertise Retrieval*, wie es unter anderem von Balog (2012) vorangetrieben wird. Auch hier werden alternative Verfahren vorgeschlagen, um informetrische Verfahren in das Web zu überführen und für das Retrieval zu verwenden.

Ein vollständig offener Punkt in den vorliegenden Untersuchungen ist die in Abschnitt 11.3 angedeutete Diskussion der Verwendung von IR-Evaluationsmethoden für die Verifizierung informetrischer Untersuchungsergebnisse. Zwar liefern erste Voruntersuchungen (Schaer u. Heck, 2013) einen Hinweis auf die generelle Machbarkeit dieser Idee, doch gibt es bisher keine gesicherten Erkenntnisse, ob diese Forschungsrichtung konkret umsetzbar ist und ob sich das rigide Evaluationskonzept der IR-Laboruntersuchung mit der Informetrie auf diese Weise verbinden lassen.





# Literaturverzeichnis

Die Gültigkeit der Internetquellen in diesem Literaturverzeichnis sowie im Rest dieser Arbeit wurde zuletzt am 6. Februar 2013 überprüft.

- [Adamic 2000] ADAMIC, Lada A.: *Zipf, Power-laws, and Pareto - a ranking tutorial*. <http://www.hpl.hp.com/research/idl/papers/ranking/ranking.html>
- [Adamic u. Huberman 2000] ADAMIC, Lada A. ; HUBERMAN, Bernardo A.: The Nature of Markets in the World Wide Web. In: *Quarterly Journal of Electronic Commerce* (2000), Nr. 1, S. 5–12. – DOI 10.2139/ssrn.166108
- [Adema u. Rutten 2010] ADEMA, Janneke ; RUTTEN, Paul: Digital Monographs in the Humanities and Social Sciences: Report on User Needs / OAPEN: Open Access Publishing in European Networks. <http://srv5.axmediatech.com/axmedis/a61/00000-a611870b-aae2-43ee-83ad-24c5777ee6b6/2/AXMEDIS7681.pdf>. Amsterdam, Januar 2010 (D 3.1.5). – Forschungsbericht
- [Albert u. a. 1999] ALBERT, Reka ; JEONG, Hawoong ; BARABASI, Albert-Laszlo: Internet: Diameter of the World-Wide Web. In: *Nature* 401 (1999), Nr. 6749, 130–131. – DOI 10.1038/43601
- [Alonso u. a. 2010] ALONSO, Omar ; SCHENKEL, Ralf ; THEOBALD, Martin: Crowdsourcing Assessments for XML Ranked Retrieval. In: GURRIN, Cathal (Hrsg.) ; HE, Yulan (Hrsg.) ; KAZAI, Gabriella (Hrsg.) ; KRUSCHWITZ, Udo (Hrsg.) ; LITTLE, Suzanne (Hrsg.) ; ROELLEKE, Thomas (Hrsg.) ; RÜGER, Stefan (Hrsg.) ; RIJSBERGEN, Keith van (Hrsg.): *Advances in Information Retrieval* Bd. 5993. Berlin, Heidelberg : Springer, 2010 (Lecture Notes in Computer Science). – ISBN 978-3-642-12274-3, S. 602–606
- [Anderson 2011] ANDERSON, Rick: The Crisis in Research Librarianship. In: *The Journal of Academic Librarianship* 37 (2011), Juli, Nr. 4, 289–290. – DOI 10.1016/j.acalib.2011.04.001
- [Arms 2000] ARMS, William: *Digital libraries*. Cambridge, MA, USA : MIT Press, 2000 <http://www.cs.cornell.edu/wya/diglib/>. – ISBN 9780585281322
- [Armstrong u. a. 2009] ARMSTRONG, Timothy G. ; MOFFAT, Alistair ; WEBBER, William ; ZOBEL, Justin: Improvements that don't add up: ad-hoc retrieval results since 1998. In: *Proceeding of the 18th ACM conference on information and knowledge management*. Hong Kong, China : ACM, 2009 (CIKM '09). – ISBN 978-1-60558-512-3, S. 601–610

- [Aula u. a. 2005] AULA, Anne ; JHAVERI, Natalie ; KÄÄKI, Mika: Information search and re-access strategies of experienced web users. In: *Proceedings of the 14th international conference on World Wide Web*. New York, NY, USA : ACM, 2005 (WWW '05). – ISBN 1-59593-046-9, 583–592
- [Bachelier 2001] BACHELIER, Günter: *Polyrepräsentation, Relevanz-Approximation und aktives Lernen im Vektorraummodell des Information-Retrievals*. Saarbrücken, Philosophische Fakultäten der Universität des Saarlandes, Dissertation, Juni 2001. <http://scidok.sulb.uni-saarland.de/volltexte/2002/66/>
- [Balog 2012] BALOG, Krisztian: Expertise Retrieval. In: *Foundations and Trends in Information Retrieval* 6 (2012), Nr. 2-3, 127–256. – DOI 10.1561/15000000024
- [Bao u. a. 2007] BAO, Shenghua ; XUE, Guirong ; WU, Xiaoyuan ; YU, Yong ; FEI, Ben ; SU, Zhong: Optimizing web search using social annotations. In: *Proceedings of the 16th international conference on World Wide Web*. New York, NY, USA : ACM, 2007 (WWW '07), 501–510
- [Bates 1990] BATES, Marcia J.: Where should the person stop and the information search interface start? In: *Information Processing & Management* 26 (1990), Nr. 5, 575–591. – DOI 10.1016/0306-4573(90)90103-9
- [Bates 2002] BATES, Marcia J.: Speculations on Browsing, Directed Searching, and Linking in Relation to the Bradford Distribution. In: BRUCE, Harry (Hrsg.) ; FIDEL, Raya (Hrsg.) ; INGWERSEN, Peter (Hrsg.) ; VAKKARI, Pertti (Hrsg.): *Emerging Frameworks and Methods: Proceedings of the Fourth International Conference on Conceptions of Library and Information Science (CoLIS 4)*. Greenwood Village, Colo. : Libraries Unlimited, 2002, S. 137–150
- [Beel u. Gipp 2009] BEEL, Jöran ; GIPP, Bela: Google Scholar's Ranking Algorithm: An Introductory Overview. In: *Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09)*, International Society for Scientometrics and Informetrics, Juli 2009, S. 230–241
- [Behrens-Neumann 2012] BEHRENS-NEUMANN, Renate: Das Projekt Gemeinsame Normdatei (GND). In: *Zeitschrift für Bibliothekswesen und Bibliographie* 59 (2012), Nr. 2, S. 96–100
- [Belkin 1980] BELKIN, Nicholas J.: Anomalous states of knowledge as a basis for information retrieval. In: *Canadian Journal of Information Science* (1980), Nr. 5, S. 133–143
- [Benz u. a. 2010] BENZ, Dominik ; HOTH, Andreas ; JÄSCHKE, Robert ; KRAUSE, Beate ; MITZLAFF, Folke ; SCHMITZ, Christoph ; STUMME, Gerd: The social bookmark and publication management system bibsonomy. In: *The VLDB Journal* 19 (2010), Dezember, Nr. 6, 849–875. – DOI 10.1007/s00778-010-0208-4

- [Berners-Lee u. a. 2001] BERNERS-LEE, Tim ; HENDLER, James ; LASSILA, Ora: The Semantic Web. In: *Scientific American* 284 (2001), Mai, Nr. 5, S. 34–43
- [Bertram u. Umlauf 2012] BERTRAM, Jutta ; UMLAUF, Konrad: Thesaurus. In: UMLAUF, Konrad (Hrsg.) ; GRADMANN, Stefan (Hrsg.): *Lexikon der Bibliotheks- und Informationswissenschaft (LBI)* Bd. 7. Stuttgart : Anton Hiersemann Verlag, 2012. – ISBN 978-3-7772-0922-7
- [Bhokal u. a. 2007] BHOGAL, J. ; MACFARLANE, A. ; SMITH, P.: A review of ontology based query expansion. In: *Information Processing & Management* 43 (2007), Nr. 4, 866–886. – DOI 10.1016/j.ipm.2006.09.003
- [Białecki u. a. 2012] BIAŁECKI, Andrzej ; MUIR, Robert ; INGERSOLL, Grant: Apache Lucene 4. In: TROTMAN, Andrew (Hrsg.) ; CLARKE, Charles L. A. (Hrsg.) ; OUNIS, Iadh (Hrsg.) ; CULPEPPER, J. S. (Hrsg.) ; CARTRIGHT, Marc-Allen (Hrsg.) ; GEVA, Shlomo (Hrsg.): *Proceedings of the SIGIR 2012 Workshop on Open Source Information Retrieval*. Portland, Oregon, USA : Department of Computer Science, University of Otago, New Zealand, August 2012. – ISBN 978-0-473-22026-6, 17–24
- [Björneborn u. Ingwersen 2004] BJÖRNEBORN, Lennart ; INGWERSEN, Peter: Toward a basic framework for webometrics. In: *Journal of the American Society for Information Science and Technology* 55 (2004), Dezember, Nr. 14, 1216–1227. – DOI 10.1002/asi.20077
- [Blair 2002] BLAIR, David C.: The challenge of commercial document retrieval, part II: a strategy for document searching based on identifiable document partitions. In: *Information Processing & Management* 38 (2002), März, Nr. 2, 293–304. – DOI 10.1016/S0306-4573(01)00025-5
- [Blair 2003] BLAIR, David C.: Information retrieval and the philosophy of language. In: *Annual Review of Information Science and Technology* 37 (2003), Nr. 1, 3–50. – DOI 10.1002/aris.1440370102
- [te Boekhorst u. a. 2003] BOEKHORST, Peter te ; KAYSS, Matthias ; POLL, Roswitha: Nutzungsanalyse des Systems der überregionalen Literatur- und Informationsversorgung: Teil I: Informationsverhalten und Informationsbedarf der Wissenschaft / Universitäts- und Landesbibliothek Münster, infas Institut für angewandte Sozialwissenschaft GmbH. [http://www.dfg.de/download/pdf/foerderung/programme/lis/ssg\\_bericht\\_teil\\_1.pdf](http://www.dfg.de/download/pdf/foerderung/programme/lis/ssg_bericht_teil_1.pdf). 2003. – Forschungsbericht
- [Borlund 2003] BORLUND, Pia: The concept of relevance in IR. In: *Journal of the American Society for Information Science and Technology* 54 (2003), Nr. 10, 913–925. – DOI 10.1002/asi.10286
- [Bradford 1934] BRADFORD, Samuel C.: Sources of information on specific subjects. In: *Engineering* (1934), Nr. 137, 85–86. <http://jis.sagepub.com/content/10/4/176.short?rss=1&ssource=mfc>

- [Büchler 2006] BÜCHLER, Marco: *Flexibles Berechnen von Kookkurrenzen auf strukturierten und unstrukturierten Daten*, Universität Leipzig, Diplomarbeit, Juni 2006. <http://asv.informatik.uni-leipzig.de/publication/file/111/dambuechler-public.pdf>
- [Buckland u. a. 1999] BUCKLAND, Michael ; CHEN, Aitao ; CHEN, Hui-Min ; KIM, Youngin ; LAM, Byron ; LARSON, Ray ; NORGARD, Barbara ; PURAT, Jacek ; GEY, Frederic: Mapping Entry Vocabulary to Unfamiliar Metadata Vocabularies. In: *D-Lib Magazine* 5 (1999), Januar, Nr. 1. – DOI 10.1045/january99-buckland
- [Buckley 2009] BUCKLEY, Chris: Why current IR engines fail. In: *Information Retrieval* 12 (2009), Dezember, Nr. 6, 652–665. – DOI 10.1007/s10791-009-9103-2
- [Buckley u. Voorhees 2000] BUCKLEY, Chris ; VOORHEES, Ellen M.: Evaluating evaluation measure stability. In: *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA : ACM, 2000 (SIGIR '00). – ISBN 1-58113-226-3, S. 33–40
- [Buckley u. Voorhees 2004] BUCKLEY, Chris ; VOORHEES, Ellen M.: Retrieval evaluation with incomplete information. In: *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA : ACM, 2004 (SIGIR '04). – ISBN 1-58113-881-4, 25–32
- [Büttcher u. a. 2007] BÜTTCHER, Stefan ; CLARKE, Charles L. A. ; YEUNG, Peter C. K. ; SOBOROFF, Ian: Reliable information retrieval evaluation with incomplete and biased judgements. In: *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA : ACM, 2007 (SIGIR '07). – ISBN 978-1-59593-597-7, 63–70
- [Calhoun u. a. 2009] CALHOUN, Karen ; CANTRELL, Joanne ; GALLAGHER, Peggy ; HAWK, Janet: Online Catalogs: What Users and Librarians Want / OCLC. <http://www.oclc.org/reports/onlinecatalogs/default.htm>. Dublin, Ohio, USA, 2009 (311870930). – OCLC Report
- [Chapelle u. a. 2011] CHAPELLE, Olivier ; CHANG, Yi ; LIU, Tie-Yan: Future directions in learning to rank. In: *Journal of Machine Learning Research - Proceedings Track 14* (2011), S. 91–100
- [Chowdhury 2010] CHOWDHURY, Gobinda G.: *Introduction to modern information retrieval*. 3rd ed. London : Facet, 2010. – ISBN 9781856046947
- [Cilibrasi u. Vitányi 2010] CILIBRASI, Rudi ; VITÁNYI, Paul M. B.: Normalized Web Distance and Word Similarity. In: INDURKHYA, Nitin (Hrsg.) ; DAMERAU, Fred J. (Hrsg.): *Handbook of Natural Language Processing*. Second Edition. Boca Raton, FL : CRC Press, Taylor and Francis Group, 2010. – ISBN 978-1420085921
- [Clauset 2007] CLAUSET, Aaron: *Power laws and all that jazz*. [http://www.cs.unm.edu/~aaron/blog/archives/2007/06/power\\_laws\\_and.htm](http://www.cs.unm.edu/~aaron/blog/archives/2007/06/power_laws_and.htm)

- [Clauset u. a. 2009] CLAUSET, Aaron ; SHALIZI, Cosma R. ; NEWMAN, M. E. J.: Power-Law Distributions in Empirical Data. In: *Society for Industrial and Applied Mathematics Review* 51 (2009), Nr. 4, 661–703. – DOI 10.1137/070710111
- [Cleverdon 1960] CLEVERDON, Cyril W.: The ASLIB Cranfield Research Project on the Comparative Efficiency of Indexing Systems. In: *Aslib Proceedings* 12 (1960), Nr. 12, 421–431. – DOI 10.1108/eb049778
- [Connaway u. Dickey 2010] CONNAWAY, Lynn S. ; DICKEY, Timothy J.: The digital information seeker: Findings from selected OCLC, RIN and JISC user behaviour projects / OCLC Research. <http://www.jisc.ac.uk/publications/reports/2010/digitalinformationseekers.aspx>. 2010. – Forschungsbericht
- [Crook u. a. 2009] CROOK, Thomas ; FRASCA, Brian ; KOHAVI, Ron ; LONGBOTHAM, Roger: Seven pitfalls to avoid when running controlled experiments on the web. In: *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (2009), S. 1105–1114. – DOI 10.1145/1557019.1557139
- [Csányi u. Szendroblaci 2004] CSÁNYI, Gábor ; SZENDROBLACI, Balázs: Structure of a large social network. In: *Physical Review E* 69 (2004), März, Nr. 3, 036131. – DOI 10.1103/PhysRevE.69.036131
- [Cummins u. a. 2011] CUMMINS, Ronan ; LALMAS, Mounia ; O’RIORDAN, Colm: The limits of retrieval effectiveness. In: *Proceedings of the 33rd European conference on Advances in information retrieval*. Berlin, Heidelberg : Springer, 2011 (ECIR’11). – ISBN 978-3-642-20160-8, 277–282
- [Day 2001] DAY, J: The quest for information: a guide to searching the Internet. In: *The Journal of Contemporary Dental Practice* 2 (2001), November, Nr. 4, S. 33–43. – PMID: 12167919
- [Demartini u. a. 2012] DEMARTINI, Gianluca ; MIKA, Peter ; TRAN, Thanh ; VRIES, Arjen P.: *From Expert Finding to Entity Search on the Web - Tutorial at ECIR 2012*. <http://diuf.unifr.ch/main/xi/EntitySearchTutorial>
- [Digital Library Federation 1998] DIGITAL LIBRARY FEDERATION: *A working definition of digital library*. <http://old.diglib.org/about/dldefinition.htm>. <http://old.diglib.org/about/dldefinition.htm>
- [Doan u. a. 2011] DOAN, Anhai ; RAMAKRISHNAN, Raghu ; HALEVY, Alon Y.: Crowdsourcing systems on the World-Wide Web. In: *Communications of the ACM* 54 (2011), April, Nr. 4, 86–96. – DOI 10.1145/1924421.1924442
- [Efthimiadis 1996] EFTHIMIADIS, Efthimis N.: Query expansion. In: *Annual Review of Information Systems and Technology (ARIST)* 31 (1996), S. 121–187
- [Egghe 2005] EGGHE, Leo: *Power Laws in the Information Production Process: Lotkaian Infometrics*. Oxford : Elsevier, 2005 (Library and Information Science Series). – ISBN 0 12 088753 3

- [Egghe 2009] EGGHE, Leo: Lotkaian informetrics and applications to social networks. In: *Bulletin of the Belgian Mathematical Society - Simon Stevin* 16 (2009), Nr. 4, 689–703. <http://hdl.handle.net/1942/9281>
- [Egghe u. a. 2011] EGGHE, Leo ; GUNS, Raf ; ROUSSEAU, Ronald: Thoughts on uncitedness: Nobel laureates and Fields medalists as case studies. In: *Journal of the American Society for Information Science and Technology* 62 (2011), August, Nr. 8, 1637–1644. – DOI 10.1002/asi.21557
- [Egghe u. Rousseau 2006] EGGHE, Leo ; ROUSSEAU, Ronald: An informetric model for the Hirsch-index. In: *Scientometrics* 69 (2006), April, Nr. 1, S. 121–129. – DOI 10.1007/s11192-006-0143-8
- [Ellis 1992] ELLIS, David: The physical and cognitive paradigms in information retrieval research. In: *Journal of Documentation* 48 (1992), Dezember, Nr. 1, 45–64. – DOI 10.1108/eb026889
- [Ellis 2010] ELLIS, John: *Will Google Instant Kill The Long Tail?* <http://searchengineland.com/will-google-instant-kill-the-long-tail-50110>
- [Elsevier 2012] ELSEVIER: *Sorting search results by date or relevance.* [http://help.sciencedirect.com/flare/sdhelp\\_Left.htm#CSHID=browse\\_open\\_doc.htm|StartTopic=Content%2Fbrowse\\_open\\_doc.htm|SkinName=svs\\_SD](http://help.sciencedirect.com/flare/sdhelp_Left.htm#CSHID=browse_open_doc.htm|StartTopic=Content%2Fbrowse_open_doc.htm|SkinName=svs_SD)
- [Emrich 2011a] EMRICH, Kerstin: International Standard Serial Number. In: UMLAUF, Konrad (Hrsg.) ; GRADMANN, Stefan (Hrsg.): *Lexikon der Bibliotheks- und Informationswissenschaft (LBI)* Bd. 6. Stuttgart : Anton Hiersemann Verlag, 2011. – ISBN 978-3-7772-0922-7, S. 461
- [Emrich 2011b] EMRICH, Kerstin: Internationale Standard-Buchnummer. In: UMLAUF, Konrad (Hrsg.) ; GRADMANN, Stefan (Hrsg.): *Lexikon der Bibliotheks- und Informationswissenschaft (LBI)* Bd. 6. Stuttgart : Anton Hiersemann Verlag, 2011. – ISBN 978-3-7772-0922-7, S. 464f
- [Ernde u. a. 2010] ERNDE, Benjamin ; LEBEL, Michael ; THIELE, Christoph ; HOLD, Alexander ; NAUMANN, Felix ; BARCZYŃSKI, Wojciech ; BRAUER, Falk: ECIR - a Lightweight Approach for Entity-centric Information Retrieval. In: *Proceedings of the Eighteenth Text REtrieval Conference (TREC 2010)*, 2010
- [Fain u. Pedersen 2006] FAIN, Daniel C. ; PEDERSEN, Jan O.: Sponsored search: A brief history. In: *Bulletin of the American Society for Information Science and Technology* 32 (2006), Dezember, Nr. 2, 12–13. – DOI 10.1002/bult.1720320206
- [Fang u. a. 2011] FANG, Hui ; TAO, Tao ; ZHAI, Chengxiang: Diagnostic Evaluation of Information Retrieval Models. In: *ACM Transactions on Information Systems* 29 (2011), April, Nr. 2, 7:1–7:42. – DOI 10.1145/1961209.1961210

- [Ferber 1996] FERBER, Reginald: Using Co-occurrence Data for Query Expansion: Wrong Paradigm or Wrong Formulas? <http://www.muenster.de/~ferber/homepage/pdf-ps/using-cooc.pdf>. 1996. – Forschungsbericht
- [Ferber 2003] FERBER, Reginald: *Information Retrieval: Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web*. Heidelberg : dpunkt Verlag, 2003. – ISBN 3-89864-213-5
- [Fox u. a. 2012] FOX, Edward A. ; GONCALVES, Marcos A. ; SHEN, Rao: *Theoretical Foundations for Digital Libraries: The 5S Approach*. Morgan & Claypool Publishers, 2012. – ISBN 9781608459100
- [Freeman 1977] FREEMAN, Linton: A Set of Measures of Centrality Based on Betweenness. In: *Sociometry* 40 (1977), März, Nr. 1, S. 35–41
- [Fronczak u. a. 2007] FRONCZAK, Piotr ; FRONCZAK, Agata ; HOLYST, Janusz A.: Analysis of scientific productivity using maximum entropy principle and fluctuation-dissipation theorem. In: *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics* 75 (2007), Februar, Nr. 2 Pt 2, S. 026103. – PMID: 17358390
- [Fu u. Aliferis 2010] FU, Lawrence D. ; ALIFERIS, Constantin F.: Using content-based and bibliometric features for machine learning models to predict citation counts in the biomedical literature. In: *Scientometrics* 85 (2010), Februar, Nr. 1, 257–270. – DOI 10.1007/s11192-010-0160-5
- [Fuhr 2004] FUHR, Norbert: Theorie des Information Retrieval I: Modelle. In: KUHLEN, Rainer (Hrsg.) ; SEEGER, T. (Hrsg.) ; STRAUCH, D. (Hrsg.): *Grundlagen der praktischen Information und Dokumentation*. 5. München : Saur, 2004, S. 207–214
- [Fuhr 2006] FUHR, Norbert: Information Retrieval Skriptum zur Vorlesung im SS 06 / Universität Duisburg. [http://www.is.informatik.uni-duisburg.de/courses/ir\\_ss06/foalien/irskall.pdf](http://www.is.informatik.uni-duisburg.de/courses/ir_ss06/foalien/irskall.pdf). 2006. – Vorlesungsskript
- [Fuhr 2008] FUHR, Norbert: A probability ranking principle for interactive information retrieval. In: *Information Retrieval* 11 (2008), Februar, Nr. 3, S. 251–265. – DOI 10.1007/s10791-008-9045-0
- [Geyer-Schulz u. a. 2003] GEYER-SCHULZ, Andreas ; NEUMANN, Andreas ; THEDE, Anke: Others Also Use: A Robust Recommender System for Scientific Libraries. In: KOCH, Traugott (Hrsg.) ; SØLVBERG, Ingeborg (Hrsg.): *Research and Advanced Technology for Digital Libraries* Bd. 2769. Berlin, Heidelberg : Springer, 2003 (Lecture Notes in Computer Science). – ISBN 978-3-540-40726-3, S. 113–125
- [Glänzel 2003] GLÄNZEL, Wolfgang: *Bibliometrics as a Research Field*. [http://nsdl.niscair.res.in/bitstream/123456789/968/1/Bib\\_Module\\_KUL.pdf](http://nsdl.niscair.res.in/bitstream/123456789/968/1/Bib_Module_KUL.pdf)
- [Glänzel 2007] GLÄNZEL, Wolfgang: Characteristic scores and scales: A bibliometric analysis of subject characteristics based on long-term citation observation. In: *Journal of Informetrics* 1 (2007), Januar, Nr. 1, 92–102. – DOI 10.1016/j.joi.2006.10.001

- [Glänzel u. a. 2008] GLÄNZEL, Wolfgang ; JANSSENS, Frizo ; THIJS, Bart: A comparative analysis of publication activity and citation impact based on the core literature in bioinformatics. In: *Scientometrics* 79 (2008), Oktober, Nr. 1, S. 109–129. – DOI 10.1007/s11192-009-0407-1
- [Goecks u. Shavlik 2000] GOECKS, Jeremy ; SHAVLIK, Jude: Learning users' interests by unobtrusively observing their normal behavior. In: *IUI '00: Proceedings of the 5th international conference on Intelligent user interfaces*. New York, NY, USA : ACM, 2000. – ISBN 1-58113-134-8, S. 129–132
- [Gollapalli u. a. 2011] GOLLAPALLI, Sujatha D. ; MITRA, Prasenjit ; GILES, C. L.: Ranking authors in digital libraries. In: *Proceedings of the 11th annual international ACM/IEEE Joint Conference on Digital Libraries*. New York, NY, USA : ACM, 2011 (JCDL '11). – ISBN 978-1-4503-0744-4, 251–254
- [Google Scholar 2011] GOOGLE SCHOLAR: *About Google Scholar*. <http://scholar.google.com/scholar/about.html?hl=en>
- [Gottron 2010] GOTTRON, Thomas: Information Retrieval / Institut für Informatik, Universität Mainz. <http://www1.informatik.uni-mainz.de/lehre/ir/skript-rose-10/IR-SoSe10.pdf>. 2010. – Vorlesungsskript
- [Hagen 2005] HAGEN, Robert: *Linguistische Pragmatik: Eine Übersicht*. <http://userpage.fu-berlin.de/~hagen66/01/ling-pragma.htm>
- [Hamers u. a. 1989] HAMERS, Lieve ; HEMERYCK, Yves ; HERWEYERS, Guido ; JANSSEN, Marc ; KETERS, Hans ; ROUSSEAU, Ronald ; VANHOUTTE, André: Similarity measures in scientometric research: The Jaccard index versus Salton's cosine formula. In: *Information Processing & Management* 25 (1989), Nr. 3, 315–318. – DOI 10.1016/0306-4573(89)90048-4
- [Hammarfelt 2012] HAMMARFELT, Björn: *Following the Footnotes : A Bibliometric Analysis of Citation Patterns in Literary Studies*. Uppsala, Uppsala University, Department of ALM, Diss., 2012. <http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-170504>. – © Björn Hammarfelt 2012
- [Hardy 2010] HARDY, Michael: Pareto's Law. In: *The Mathematical Intelligencer* 32 (2010), September, Nr. 3, S. 38–43. – DOI 10.1007/s00283-010-9159-2
- [Harzing 2010] HARZING, Anne-Wil: *The Publish or perish book : your guide to effective and responsible citation analysis*. Melbourne : Tarma Software Research Pty Ltd, 2010. – ISBN 9780980848519 0980848512 9780980848526 0980848520
- [Hätscher u. a. 2007] HÄTSCHER, Petra ; OBERLÄNDER, Anja ; KOHL-FREY, Oliver: *Perspektiven der Literatur- und Informationsversorgung : Ergebnisse der Befragung der Wissenschaftlerinnen und Wissenschaftler der Universität Konstanz 2007*. 2007 (Bibliothek aktuell Sonderheft 16). <http://kops.ub.uni-konstanz.de/handle/urn:nbn:de:bsz:352-opus-39615>



- [Havemann 2009] HAVEMANN, Frank: *Einführung in die Bibliometrie*. Berlin : Gesellschaft für Wissenschaftsforschung, 2009 <http://edoc.hu-berlin.de/oa/books/reMKADKkid1Wk/PDF/20uf7RZtM6ZJk.pdf>. – ISBN 978–3–934682–46–7
- [He 2009] HE, Zi-Lin: International collaboration does not have greater epistemic authority. In: *Journal of the American Society for Information Science and Technology* 60 (2009), Oktober, Nr. 10, 2151–2164. – DOI 10.1002/asi.v60:10
- [Head u. Eisenberg 2009] HEAD, Alison J. ; EISENBERG, Michael B.: Lessons Learned: How College Students Seek Information in the Digital Age / University of Washington's Information School. [http://projectinfolit.org/pdfs/PIL\\_Fall2009\\_finalv\\_YR1\\_12\\_2009v2.pdf](http://projectinfolit.org/pdfs/PIL_Fall2009_finalv_YR1_12_2009v2.pdf). 2009. – Forschungsbericht
- [Heck u. a. 2011] HECK, Tamara ; PETERS, Isabella ; STOCK, Wolfgang G.: Testing collaborative filtering against co-citation analysis and bibliographic coupling for academic author recommendation. In: *Proceedings of the 3rd ACM RecSys'11 Workshop on Recommender Systems and the Social Web*, 2011, S. 16–23
- [Hellweg u. a. 2001] HELLWEG, Heiko ; KRAUSE, Jürgen ; MANDL, Thomas ; MARX, Jutta ; MÜLLER, Matthias ; MUTSCHKE, Peter ; STRÖTGEN, Robert: Treatment of Semantic Heterogeneity in Information Retrieval / IZ-Sozialwissenschaften. [http://www.thesis.org/fileadmin/upload/forschung/publikationen/thesis\\_reihen/iz\\_arbeitsberichte/ab\\_23.pdf](http://www.thesis.org/fileadmin/upload/forschung/publikationen/thesis_reihen/iz_arbeitsberichte/ab_23.pdf). Bonn, 2001 (23). – IZ-Arbeitsbericht
- [Hess 2012] HESS, Kirk: Discovering Digital Library User Behavior with Google Analytics. In: *The Code4Lib Journal* (2012), Juni, Nr. 17. <http://journal.code4lib.org/articles/6942>
- [Hienert u. a. 2011a] HIENERT, Daniel ; SCHAER, Philipp ; SCHAIBLE, Johann ; MAYR, Philipp: A Novel Combined Term Suggestion Service for Domain-Specific Digital Libraries. In: GRADMANN, Stefan (Hrsg.) ; BORRI, Francesca (Hrsg.) ; MEGHINI, Carlo (Hrsg.) ; SCHULDT, Heiko (Hrsg.): *Research and Advanced Technology for Digital Libraries* Bd. 6966. Berlin, Heidelberg : Springer, 2011 (Lecture Notes in Computer Science). – ISBN 978–3–642–24468–1, S. 192–203
- [Hienert u. a. 2011b] HIENERT, Daniel ; ZAPILKO, Benjamin ; SCHAER, Philipp ; MATHIAK, Brigitte: VIZGR: Combining Data on a Visual Level. In: CORDEIRO, José (Hrsg.) ; FILIPE, Joaquim (Hrsg.): *Proceedings of the 7th International Conference on Web Information Systems and Technologies (WEBIST)*, SciTePress, 2011. – ISBN 978–989–8425–51–5, 202–211
- [Hienert u. a. 2011c] HIENERT, Daniel ; ZAPILKO, Benjamin ; SCHAER, Philipp ; MATHIAK, Brigitte: Vizgr: Linking Data in Visualizations. In: CORDEIRO, José (Hrsg.) ; FILIPE, Joaquim (Hrsg.): *WEBIST 2011 Selected and Revised Papers* Bd. 101. Berlin, Heidelberg : Springer, 2011 (Lecture Notes in Business Information Processing), S. 177–191

- [Hirsch 2005] HIRSCH, J. E.: An index to quantify an individual's scientific research output. In: *Proceedings of the National Academy of Sciences of the United States of America* 102 (2005), November, Nr. 46, 16569–16572. – DOI 10.1073/pnas.0507655102
- [Hitzler u. a. 2008] HITZLER, Pascal ; KRÖTZSCH, Markus ; RUDOLPH, Sebastian ; SURE, York: *Semantic Web : Grundlagen*. Berlin : Springer, 2008. – ISBN 978-3-540-33993-9
- [Hjørland 2000] HJØRLAND, Birger: Library and information science: practice, theory, and philosophical basis. In: *Information Processing and Management* 36 (2000), Nr. 3, S. 501–531. – DOI 10.1016/S0306-4573(99)00038-2
- [Hjørland 2006] HJØRLAND, Birger: *Polyrepresentation*. [http://www.iva.dk/bh/Lifeboat\\_K0/CONCEPTS/polyrepresentation.htm](http://www.iva.dk/bh/Lifeboat_K0/CONCEPTS/polyrepresentation.htm)
- [Hjørland u. Nicolaisen 2005] HJØRLAND, Birger ; NICOLAISEN, Jeppe: Bradford's Law of Scattering: Ambiguities in the Concept of "Subject". In: CRESTANI, Fabio (Hrsg.) ; RUTHVEN, Ian (Hrsg.): *Context: Nature, Impact, and Role* Bd. 3507. Berlin, Heidelberg : Springer, Januar 2005 (Lecture Notes in Computer Science). – ISBN 978-3-540-26178-0, S. 96–106
- [Horvitz 1999] HORVITZ, Eric: Principles of mixed-initiative user interfaces. In: *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA : ACM, 1999. – ISBN 0-201-48559-1, S. 159–166
- [Hotho u. a. 2006] HOTHO, Andreas ; JÄSCHKE, Robert ; SCHMITZ, Christoph ; STUMME, Gerd: Information Retrieval in Folksonomies: Search and Ranking. In: SURE, York (Hrsg.) ; DOMINGUE, John (Hrsg.): *The Semantic Web: Research and Applications* Bd. 4011. Heidelberg : Springer, Juni 2006 (Lecture Notes in Artificial Intelligence). – ISBN 978-3-540-34544-2, 411–426
- [Huberman u. Adamic 1999] HUBERMAN, Bernardo A. ; ADAMIC, Lada A.: Internet: Growth dynamics of the World-Wide Web. In: *Nature* 401 (1999), Nr. 6749, 131. – DOI 10.1038/43604
- [Ingwersen 1996] INGWERSEN, Peter: Cognitive Perspectives of Information Retrieval Interaction: Elements of a Cognitive IR Theory. In: *Journal of Documentation* 52 (1996), Nr. 1, S. 3–50. – DOI 10.1108/eb026960
- [Ingwersen 2012a] INGWERSEN, Peter: *Bibliometrics/Scientometrics and IR A methodological bridge through visualization*. <http://www.promise-noe.eu/documents/10156/028a48d8-4ba8-463c-acbc-db75db67ea4d>
- [Ingwersen 2012b] INGWERSEN, Peter: Citations and references as keys to relevance ranking in interactive IR. In: *Proceedings of the 4th Information Interaction in Context Symposium*. New York, NY, USA : ACM, 2012 (IIIX '12). – ISBN 978-1-4503-1282-0, S. 1–1. – Video des Vortrags: <http://iiix2012.cs.ru.nl/videos.html>

- [Ingwersen 2012c] INGWERSEN, Peter: *Scientometric Indicators and Webometrics and the Polyrepresentation Principle in Information Retrieval*. Ess Ess Publication, 2012. – ISBN 9788170006572
- [Ingwersen u. Järvelin 2005] INGWERSEN, Peter ; JÄRVELIN, Kalervo: *The turn - integration of information seeking and retrieval in context*. Dordrecht : Springer, 2005. – ISBN 9781402038518
- [Ingwersen u. Järvelin 2007] INGWERSEN, Peter ; JÄRVELIN, Kalervo: On the Holistic Cognitive Theory for Information Retrieval: Drifting Outside the Border of the Laboratory Framework. In: DOMINIC, S. (Hrsg.) ; KISS, F. (Hrsg.): *Studies in the Theory of Information Retrieval (ICTIR 2007)*. Budapest, Hungary: Foundation for Information Society, 2007, 135–147
- [Iribarren-Maestro u. a. 2009] IRIBARREN-MAESTRO, Isabel ; LASCURAIN-SÁNCHEZ, Mariá L. ; SANZ-CASADO, Elias: The Use of Bibliometric Techniques in Evaluating Social Sciences and Humanities. <http://www8.umu.se/inforsk/Bibexcel/ollepersson60.pdf>. In: ASTRÖM, Fredrik (Hrsg.) ; DANELL, Richard (Hrsg.) ; LARSEN, Birger (Hrsg.) ; SCHNEIDER, Jesper W. (Hrsg.): *Celebrating Scholarly Communication Systems: A Festschrift for Olle Persson at his 60th Birthday*. ISSI, 2009 (E-Newsletter of the International Society for Scientometrics and Informetrics 05-S), 25–37
- [Jaccard 1901] JACCARD, Paul: Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. In: *Bulletin de la Société Vaudoise des Sciences Naturelles* (1901), Nr. 37, S. 241–272
- [Jacsó 2008] JACSÓ, Péter: Google Scholar revisited. In: *Online Information Review* 32 (2008), Februar, Nr. 1, 102–114. – DOI 10.1108/14684520810866010
- [Jameson 2003] JAMESON, Anthony: Adaptive interfaces and agents. In: JACKO, Julie A. (Hrsg.) ; SEARS, Andrew (Hrsg.): *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*. Mahwah, NJ, USA : Lawrence Erlbaum Associates, Inc., 2003. – ISBN 0–8058–3838–4, S. 305–330
- [Jameson 2006] JAMESON, Anthony: Adaptive Interfaces and Agents. In: JACKO, Julie A. (Hrsg.) ; SEARS, Andrew (Hrsg.): *Human-computer interaction handbook*. 2nd ed. Mahwah, NJ : Erlbaum, 2006
- [Jansen u. a. 2007] JANSEN, Bernard J. ; SPINK, Amanda ; KOSHMAN, Sherry: Web searcher interaction with the Dogpile.com metasearch engine. In: *Journal of the American Society for Information Science and Technology* 58 (2007), März, 744–755. – DOI 10.1002/asi.v58:5
- [Järvelin u. Kekäläinen 2002] JÄRVELIN, Kalervo ; KEKÄLÄINEN, Jaana: Cumulated gain-based evaluation of IR techniques. In: *ACM Transactions on Information Systems* 20 (2002), Oktober, Nr. 4, 422–446. – DOI 10.1145/582415.582418

- [Jones u. a. 2000] JONES, K. S. ; WALKER, S. ; ROBERTSON, S. E.: A probabilistic model of information retrieval: development and comparative experiments. In: *Information Processing & Management* 36 (2000), November, Nr. 6, 779–808. – DOI 10.1016/S0306-4573(00)00015-7
- [Jordy u. a. 1999] JORDY, Matthew L. ; MCGRATH, Eileen L. ; RUTLEDGE, John B.: Book Reviews As a Tool for Assessing Publisher Reputation. In: *College & Research Libraries* 60 (1999), März, Nr. 2, 132–142. <http://crl.acrl.org/content/60/2/132.abstract>
- [Juran 1954] JURAN, Joseph M.: Universals in management planning and controlling. In: *Management Review* 43 (1954), Nr. 11, S. 748–761
- [Kahn u. Cerf 1988] KAHN, Robert E. ; CERF, Vinton G.: The Digital Library Project Volume I: The World of Knowbots (DRAFT): An Open Architecture For a Digital Library System and a Plan For Its Development / Corporation for National Research Initiatives. <http://hdl.handle.net/4263537/2091>. 1988. – Forschungsbericht
- [Kamps 2004] KAMPS, Jaap: Improving Retrieval Effectiveness by Reranking Documents Based on Controlled Vocabulary. In: MCDONALD, Sharon (Hrsg.) ; TAIT, John (Hrsg.): *Advances in Information Retrieval* Bd. 2997. Berlin, Heidelberg : Springer, 2004 (Lecture Notes in Computer Science). – ISBN 978-3-540-21382-6, S. 283–295
- [Karatzoglou u. a. 2006] KARATZOGLOU, Alexandros ; MEYER, David ; HORNIK, Kurt: Support Vector Machines in R. In: *Journal of Statistical Software* 15 (2006), Nr. 9. <http://ideas.repec.org/a/jss/jstsof/15i09.html>
- [Kay 1990] KAY, Alan: User Interface: A Personal View. In: LAUREL, B. (Hrsg.): *The Art of Human-Computer Interface Design*. Reading, Mass : Addison-Wesley, 1990
- [Keidel u. Müller 2010] KEIDEL, Petra ; MÜLLER, Christian: Der Journal Recommender an der TH Wildau. In: MITTERMAIER, Bernhard (Hrsg.): *eLibrary - den Wandel gestalten*, Forschungszentrum Jülich GmbH, Zentralbibliothek, Verlag, 2010 (Schriften des Forschungszentrums Jülich / Reihe Bibliothek/Library). – ISBN 978-3-89336-668-2, S. 95–108
- [Keil 2012] KEIL, Stefan: Terminologie Mapping: Grundlagen und aktuelle Normungsvorhaben. In: *Information - Wissenschaft und Praxis* 63 (2012), Nr. 1, 45–55. – DOI 10.1515/iwp-2012-0004
- [Kelly 2009] KELLY, Diane: Methods for Evaluating Interactive Information Retrieval Systems with Users. In: *Foundations and Trends in Information Retrieval* 3 (2009), Nr. 1–2, 1–224. – DOI 10.1561/15000000012
- [Kleinberg 1999] KLEINBERG, Jon M.: Authoritative sources in a hyperlinked environment. In: *J. ACM* 46 (1999), September, Nr. 5, S. 604–632. – DOI 10.1145/324133.324140

- [Kluck 2003] KLUCK, Michael: The GIRT Data in the Evaluation of CLIR Systems - from 1997 Until 2003. In: PETERS, Carol (Hrsg.) ; GONZALO, Julio (Hrsg.) ; BRASCHLER, Martin (Hrsg.) ; KLUCK, Michael (Hrsg.): *Comparative Evaluation of Multilingual Information Access Systems* Bd. 3237. Berlin, Heidelberg : Springer, 2003 (Lecture Notes in Computer Science). – ISBN 3-540-24017-9, S. 376-390
- [Kluck u. Winter 2006] KLUCK, Michael ; WINTER, Marco: Topic-Entwicklung und Relevanzbewertung bei GIRT: ein Werkstattbericht. In: MANDL, Thomas (Hrsg.) ; WOMSER-HACKER, Christa (Hrsg.): *Effektive Information Retrieval Verfahren in Theorie und Praxis: ausgewählte und erweiterte Beiträge des Vierten Hildesheimer Evaluierungs- und Retrievalworkshop (HIER 2005), Hildesheim, 20.7.2005*. Bd. 45. Konstanz : UVK Verlagsgesellschaft, 2006 (Schriften zur Informationswissenschaft), S. 103-116
- [Krause 1996] KRAUSE, Jürgen: Informationserschließung und -bereitstellung zwischen Deregulation, Kommerzialisierung und weltweiter Vernetzung - Schalenmodell / IZ-Sozialwissenschaften. [http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/iz\\_arbeitsberichte/ab6.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/iz_arbeitsberichte/ab6.pdf). Bonn, 1996 (6). – IZ-Arbeitsbericht
- [Krause 2003] KRAUSE, Jürgen: Standardisierung von der Heterogenität her denken: Zum Entwicklungsstand Bilateraler Transferkomponenten für digitale Fachbibliotheken / IZ-Sozialwissenschaften. [http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/iz\\_arbeitsberichte/ab\\_28.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/iz_arbeitsberichte/ab_28.pdf). Bonn, 2003 (28). – IZ-Arbeitsbericht
- [Krause 2004] KRAUSE, Jürgen: Konkretes zur These, die Standardisierung von der Heterogenität her zu denken. In: *Zeitschrift für Bibliothekswesen und Bibliographie* 51 (2004), Nr. 2, S. 76-89
- [Krause 2006] KRAUSE, Jürgen: Shell Model, Semantic Web and Web Information Retrieval. In: HARMS, Ilse (Hrsg.) ; LUCKHARDT, Heinz-Dirk (Hrsg.) ; GIESSEN, Hans W. (Hrsg.): *Information und Sprache : Beiträge zu Informationswissenschaft, Computerlinguistik, Bibliothekswesen und verwandten Fächern ; Festschrift für Harald H. Zimmermann*. München : Saur, 2006. – ISBN 3-598-11754-X, S. 95 - 106
- [Krause 2008] KRAUSE, Jürgen: Semantic heterogeneity: comparing new semantic web approaches with those of digital libraries. In: *Library Review* 57 (2008), März, Nr. 3, 235-248. – DOI 10.1108/00242530810865501
- [Krause 2012] KRAUSE, Jürgen: Vascoda goes... In: PETROVIC, Otto (Hrsg.) ; SCHLÖGL, Christian (Hrsg.) ; REICHMANN, Gerhard (Hrsg.): *Informationswissenschaft. Begegnungen mit Wolf Rauch*. Wien, Köln, Weimar : Böhlau Verlag, 2012. – ISBN 978-3-205-78799-0, S. 181-194
- [Krause u. a. 2008] KRAUSE, Jürgen ; PETRAS, Vivien ; STEMPFHUBER, Maximilian: Retrieval-Mehrwertdienste zur Weiterentwicklung wissenschaftlicher Fachportale wie

- vascoda und sowiport. Suchexpandierung und Re-Ranking. Bonn, März 2008. – DFG-Projektantrag
- [Kriewel u. Fuhr 2010] KRIEWEL, Sascha ; FUHR, Norbert: Evaluation of an Adaptive Search Suggestion System. In: GURRIN, Cathal (Hrsg.) ; HE, Yulan (Hrsg.) ; KAZAI, Gabriella (Hrsg.) ; KRUSCHWITZ, Udo (Hrsg.) ; LITTLE, Suzanne (Hrsg.) ; ROELLEKE, Thomas (Hrsg.) ; RÜGER, Stefan (Hrsg.) ; RIJSBERGEN, Keith (Hrsg.): *Advances in Information Retrieval* Bd. 5993. Berlin, Heidelberg : Springer, 2010 (Lecture Notes in Computer Science). – ISBN 978-3-642-12274-3, S. 544–555
- [Kriewel u. a. 2004] KRIEWEL, Sascha ; KLAS, Claus-Peter ; SCHAEFER, Andre ; FUHR, Norbert: DAFFODIL - Strategic Support for User-Oriented Access to Heterogeneous Digital Libraries. In: *D-Lib Magazine* 10 (2004), Juni, Nr. 6. – DOI 10.1045/june2004-kriewel
- [Kruk u. McDaniel 2009] KRUK, Sebastian R. (Hrsg.) ; MCDANIEL, Bill (Hrsg.): *Semantic Digital Libraries*. Berlin, Heidelberg : Springer, 2009. – ISBN 978-3-540-85433-3
- [Kuhlen 1999] KUHLEN, Rainer: *Die Konsequenzen der Informationsassistenten. Was bedeutet informationelle Autonomie oder wie kann Vertrauen in elektronische Dienste in offenen Informationsmärkten gesichert werden?* Frankfurt : Suhrkamp-Verlag, 1999 (Suhrkamp Taschenbuch wissenschaft stw 1443)
- [Kunegis u. Preusse 2012] KUNEGIS, Jérôme ; PREUSSE, Julia: Fairness on the web: alternatives to the power law. In: *Proceedings of the 3rd Annual ACM Web Science Conference*. New York, NY, USA : ACM, 2012 (WebSci '12). – ISBN 978-1-4503-1228-8, 175–184
- [Kuropka 2005] KUROPKA, Dominik: Uselessness of simple co-occurrence measures for IF&IR - a linguistic point of view. In: *8th International Conference on Business Information Systems*. Poznań, Poland : Wydawnictwo Akademii Ekonomicznej w Poznaniu, 2005. – ISBN 83-7417-094-8, S. 147–156
- [Kürsten u. Eibl 2011] KÜRSTEN, Jens ; EIBL, Maximilian: A Large-Scale System Evaluation on Component-Level. In: CLOUGH, Paul (Hrsg.) ; FOLEY, Colum (Hrsg.) ; GURRIN, Cathal (Hrsg.) ; JONES, Gareth (Hrsg.) ; KRAAIJ, Wessel (Hrsg.) ; LEE, Hyowon (Hrsg.) ; MUDUCH, Vanessa (Hrsg.): *Advances in Information Retrieval* Bd. 6611. Berlin, Heidelberg : Springer, 2011 (Lecture Notes in Computer Science). – ISBN 978-3-642-20160-8, S. 679–682
- [Lagoze u. a. 2005] LAGOZE, Carl ; KRAFT, Dean B. ; PAYETTE, Sandy ; JESUROGA, Susan: What Is a Digital Library Anyway? In: *D-Lib Magazine* 11 (2005), November, Nr. 11. – DOI 10.1045/november2005-lagoze
- [Lange u. a. 2010] LANGE, Matthias ; SPIES, Karl ; BARGSTEN, Joachim ; HABERHAUER, Gregor ; KLAPPERSTÜCK, Matthias ; LEPS, Michael ; WEINEL, Christian ; WÜNSCHERS, Röbbbe ; WEISSBACH, Mandy ; STEIN, Jens ; SCHOLZ, Uwe: The LAILAPS search engine:

- relevance ranking in life science databases. In: *Journal of Integrative Bioinformatics* 7 (2010), Nr. 2, S. 110. – DOI 10.2390/biecoll-jib-2010-110. – PMID: 20134080
- [Larsen 2004] LARSEN, Birger: *References and citations in automatic indexing and retrieval systems - experiments with the boomerang effect*. Copenhagen, Denmark, Department of Information Studies, Royal School of Library and Information Science, PhD Thesis, 2004. [http://pure.iva.dk/files/31034810/birger\\_larsen\\_phd.pdf](http://pure.iva.dk/files/31034810/birger_larsen_phd.pdf)
- [Larsen 2005] LARSEN, Birger: Practical Implications of Handling Multiple Contexts in the Principle of Polyrepresentation. In: CRESTANI, Fabio (Hrsg.) ; RUTHVEN, Ian (Hrsg.): *Context: Nature, Impact, and Role* Bd. 3507. Berlin, Heidelberg : Springer, 2005 (Lecture Notes in Computer Science). – ISBN 978-3-540-26178-0, S. 20–31
- [Larsen u. Ingwersen 2006] LARSEN, Birger ; INGWERSEN, Peter: Using citations for ranking in digital libraries. In: *Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries - JCDL '06*. Chapel Hill, NC, USA, 2006, 370
- [Larsen u. a. 2006] LARSEN, Birger ; INGWERSEN, Peter ; KEKÄLÄINEN, Jaana: The polyrepresentation continuum in IR. In: *Proceedings of the 1st international conference on Information interaction in context - IiX*. New York, NY, USA : ACM, 2006, 88–96
- [Larsen u. a. 2009] LARSEN, Birger ; INGWERSEN, Peter ; LUND, Berit: Data fusion according to the principle of polyrepresentation. In: *Journal of the American Society for Information Science and Technology* 60 (2009), April, Nr. 4, 646–654. – DOI 10.1002/asi.21028
- [Lawrence 2003] LAWRENCE, Peter A.: The politics of publication. In: *Nature* 422 (2003), März, Nr. 6929, 259–261. – DOI 10.1038/422259a
- [Lewandowski 2005] LEWANDOWSKI, Dirk: *Web Information Retrieval Technologien zur Informationssuche im Internet*. Frankfurt am Main : DGI, 2005. – ISBN 9783925474552
- [Lewandowski 2009] LEWANDOWSKI, Dirk: Ranking library materials. In: *Library Hi Tech* 27 (2009), Nr. 4, 584–593. – DOI 10.1108/07378830911007682
- [Lewandowski 2011] LEWANDOWSKI, Dirk: Evaluierung von Suchmaschinen. In: LEWANDOWSKI, Dirk (Hrsg.): *Handbuch Internet-Suchmaschinen: Neue Entwicklungen in der Web-Suche* Bd. 2. 1. Aufl. Heidelberg : Akademische Verlagsgesellschaft Aka, 2011. – ISBN 9783898386517, S. 203–228
- [Liu 2007] LIU, Tie-Yan: Learning to Rank for Information Retrieval. In: *Foundations and Trends in Information Retrieval* 3 (2007), Nr. 3, 225–331. – DOI 10.1561/1500000016
- [Lotka 1926] LOTKA, Alfred J.: The frequency distribution of scientific productivity. In: *Journal of the Washington Academy of Sciences* 16 (1926), Nr. 12, S. 317–323
- [Lüke u. a. 2012] LÜKE, Thomas ; SCHAER, Philipp ; MAYR, Philipp: Improving Retrieval Results with Discipline-Specific Query Expansion. In: ZAPHIRIS, Panayiotis (Hrsg.) ;

- BUCHANAN, George (Hrsg.) ; RASMUSSEN, Edie (Hrsg.) ; LOIZIDES, Fernando (Hrsg.): *Theory and Practice of Digital Libraries* Bd. 7489. Berlin, Heidelberg : Springer, 2012 (Lecture Notes in Computer Science). – ISBN 978-3-642-33289-0, S. 408-413
- [Lykke u. a. 2010] LYKKE, Marianne ; LARSEN, Birger ; LUND, Haakon ; INGWERSEN, Peter: Developing a Test Collection for the Evaluation of Integrated Search. In: GURRIN, Cathal (Hrsg.) ; HE, Yulan (Hrsg.) ; KAZAI, Gabriella (Hrsg.) ; KRUSCHWITZ, Udo (Hrsg.) ; LITTLE, Suzanne (Hrsg.) ; ROELLEKE, Thomas (Hrsg.) ; RÜGER, Stefan (Hrsg.) ; RIJSBERGEN, Keith (Hrsg.): *Advances in Information Retrieval* Bd. 5993. Berlin, Heidelberg : Springer, 2010 (Lecture Notes in Computer Science). – ISBN 978-3-642-12274-3, S. 627-630
- [Maes 1994] MAES, Pattie: Agents that reduce work and information overload. In: *Commun. ACM* 37 (1994), Nr. 7, S. 30-40. – DOI 10.1145/176789.176792
- [Mandl 2006] MANDL, Thomas: *Die automatische Bewertung der Qualität von Internet-Seiten in Information Retrieval*, Universität Hildesheim, Habilitationsschrift, 2006
- [Manning u. a. 2008] MANNING, Christopher D. ; RAGHAVAN, Prabhakar ; SCHÜTZE, Hinrich: *Introduction to Information Retrieval*. Cambridge, UK : Cambridge University Press, 2008 <http://nlp.stanford.edu/IR-book/information-retrieval-book.html>. – ISBN 978-0-521-86571-5
- [Manning u. Schütze 1999] MANNING, Christopher D. ; SCHÜTZE, Hinrich: *Foundations of Statistical Natural Language Processing*. Cambridge, MA, USA : MIT Press, 1999. – ISBN 0262133601
- [Marchionini 2006] MARCHIONINI, Gary: Exploratory search: from finding to understanding. In: *Communications of the ACM* 49 (2006), April, Nr. 4, 41-46. – DOI 10.1145/1121949.1121979
- [Mayr 2009a] MAYR, Philipp: *Non-textual ranking in digital libraries*. <http://www.ib.hu-berlin.de/~mayr/talks/jourfixe-18-11-09.pdf>
- [Mayr 2009b] MAYR, Philipp: *Re-Ranking auf Basis von Bradfordizing für die verteilte Suche in Digitalen Bibliotheken*, Philosophische Fakultät I, Institut für Bibliotheks- und Informationswissenschaft. Humboldt-Universität zu Berlin, Dissertation, 2009. <http://edoc.hu-berlin.de/dissertationen/mayr-philipp-2009-02-18/PDF/mayr.pdf>
- [Mayr 2010] MAYR, Philipp: *Information Retrieval- Mehrwertdienste für Digitale Bibliotheken Crosskonkordanzen und Bradfordizing*. Bonn : GESIS, 2010 (GESIS-Schriftenreihe 5). <http://www.gesis.org/forschung-lehre/gesis-publikationen/gesis-reihen/gesis-schriftenreihe/#c11846>. – ISBN 978-3-86819-012-0
- [Mayr u. a. 2011a] MAYR, Philipp ; MUTSCHKE, Peter ; PETRAS, Vivien ; SCHAER, Philipp ; SURE, York: Applying Science Models for Search. In: *Information und*



- Wissen: global, sozial und frei? - Proceedings des 12. Internationalen Symposiums für Informationswissenschaft (ISI 2011)* Bd. 58. Boizenburg : Verlag Werner Hülsbusch, 2011 (Schriften zur Informationswissenschaft), 184—196
- [Mayr u. a. 2011b] MAYR, Philipp ; MUTSCHKE, Peter ; SCHAER, Philipp ; SURE, York: Mehrwertdienste für das Information Retrieval: das Projekt IRM. In: *Wissen - Wissenschaft - Organisation* Bd. 12. Würzburg : Ergon-Verlag, 2011 (Fortschritte in der Wissensorganisation)
- [Mayr u. Petras 2008] MAYR, Philipp ; PETRAS, Vivien: Cross-concordances: terminology mapping and its effectiveness for information retrieval. In: *74th IFLA World Library and Information Congress, Québec, Canada, 2008*
- [Mayr u. a. 2011c] MAYR, Philipp ; SCHAER, Philipp ; MUTSCHKE, Peter: A Science Model Driven Retrieval Prototype. In: BOTERAM, Felix (Hrsg.) ; GÖDERT, Winfried (Hrsg.) ; HUBRICH, Jessica (Hrsg.): *Proceedings of the Cologne Conference on Interoperability and Semantics in Knowledge Organization* Bd. 1. Würzburg : Ergon Verlag, 2011 (Reihe Informations- und Bibliothekswissenschaften), 111–122
- [McCallum u. Godwin 1981] MCCALLUM, Sally H. ; GODWIN, James L.: Statistics on Headings in the MARC File. In: *Journal of Library Automation* 14 (1981), Nr. 3, S. 194–201
- [McCandless u. Hatcher 2010] MCCANDLESS, Michael ; HATCHER, Erik: *Lucene in action*. 2nd ed. Greenwich : Manning, 2010. – ISBN 9781933988177
- [Metzler 2011] METZLER, Donald: *The Information Retrieval Series*. Bd. 27: *A Feature-Centric View of Information Retrieval*. Berlin, Heidelberg : Springer, 2011. – ISBN 9783642228971
- [Meurer 2003] MEURER, Anke: *Barrierefreiheit im WWW - Ein Mehrwert für alle?*, Universität Koblenz-Landau, Diplomarbeit, September 2003
- [Milojević 2010a] MILOJEVIĆ, Staša: Modes of collaboration in modern science: Beyond power laws and preferential attachment. In: *Journal of the American Society for Information Science and Technology* 61 (2010), Juli, 1410–1423. – DOI 10.1002/asi.v61:7
- [Milojević 2010b] MILOJEVIĆ, Staša: Power law distributions in information science: Making the case for logarithmic binning. In: *Journal of the American Society for Information Science and Technology* 61 (2010), Dezember, 2417–2425. – DOI 10.1002/asi.v61:12
- [Moreira u. a. 2011] MOREIRA, Catarina ; CALADO, Pável ; MARTINS, Bruno: Learning to Rank for Expert Search in Digital Libraries of Academic Publications. In: ANTUNES, Luis (Hrsg.) ; PINTO, H. (Hrsg.): *Progress in Artificial Intelligence* Bd. 7026. Berlin, Heidelberg : Springer, 2011 (Lecture Notes in Computer Science). – ISBN 978-3-642-24768-2, S. 431–445

- [Morris u. Yen 2004] MORRIS, Steven. A. ; YEN, Gary G.: Crossmaps: Visualization of overlapping relationships in collections of journal papers. In: *Proceedings of the National Academy of Sciences* 101 (2004), Januar, Nr. Suppl 1, 5291–5296. – DOI 10.1073/pnas.0307604100
- [Murphy 1973] MURPHY, Larry J.: Lotka's law in the humanities? In: *Journal of the American Society for Information Science* 24 (1973), November, Nr. 6, 461–462. – DOI 10.1002/asi.4630240607
- [Mutschke 2009] MUTSCHKE, Peter: Zentralitäts- und Prestigemaße. In: STEGBAUER, Christian (Hrsg.) ; HÄUSSLING, Roger (Hrsg.): *Handbuch Netzwerkforschung*. 1. Aufl. Wiesbaden : VS Verlag für Sozialwissenschaften, 2009. – ISBN 9783531158082, S. 365–378
- [Mutschke u. a. 2011] MUTSCHKE, Peter ; MAYR, Philipp ; SCHAER, Philipp ; SURE, York: Science models as value-added services for scholarly information systems. In: *Scientometrics* 89 (2011), Juni, Nr. 1, S. 349–364. – DOI 10.1007/s11192-011-0430-x
- [Nacke 1979] NACKE, Otto: Informetrie: Ein neuer Name für eine neue Disziplin. In: *Nachrichten für Dokumentation* 30 (1979), Nr. 6, S. 219–226
- [Naranan 1971] NARANAN, S.: Power Law Relations in Science Bibliography - A Self-Consistent Interpretation. In: *Journal of Documentation* 27 (1971), Nr. 2, 83–97. – DOI 10.1108/eb026510
- [Neal 2012] NEAL, Diane R. (Hrsg.): *Indexing and Retrieval of Non-Text Information*. De Gruyter - Saur, 2012 (Knowledge and Information / Studies in Information Science). – ISBN 978-3-11-026057-1
- [Newman 2005] NEWMAN, M. E. J.: Power laws, Pareto distributions and Zipf's law. In: *Contemporary Physics* 46 (2005), 323 – 351. – DOI 10.1080/00107510500052444
- [van Noorden 2011] NOORDEN, Richard van: *Most-published researchers in 2011*. <http://twitter.com/#!/Richvn/status/144414256610283520>
- [Norman 2008] NORMAN, A.: Simplicity is not the answer. In: *ACM interactions* 15 (2008), Nr. 5, 45–46. – DOI 10.1145/1390085.1390094
- [Nwana 1996] NWANA, Hyacinth S.: Software Agents: An Overview. In: *Knowledge Engineering Review* 11 (1996), September, Nr. 3, 1–40. <http://agents.umbc.edu/introduction/ao/>
- [Odell 2000] ODELL, James: Agent Technology Green Paper, Version 1.0 / Agent Working Group OMG. [http://www.objs.com/agent/agents\\_Green\\_Paper\\_v100.doc](http://www.objs.com/agent/agents_Green_Paper_v100.doc). 2000 (Document ec/2000-08-01). – Forschungsbericht
- [Page u. a. 1999] PAGE, Lawrence ; BRIN, Sergey ; MOTWANI, Rajeev ; WINOGRAD, Terry: The PageRank Citation Ranking: Bringing Order to the Web. / Stanford InfoLab. <http://ilpubs.stanford.edu:8090/422/>. 1999 (1999-66). – Technical Report

- [Peters 2010] PETERS, Isabella: *Folksonomies. Indexing and Retrieval in Web 2.0*. Berlin : De Gruyter - Saur, 2010. – ISBN 978-3-598-25179-5
- [Peters u. Stock 2010] PETERS, Isabella ; STOCK, Wolfgang G.: Power tags in information retrieval. In: *Library Hi Tech* 28 (2010), Nr. 1, S. 81–93
- [Petras 2005] PETRAS, Vivien: How one word can make all the difference - using subject metadata for automatic query expansion and reformulation. In: *Working Notes for the CLEF 2005 Workshop, 21–23 September, Vienna, Austria, 2005*
- [Petras 2006] PETRAS, Vivien: *Translating Dialects in Search: Mapping between Specialized Languages of Discourse and Documentary Languages*. Berkeley, University of California, PhD Thesis, 2006. <http://www.sims.berkeley.edu/~vivienp/diss/>
- [Plaunt u. Norgard 1998] PLAUNT, Christian ; NORGDARD, Barbara A.: An association-based method for automatic indexing with a controlled vocabulary. In: *Journal of the American Society for Information Science* 49 (1998), August, Nr. 10, 888–902. – DOI 10.1002/(SICI)1097-4571(199808)49:10<888::AID-ASI5>3.0.CO;2-Z
- [Potter 1981] POTTER, William G.: Lotka's Law Revisited. In: *Library Trends* 30 (1981), Nr. 1, S. 21–39
- [Priem u. a. 2011] PRIEM, Jason ; TARABORELLI, Dario ; GROTH, Paul ; NEYLON, Cameron: *altmetrics: a manifesto*. <http://altmetrics.org/manifesto/>
- [Radhakrishnan u. Kernizan 1979] RADHAKRISHNAN, T. ; KERNIZAN, R.: Lotka's Law and Computer Science Literature. In: *Journal of the American Society for Information Science* 30 (1979), Nr. 1, 51–54. – DOI 10.1002/asi.4630300109
- [Raghavan 2008] RAGHAVAN, Prabhakar: *The changing face of web search*. <http://www.di.unipi.it/~cpm08/prabh.html>
- [Redner 2010] REDNER, Sidney: On the meaning of the h-index. In: *Journal of Statistical Mechanics: Theory and Experiment* 2010 (2010), März, L03005. – DOI 10.1088/1742-5468/2010/03/L03005
- [Rhodes u. Maes 2000] RHODES, Bradley J. ; MAES, Pattie: Just-in-time information retrieval agents. In: *IBM Systems Journal* 39 (2000), Nr. 3&4, S. 685–704. – DOI 10.1147/sj.393.0685
- [van Rijsbergen 1974] RIJSBERGEN, C. J.: Foundation of Evaluation. In: *Journal of Documentation* 30 (1974), Nr. 4, 365–373. – DOI 10.1108/eb026584
- [Robertson 1977a] ROBERTSON, Stephen E.: The Probability Ranking Principle in IR. In: *Journal of Documentation* 33 (1977), Nr. 4, S. 294–304
- [Robertson 1977b] ROBERTSON, Stephen E.: Theories and Models in Information Retrieval. In: *Journal of Documentation* 33 (1977), Nr. 2, S. 126–148

- [Robertson u. a. 1995] ROBERTSON, Stephen E. ; WALKER, Steve ; JONES, Susan ; HANCOCK-BEAULIEU, Micheline ; GATFORD, Mike: Okapi at TREC-3. In: *Proceedings of 3rd Text Retrieval Conference*, 1995, 109–126
- [Robertson u. Zaragoza 2010] ROBERTSON, Stephen E. ; ZARAGOZA, Hugo: The Probabilistic Relevance Framework: BM25 and Beyond. In: *Foundations and Trends in Information Retrieval* 3 (2010), Nr. 4, 333–389. – DOI 10.1561/15000000019
- [Rousseau u. Rousseau 2000] ROUSSEAU, B. ; ROUSSEAU, R.: LOTKA: A program to fit a power law distribution to observed frequency data. In: *Cybermetrics* 4 (2000), Nr. 1, Paper No. 4. <http://cybermetrics.cindoc.csic.es/articles/v4i1p4.html>
- [Salton u. McGill 1983] SALTON, G. ; MCGILL, M.: *Introduction to Modern Information Retrieval*. McGraw-Hill, 1983
- [Sanderson 2010] SANDERSON, Mark: Test Collection Based Evaluation of Information Retrieval Systems. In: *Foundations and Trends in Information Retrieval* 4 (2010), Nr. 4, 247–375. – DOI 10.1561/15000000009
- [Sanderson u. Soboroff 2007] SANDERSON, Mark ; SOBOROFF, Ian: Problems with Kendall's tau. In: *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA : ACM, 2007 (SIGIR '07). – ISBN 978–1–59593–597–7, 839–840
- [Saracevic 2007a] SARACEVIC, Tefko: Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance. In: *Journal of the American Society for Information Science and Technology* 58 (2007), November, Nr. 13, 1915–1933. – DOI 10.1002/asi.v58:13
- [Saracevic 2007b] SARACEVIC, Tefko: Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. In: *Journal of the American Society for Information Science and Technology* 58 (2007), November, Nr. 13, 2126–2144. – DOI 10.1002/asi.v58:13
- [Saracevic 2009] SARACEVIC, Tefko: Introduction: the framework for digital library evaluation. In: TSAKONAS, Giannis (Hrsg.) ; PAPTAEODOROU, Christos (Hrsg.): *Evaluation of Digital Libraries: An Insight to Useful Applications and Methods*. Chandos Publishing (Oxford) Ltd, Juli 2009. – ISBN 184334484X, S. 1–13
- [Schaefer 2008] SCHAEFER, Andre: *Arbeitsteilung zwischen direkter Manipulation und proaktiven Software-Agenten in Informationssystemen*. Duisburg, Universität Duisburg-Essen, Dissertation, 2008. <http://duepublico.uni-duisburg-essen.de/servlets/DocumentServlet?id=19419>
- [Schaer 2010] SCHAER, Philipp: Integration von Open Access Repositorien in Fachportale. In: SIEGLERSCHMIDT, Jörn (Hrsg.) ; OHLY, Peter (Hrsg.): *Wissensspeicherung in digitalen Räumen. Nachhaltigkeit, Verfügbarkeit, semantische Interoperabilität*. Würzburg : Ergon-Verlag, 2010 (Fortschritte in der Wissensorganisation), 245–251

- [Schaer 2011] SCHAER, Philipp: Using Lotkaian Informetrics for Ranking in Digital Libraries. In: HOARE, Cathal (Hrsg.) ; O’RIORDAN, Adrian (Hrsg.): *Proceedings of the ASIS&T European Workshop 2011 (AEW 2011)*. Cork, Ireland : ASIS&T, 2011
- [Schaer 2012] SCHAER, Philipp: Better than Their Reputation? On the Reliability of Relevance Assessments with Students. In: CATARCI, Tiziana (Hrsg.) ; FORNER, Pamela (Hrsg.) ; HIEMSTRA, Djoerd (Hrsg.) ; PEÑAS, Anselmo (Hrsg.) ; SANTUCCI, Giuseppe (Hrsg.): *Information Access Evaluation. Multilinguality, Multimodality, and Visual Analytics* Bd. 7488. Berlin, Heidelberg : Springer, 2012 (Lecture Notes in Computer Science). – ISBN 978-3-642-33246-3, S. 124-135
- [Schaer u. Heck 2013] SCHAER, Philipp ; HECK, Tamara: Performing Informetric Analysis on Information Retrieval Test Collections: Preliminary Experiments in the Physics Domain. In: *Proceedings of the 14th International Society of Scientometrics and Informetrics Conference*, 2013. – im Druck
- [Schaer u. Heuser 2006] SCHAER, Philipp ; HEUSER, Holger: Grundlagen der Kognition und Perzeption für die Software-Ergonomie / Universität Koblenz-Landau, Institut für Computervisualistik. <http://userpages.uni-koblenz.de/~schaer/publications/KognitionPerzeptionSoftwareErgonomie.pdf>. 2006. – Research Report
- [Schaer u. a. 2012a] SCHAER, Philipp ; HIENERT, Daniel ; SAWITZKI, Frank ; WIRA-ALAM, Andias ; LÜKE, Thomas: Dealing with Sparse Document and Topic Representations: Lab Report for CHiC 2012. In: FORNER, Pamela (Hrsg.) ; KARLGREN, Jussi (Hrsg.) ; WOMSER-HACKER, Christa (Hrsg.): *CLEF (Online Working Notes/Labs/Workshop)*, 2012. – ISBN 978-88-904810-3-1
- [Schaer u. a. 2012b] SCHAER, Philipp ; LÜKE, Thomas ; HOEK, Wilko van: Building Custom Term Suggestion Web Services with OAI-Harvested Open Data. In: OCKENFELD, Marlies (Hrsg.) ; PETERS, Isabella (Hrsg.) ; WELLER, Katrin (Hrsg.): *Proceedings of the 64. DGI Annual Meeting and 2nd DGI-Conference*, Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis, 2012 (Tagungen der Deutschen Gesellschaft für Informationswissenschaft und Informationspraxis), S. 389-396
- [Schaer u. a. 2012c] SCHAER, Philipp ; MAYR, Philipp ; LÜKE, Thomas: Extending Term Suggestion with Author Names. In: ZAPHIRIS, Panayiotis (Hrsg.) ; BUCHANAN, George (Hrsg.) ; RASMUSSEN, Edie (Hrsg.) ; LOIZIDES, Fernando (Hrsg.): *Theory and Practice of Digital Libraries* Bd. 7489. Berlin, Heidelberg : Springer, 2012 (Lecture Notes in Computer Science). – ISBN 978-3-642-33289-0, S. 317-322
- [Schaer u. a. 2010a] SCHAER, Philipp ; MAYR, Philipp ; MUTSCHKE, Peter: Demonstrating a Service-Enhanced Retrieval System. In: *Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem* Bd. 47. Silver Springs, MD, USA : American Society for Information Science, 2010, 146:1-146:2
- [Schaer u. a. 2010b] SCHAER, Philipp ; MAYR, Philipp ; MUTSCHKE, Peter: Implications of Inter-Rater Agreement on a Student Information Retrieval Evaluation. In: ATZMÜLLER,

- Martin (Hrsg.) ; BENZ, Dominik (Hrsg.) ; HOTHO, Andreas (Hrsg.) ; STUMME, Gerd (Hrsg.): *Proceedings of LWA2010 - Workshop-Woche: Lernen, Wissen & Adaptivitaet*. Kassel, Germany : Fachgruppe Information Retrieval der Gesellschaft für Informatik, 2010
- [Schaer u. a. 2011] SCHAER, Philipp ; MAYR, Philipp ; MUTSCHKE, Peter: Retrieval-Mehrwertdienste zur Weiterentwicklung wissenschaftlicher Fachportale wie vascoda und sowiport. Suchexpandierung und Re-Ranking. 2011. – DFG-Abschlussbericht
- [Schamber u. a. 1990] SCHAMBER, Linda ; EISENBERG, Michael B. ; NILAN, Michael S.: A re-examination of relevance: toward a dynamic, situational definition. In: *Information Processing & Management* 26 (1990), Nr. 6, 755–776. – DOI 10.1016/0306-4573(90)90050-C
- [Scharnhorst 2012] SCHARNHORST, Andrea: Preface. In: SCHARNHORST, Andrea (Hrsg.) ; BÖRNER, Katy (Hrsg.) ; BESSELAAR, Peter van d. (Hrsg.): *Models of Science Dynamics Encounters Between Complexity Theory and Information Sciences*. Berlin : Springer, 2012 (Understanding Complex Systems), S. xi–xix. – DOI 10.1007/978-3-642-23068-4
- [Scharnhorst u. Garfield 2010] SCHARNHORST, Andrea ; GARFIELD, Eugene: Tracing scientific influence. In: *Journal of Dynamics of Socio-Economic Systems* 2 (2010), Nr. 1
- [Schatz u. a. 1996] SCHATZ, Bruce R. ; JOHNSON, Eric H. ; COCHRANE, Pauline A. ; CHEN, Hsinchun: Interactive term suggestion for users of digital libraries: using subject thesauri and co-occurrence lists for information retrieval. In: *Proceedings of the first ACM international conference on Digital libraries*. New York, NY, USA : ACM, 1996 (DL '96). – ISBN 0-89791-830-4, 126–133
- [Schlögl u. Gorraiz 2012] SCHLÖGL, Christian ; GORRAIZ, Juan: Sind Downloads die besseren Zeitschriftennutzungsdaten? Ein Vergleich von Download und Zitationsindikatoren. In: *Zeitschrift für Bibliothekswesen und Bibliographie* 59 (2012), Nr. 2, S. 87–95
- [Schmidt u. a. 2011] SCHMIDT, Thomas C. ; WÄHLISCH, Matthias ; BRÜN, Markus de ; HÄBERLEN, Thomas: Ein Routing-Atlas für die strukturelle und visuelle Exposition des deutschen Internets. In: *PIK - Praxis der Informationsverarbeitung und Kommunikation* 34 (2011), Juni, Nr. 2, 60–72. – DOI 10.1515/piko.2011.012
- [Schorr 1975] SCHORR, Alan E.: Lotka's Law and Map Librarianship. In: *Journal of the American Society for Information Science* 26 (1975), Mai, Nr. 3, 189–190. – DOI 10.1002/asi.4630260308
- [Schütz 2011] SCHÜTZ, Christian: *Nationales ISSN-Zentrum für Deutschland - Häufig gestellte Fragen*. <http://www.dnb.de/DE/Wir/Kooperation/ISSN/issnFAQ.html>

- [Seadle 2009] SEADLE, Michael: Digitale Bibliothek. In: UMLAUF, Konrad (Hrsg.) ; GRADMANN, Stefan (Hrsg.): *Lexikon der Bibliotheks- und Informationswissenschaft (LBI)* Bd. 3. Stuttgart : Anton Hiersemann Verlag, 2009. – ISBN 978-3-7772-0922-7, S. 216–217
- [Sean u. a. 2011] SEAN, Owen ; ROBIN, Anil ; TED, Dunning ; ELLEN, Friedman: *Mahout in Action*. Manning, 2011. – ISBN 9781935182689
- [Shannon u. Weaver 1963] SHANNON, Claude ; WEAVER, Warren: *The mathematical theory of communication*. Urbana : University of Illinois Press, 1963. – ISBN 9780252725487
- [Shiri u. Revie 2006] SHIRI, Ali ; REVIE, Crawford: Query expansion behavior within a thesaurus-enhanced search environment: A user-centered evaluation. In: *Journal of the American Society for Information Science and Technology* 57 (2006), Februar, 462–478. – DOI 10.1002/asi.v57:4
- [Shneiderman u. Maes 1997] SHNEIDERMAN, Ben ; MAES, Pattie: Direct manipulation vs. interface agents. In: *interactions* 4 (1997), Nr. 6, S. 42–61. – DOI 10.1145/267505.267514
- [Siegfried u. Flieger 2011] SIEGFRIED, Doreen ; FLIEGER, Elisabeth: World Wide Wissenschaft – Wie professionell Forschende im Internet arbeiten / ZBW - Leibniz-Informationszentrum Wirtschaft. [http://www.zbw.eu/presse/pressemitteilungen/docs/world\\_wide\\_wissenschaft\\_zbw\\_studie.pdf](http://www.zbw.eu/presse/pressemitteilungen/docs/world_wide_wissenschaft_zbw_studie.pdf). 2011. – Forschungsbericht. – 15 S.
- [Skov u. a. 2008] SKOV, Mette ; LARSEN, Birger ; INGWERSEN, Peter: Inter and intra-document contexts applied in polyrepresentation for best match IR. In: *Information Processing & Management* 44 (2008), September, Nr. 5, 1673–1683. – DOI 10.1016/j.ipm.2008.05.006
- [Smiley u. Pugh 2009] SMILEY, David ; PUGH, Eric: *Solr 1.4 Enterprise Search Server enhance your search with faceted navigation, result highlighting, fuzzy queries, ranked scoring, and more*. Birmingham UK : Packt Publishing, 2009. – ISBN 9781847195890
- [de Solla Price 1963] SOLLA PRICE, Derek J.: *Little Science, Big Science*. New York : Columbia University Press, 1963
- [Song u. a. 2011] SONG, Ruihua ; GUO, Qingwei ; ZHANG, Ruochi ; XIN, Guomao ; WEN, Ji-Rong ; YU, Yong ; HON, Hsiao-Wuen: Select-the-Best-Ones: A new way to judge relative relevance. In: *Information Processing & Management* 47 (2011), Januar, Nr. 1, 37–52. – DOI 10.1016/j.ipm.2010.02.005
- [Spärck Jones 1972] SPÄRCK JONES, Karen: A statistical interpretation of term specificity and its application in retrieval. In: *Journal of Documentation* 28 (1972), Nr. 1, S. 11–21
- [Spärck Jones 1999] SPÄRCK JONES, Karen: What is the Role of NLP in Text Retrieval? In: STRZALKOWSKI, Tomek (Hrsg.): *Natural language information retrieval* Bd. 7. Dordrecht, Boston : Kluwer Academic, 1999. – ISBN 9780792356851, S. 1–24

- [Stempfhuber u. a. 2008] STEMPFHUBER, Maximilian ; SCHAER, Philipp ; SHEN, Wei: Enhancing visibility: integrating Grey Literature in the SOWIPORT Information Cycle. In: *The Grey Journal* (2008), Nr. 3, S. 121–128
- [Stock u. Weber 2006] STOCK, Wolfgang ; WEBER, Sonja: Facets of Informetrics. In: *Information - Wissenschaft & Praxis* 57 (2006), Nr. 6, S. 385–389
- [Sühl-Strohmenger 2008] SÜHL-STROHMENGER, Wilfried: *Digitale Welt und Wissenschaftliche Bibliothek-Informationsspraxis im Wandel : Determinanten, Ressourcen, Dienste, Kompetenzen : eine Einführung*. Wiesbaden : Harrassowitz, 2008. – ISBN 978-3-447-05741-7
- [Sullivan 2010] SULLIVAN, Danny: *What Social Signals Do Google and Bing Really Count?* <http://searchengineland.com/what-social-signals-do-google-bing-really-count-55389>
- [Tijssen u. Van Raan 1994] TIJSSSEN, Robert J. W. ; VAN RAAN, Anthony F. J.: Mapping Changes in Science and Technology: Bibliometric Co-Occurrence Analysis of the R&D Literature. In: *Evaluation Review* 18 (1994), Nr. 1, S. 98–115
- [Tran u. Fuhr 2012] TRAN, Vu T. ; FUHR, Norbert: Using Eye-Tracking with Dynamic Areas of Interest for Analyzing Interactive Information Retrieval. In: *Proceedings of the 35th international ACM SIGIR conference on Research and development in Information Retrieval*. New York, NY, USA : ACM, 2012. – ISBN 978-1-4503-1472-5, S. 1165–1166
- [Trotman 2005] TROTMAN, Andrew: Learning to Rank. In: *Information Retrieval* 8 (2005), Mai, Nr. 3, 359–381. – DOI 10.1007/s10791-005-6991-7
- [Trotman u. Keeler 2011] TROTMAN, Andrew ; KEELER, David: Ad Hoc IR - Not Much Room for Improvement. In: *Proceeding of the 34th international ACM SIGIR conference on Research and development in information retrieval - SIGIR '11*. New York, NY, USA : ACM, 2011. – ISBN 978-1-4503-0757-4, S. 1095–1096
- [Tunkelang 2009] TUNKELANG, Daniel: *Faceted Search*. Morgan & Claypool Publishers, 2009 (Synthesis Lectures on Information Concepts, Retrieval, and Services). – ISBN 978-1598299991
- [Umlauf u. Gradmann 2009] UMLAUF, Konrad (Hrsg.) ; GRADMANN, Stefan (Hrsg.): *Lexikon der Bibliotheks- und Informationswissenschaft (LBI)*. Bd. 1 ff. Stuttgart : Anton Hiersemann Verlag, 2009. – ISBN 978-3-7772-0922-7
- [Umstätter 1999] UMSTÄTTER, Walther: *Zum Thema Lotka's law*. <http://www.ib.hu-berlin.de/~wumsta/price52.html>
- [Umstätter 2003] UMSTÄTTER, Walther: *Das Principle of Least Effort in der Wissenschaft*. <http://www.ib.hu-berlin.de/~wumsta/infopub/lectures/leasteffort03a.pdf>



- [Umstätter 2005a] UMSTÄTTER, Walther: Anmerkungen zu Birger Hjørland und Jeppe Nicolaisen: Bradford's Law of Scattering: Ambiguities in the Concept of Subject. In: *LIBREAS - Library Ideas* (2005), Nr. 3. [http://www.ib.hu-berlin.de/~libreas/libreas\\_neu/ausgabe3/008ums.htm](http://www.ib.hu-berlin.de/~libreas/libreas_neu/ausgabe3/008ums.htm)
- [Umstätter 2005b] UMSTÄTTER, Walther: *Die Bedeutung des Bradford's Law of Scattering für die Bibliothekswissenschaft*. <http://www.ib.hu-berlin.de/~wumsta/infopub/pub2001f/Bradford05fold.pdf>
- [Umstätter 2006] UMSTÄTTER, Walther: *Ist es notwendig, eine neue Zeitschrift für Informetrie zu gründen?* <http://www.ib.hu-berlin.de/~wumsta/infopub/pub2001f/BBKInformetrics.pdf>
- [Vechtomova u. Karamuftuoglu 2004] VECHTOMOVA, Olga ; KARAMUFTUOGLU, Murat: Comparison of two interactive search refinement techniques. In: *HLT-NAACL'04*, 2004, S. 225–232
- [Vechtomova u. a. 2003] VECHTOMOVA, Olga ; ROBERTSON, Stephen ; JONES, Susan: Query Expansion with Long-Span Collocates. In: *Inf. Retr.* 6 (2003), April, 251–273. – DOI 10.1023/A:1023936321956
- [Voorhees 1998] VOORHEES, Ellen M.: Variations in Relevance Judgments and the Measurement of Retrieval Effectiveness. In: *SIGIR '98: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, August 24-28 1998, Melbourne, Australia*. New York, NY, USA : ACM, 1998. – ISBN 1-58113-015-5, S. 315–323
- [Voorhees 2000] VOORHEES, Ellen M.: Variations in relevance judgments and the measurement of retrieval effectiveness. In: *Information Processing & Management* 36 (2000), September, Nr. 5, 697–716. – DOI 10.1016/S0306-4573(00)00010-8
- [Voorhees 2009] VOORHEES, Ellen M.: Topic set size redux. In: *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA : ACM, 2009 (SIGIR '09). – ISBN 978-1-60558-483-6, 806–807
- [Voorhees u. Harmann 2005] VOORHEES, Ellen M. (Hrsg.) ; HARMANN, Donna (Hrsg.): *TREC : experiment and evaluation in information retrieval*. Cambridge, MA, USA : MIT Press, 2005. – ISBN 9780262220736
- [Voß 2007] VOSS, Jakob: *Visualisierung der ZDB*. [http://www.gbv.de/wikis/cls/Visualisierung\\_der\\_ZDB](http://www.gbv.de/wikis/cls/Visualisierung_der_ZDB)
- [Voß 2012] VOSS, Jakob: Metadaten. In: UMLAUF, Konrad (Hrsg.) ; GRADMANN, Stefan (Hrsg.): *Lexikon der Bibliotheks- und Informationswissenschaft (LBI)* Bd. 7. Stuttgart : Anton Hiersemann Verlag, 2012. – ISBN 978-3-7772-0922-7

- [Webber u. a. 2008] WEBBER, William ; MOFFAT, Alistair ; ZOBEL, Justin ; SAKAI, Tetsuya: Precision-at-ten considered redundant. In: *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. Singapore, Singapore : ACM, 2008 (SIGIR '08). – ISBN 978–1–60558–164–4, S. 695–696
- [White 1981] WHITE, Howard D.: 'Bradfordizing' search output: how it would help online users. In: *Online Review* 5 (1981), Nr. 1, S. 47–54
- [White 2007a] WHITE, Howard D.: Combining bibliometrics, information retrieval, and relevance theory, Part 1: First examples of a synthesis: Research Articles. In: *Journal of the American Society for Information Science and Technology* 58 (2007), Februar, 536–559. – DOI 10.1002/asi.v58:4
- [White 2007b] WHITE, Howard D.: Combining bibliometrics, information retrieval, and relevance theory, Part 2: Some implications for information science: Research Articles. In: *Journal of the American Society for Information Science and Technology* 58 (2007), Februar, Nr. 4, 583–605. – DOI 10.1002/asi.v58:4
- [White 2006] WHITE, Ryan: Using searcher simulations to redesign a polyrepresentative implicit feedback interface. In: *Information Processing & Management* 42 (2006), September, Nr. 5, 1185–1202. – DOI 10.1016/j.ipm.2006.02.005
- [White u. Marchionini 2007] WHITE, Ryan W. ; MARCHIONINI, Gary: Examining the effectiveness of real-time query expansion. In: *Information Processing & Management* 43 (2007), Mai, 685–704. – DOI 10.1016/j.ipm.2006.06.005
- [Wilde u. a. 2010] WILDE, Anja ; WENNINGER, Agnieszka ; HOPT, Oliver ; SCHAER, Philipp ; ZAPILKO, Benjamin: Aktivitäten von GESIS im Kontext von Open Data und Zugang zu sozialwissenschaftlichen Forschungsergebnissen. In: *1. DGI-Konferenz, 62. DGI Jahrestagung - Semantic Web & Linked Data Elemente zukünftiger Informationsinfrastrukturen*. Frankfurt am Main : Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis, 2010, 183–192
- [Wong u. a. 2010] WONG, William ; STELMASZEWSKA, Hanna ; BARN, Balbir ; BHIMANI, Nazlin ; SUKHBINDER, Barn: JISC user behaviour observational study: User behaviour in resource discovery / JISC. <http://www.jisc.ac.uk/publications/programmerelated/2010/ubirdfinalreport.aspx>. 2010. – Forschungsbericht
- [Wooldridge u. Jennings 1995] WOOLDRIDGE, Michael ; JENNINGS, Nicholas R.: Intelligent Agents: Theory and Practice. In: *Knowledge Engineering Review* 10 (1995), Nr. 2, S. 115–152
- [Xie 2008] XIE, Hong I.: Users' evaluation of digital libraries (DLs): Their uses, their criteria, and their assessment. In: *Information Processing & Management* 44 (2008), Mai, Nr. 3, 1346–1373. – DOI 10.1016/j.ipm.2007.10.003

- [Yan u. Ding 2009] YAN, Erjia ; DING, Ying: Applying centrality measures to impact analysis: A coauthorship network analysis. In: *Journal of the American Society for Information Science and Technology* 60 (2009), Oktober, Nr. 10, 2107–2118. – DOI 10.1002/asi.v60:10
- [Yin u. a. 2006] YIN, Li-chun ; KRETSCHMER, Hildrun ; HANNEMAN, Robert A. ; LIU, Ze-yuan: Connection and stratification in research collaboration: an analysis of the COLLNET network. In: *Information Processing & Management* 42 (2006), Dezember, Nr. 6, 1599–1613. – DOI 10.1016/j.ipm.2006.03.021
- [Zanardi u. Capra 2008] ZANARDI, Valentina ; CAPRA, Licia: Social ranking: uncovering relevant content using tag-based recommender systems. In: *Proceedings of the 2008 ACM conference on Recommender systems*. Lausanne, Switzerland : ACM, 2008 (RecSys '08). – ISBN 978-1-60558-093-7, S. 51–58
- [Zapilko u. Sure 2009] ZAPILKO, Benjamin ; SURE, York: Converting the TheSoz to SKOS / GESIS. [http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/gesis\\_methodenberichte/2009/TechnicalReport\\_09\\_07.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2009/TechnicalReport_09_07.pdf). Bonn, 2009 (2009/07). – GESIS Technical Reports
- [Zens 2010] ZENS, Maria: *Informationsverhalten von WissenschaftlerInnen*. Bonn, November 2010
- [Zhai 2007] ZHAI, ChengXiang: Statistical Language Models for Information Retrieval A Critical Review. In: *Foundations and Trends in Information Retrieval* 2 (2007), Nr. 3, 137–213. – DOI 10.1561/15000000008
- [Zhang 2008] ZHANG, Yan: Undergraduate students mental models of the Web as an information retrieval system. In: *Journal of the American Society for Information Science and Technology* 59 (2008), November, Nr. 13, 2087–2098. – DOI 10.1002/asi.v59:13
- [Zhao 2011] ZHAO, Dangzhi: *Mapping research fields with author co-citation analysis and author bibliographic coupling analysis*. GESIS - Leibniz Institute for the Social Sciences, Bonn, Germany, Juli 2011
- [Zhu u. a. 2009] ZHU, Jianhan ; SONG, Dawei ; RÜGER, Stefan: Integrating multiple windows and document features for expert finding. In: *Journal of the American Society for Information Science and Technology* 60 (2009), April, Nr. 4, 694–715. – DOI 10.1002/asi.21012
- [Zimmer u. Rehse 2011] ZIMMER, Monika (Hrsg.) ; REHSE, Brigitte (Hrsg.): *Sozialwissenschaftliche Zeitschriftenliste*. Bonn : GESIS, 2011 <http://vt-www.gesis.org/pdf/Zeitschriften.pdf?lang=de>
- [Zipf 1949] ZIPF, George K.: *Human Behaviour and the Principle of Least Effort*. Cambridge, MA, USA : Addison-Wesley, 1949

*Literaturverzeichnis*

[Zobel u. a. 2009] ZOBEL, Justin ; MOFFAT, Alistair ; PARK, Laurence A. F.: Against recall: is it persistence, cardinality, density, coverage, or totality? In: *SIGIR Forum* 43 (2009), Juni, Nr. 1, S. 3–8. – DOI 10.1145/1670598.1670600

# Abkürzungsverzeichnis

Im folgenden Verzeichnis werden die in der Arbeit verwendeten Abkürzungen aufgelöst und ggf. durch eine Übersetzung oder einen Verweis auf eine relevante Stelle in der Arbeit ergänzt.

bpref	Binary Preference (s. Abschnitt 5.2)
CLEF	Conference and Labs of the Evaluation Forum (s. Abschnitt 5.1)
DCG	Discounted Cumulative Gain (s. Abschnitt 5.2)
DL	Digital Library, digitale Bibliotheken (s. Kap. 2)
GIRT	German Indexing and Retrieval Test database (s. Abschnitt 5.1)
IIR	Interactive Information Retrieval (s. Abschnitt 5.1)
IR	Information Retrieval
iSearch	integrated Search (s. Abschnitt 5.1)
QE	Query Expansion, Anfrageerweiterung (s. Kap. 3)
MAP	Mean Average Precision (s. Abschnitt 5.2)
P	Precision (s. Abschnitt 5.2)
P@x	Cut-off-Precision (s. Abschnitt 5.2)
R	Recall (s. Abschnitt 5.2)
RP	R-Precision (s. Abschnitt 5.2)
TF*IDF	Term Frequency - Invers Document Frequency (s. Abschnitt 4.3.2)
TREC	Text Retrieval Evaluation Conference (s. Abschnitt 5.1)



# Anhang





# A Ermittelte Entitäten der Kookkurenzanalyse

Die folgenden Tabellen enthalten die jeweils für die einzelnen Topics ermittelten Autorennamen, ISSN-Codes und Thesaurusterme für GIRT und die Autorennamen und Publikationsquellen für iSearch.

## A.1 Ermittelte Entitäten für GIRT

topic	author	issn	subject
76	Beyme, Klaus von; Pelinka, Anton; Gabriel, Oscar W.; Butterwegge, Christoph	0479-611X; 0006-4416; 0340-0425; 0032-3470	Bundesrepublik Deutschland; Politik; historische Entwicklung; Entwicklungsland
77	Gabriel, Oscar W.; Kißler, Leo; Schröter, Ursula; Beyme, Klaus von	0479-611X; 0340-1758; 0340-0425; 0004-1157	Bundesrepublik Deutschland; Frau; DDR; neue Bundesländer
78	Nauck, Bernhard; Seifert, Wolfgang; Sen, Faruk; Herwartz-Emden, Leonie	0479-611X; 0723-8525; 0044-3247; 0930-9381	Asien; Entwicklungsland; Nahost; Bundesrepublik Deutschland
79	Jung, Otmar; Beck, Ulrich; Evers, Tilman; Narr, Wolf-Dieter	0479-611X; 0340-1758; 1430-6387; 0006-4416	Volksentscheid; direkte Demokratie; Verfassung; politisches System
80	Audretsch, David B.; Schmid, Josef; Hirsch-Kreinsen, Hartmut; Häussermann, Hartmut	0479-611X; 0342-300X; 0006-4416; 0342-8176	Entwicklungsland; Asien; Bundesrepublik Deutschland; historische Entwicklung
81	Faßmann, Hendrik	0341-4515; 0931-8895	Abbrecher; Berufsbildung; Determinanten; Ausbildung
82	Seifert, Wolfgang; Schober, Karen; Tessaring, Manfred; Lewin, Karl	0479-611X; 0341-4515; 0723-8525; 0340-3254	Bundesrepublik Deutschland; Arbeitsmarkt; Berufsbildung; Ausländer
83	Wilke, Jürgen; Hickethier, Knut; Gleich, Uli; Groebel, Jo	0479-611X; 0025-8350; 0941-5491; 0171-3957	Massenmedien; Bundesrepublik Deutschland; Berichterstattung; historische Entwicklung
84	Gleich, Uli; Groebel, Jo; Wilke, Jürgen; Aufenanger, Stefan	0171-3957; 0723-399X; 0479-611X; 0170-1754	Bundesrepublik Deutschland; Jugendlicher; Schule; Massenmedien

## A Ermittelte Entitäten der Kookkurenzanalyse

topic	author	issn	subject
85	Walwei, Ulrich; Werner, Heinz; Schmid, Josef; Teichler, Ulrich	0479-611X; 0723-8525; 0170-1754; 0930-9381	Bundesrepublik Deutschland; Europa; internationaler Vergleich; Bildungswesen
86	Beyme, Klaus von; Müller, Klaus; Hitzler, Ronald; Sterbling, Anton	0479-611X; 0006-4416; 0340-0425; 0863-1808	Bundesrepublik Deutschland; postsozialistisches Land; Europa; historische Entwicklung
87	Hurrelmann, Klaus; Jeschek, Wolfgang; Stadler, Hans; Schmidt, Bernd	0479-611X; 0513-9066; 0723-8525; 0722-8333	Bundesrepublik Deutschland; Jugendlicher; Schule; soziale Integration
88	Gerhardt, Uta; Brinkhoff, Klaus-Peter; Bette, Karl-Heinrich; Lüschen, Günther	1012-6902; 0479-611X; 0170-1754; 0259-7446	historische Entwicklung; Bundesrepublik Deutschland; Drittes Reich; DDR
89	Pöttker, Horst; Weischenberg, Siegfried; Ruß-Mohl, Stephan; Altmeppen, Klaus-Dieter	0033-4006; 1438-499X; 0025-8350; 0267-3231	Berufsethos; Journalist; journalistischer Beruf; Ethik
90	Peiser, Wolfram; Schönbach, Klaus	0033-4006	Rezeption; Presse; Rezipientenforschung; Leser
91	Büchel, Felix; Weißhuhn, Gernot; Schimpl-Neimanns, Bernhard; Allmendinger, Jutta	0479-611X; 0723-8525; 0044-3247; 0172-2875	Bildungswesen; Bundesrepublik Deutschland; Berufsbildung; Bildungspolitik
92	Niedermayer, Oskar; Bergstedt, Jörg; Geißel, Brigitte; Hoecker, Beate	0340-1758; 0479-611X; 0863-4890	Mitgliedschaft; Partei; innerparteiliche Demokratie; SPD
93	-	-	Burnout; Lehrer; Hauptschule; Krankenpflege
94	Starke, Kurt; Hutter, Jörg; Weller, Konrad; Lautmann, Rüdiger	0932-8114; 0722-0189; 0323-3227	Sexualität; Sexualforschung; AIDS; Diskriminierung
95	Seifert, Hartmut	0342-300X	Wochenende; Schichtarbeit; Arbeitszeit; Nachtarbeit
96	Bardeleben, Richard von; Beicht, Ursula; Lewin, Karl; Schober, Karen	0341-4515; 0340-3254; 0479-611X; 0723-8525	Berufsbildung; Bundesrepublik Deutschland; Betrieb; berufliche Weiterbildung
97	Eiben, Jürgen; Buss, Andreas; Hemminger, Hansjörg; Rohr, Elisabeth	0479-611X	Sekte; Religion; religiöse Bewegung; Christentum
98	Czempiel, Ernst-Otto; Timmermann, Heinz; Hacke, Christian; Alexandrova, Olga	0479-611X; 0006-4416; 0340-1758; 0030-6428	Nordamerika; internationale Beziehungen; USA; Kanada

## A.1 Ermittelte Entitäten für GIRT

topic	author	issn	subject
99	Krafeld, Franz Josef; Hafeneger, Benno; Brenner, Gerd; Pilz, Gunter A.	1012-6902; 0342-9857; 0012-0332; 0723-399X	Jugendlicher; Jugend; Freizeit; Bundesrepublik Deutschland
100	-	-	-
101	Bach, Heinz Willi; Schaeffer, Doris; Krebs, Angelika; Irrgang, Bernhard	0479-611X; 0721-7234; 0932-8114; 0038-609X	Medizin; Ethik; Tier; Bundesrepublik Deutschland
102	Heise, Arne; Keller, Berndt; Schmid, Günther; Kress, Ulrike	0342-300X; 0479-611X; 0267-3231; 0170-1754	Liberalisierung; Wettbewerb; Flexibilität; Bundesrepublik Deutschland
103	-	-	Wohnverhalten; Wohnungspolitik; Wohnen; Architektur
104	Rosenbrock, Rolf; Hurrelmann, Klaus; Jacob, Rüdiger; Schaeffer, Doris	0170-2602; 0341-2059; 0941-3790; 0943-1853	Gesundheit; Krankheit; Gesundheitsvorsorge; Prophylaxe
105	Parmentier, Klaus; Schreyer, Franziska; Gleiser, Sigmar; Minks, Karl-Heinz	0340-3254; 0479-611X; 0723-8525; 0724-3464	Bundesrepublik Deutschland; Arbeitslosigkeit; Hochschule; Berufsaussicht
106	Merz, Joachim; Schröer, Norbert; Wagner, Gert; Busse, Horst	0044-3514; 0023-2653; 0941-1670; 0723-5607	Multiple Sklerose; Krankheit; Bundesrepublik Deutschland; Behinderter
107	Siedschlag, Alexander	0479-611X; 0177-7521	Konfliktbewältigung; Konfliktregelung; UNO; Konflikt
108	-	-	-
109	Weizsäcker, Ernst Ulrich von; Winkler, Gerhard; Kretschmar, Gotthard; Brown, Lester R.	0722-8333; 0479-611X; 0176-2389; 0005-9080	Umweltpolitik; Ökologie; Umweltschutz; Bundesrepublik Deutschland
110	Hurrelmann, Klaus; Stiehler, Hans-Jörg; Förster, Peter; Wiedemann, Dieter	0723-399X; 0171-3957; 0937-9614; 0012-0332	Jugendlicher; Bundesrepublik Deutschland; DDR; Kind
111	Hurrelmann, Klaus; Schubarth, Wilfried; Jeschek, Wolfgang; Tillmann, Klaus-Jürgen	0513-9066; 0028-3355; 0171-3957; 0479-611X	Kind; Bundesrepublik Deutschland; Jugendlicher; Eltern
112	Gregory, Abigail	0025-8350; 0173-5993; 0941-5491; 0170-1754	Einzelhandel; Kaufverhalten; Bundesrepublik Deutschland; Marketing
113	Bingen, Dieter; Fehr, Helmut; Tatur, Melanie; Krasnodebski, Zdzislaw	0010-3497; 0479-611X; 0012-0812; 0006-4416	postsozialistisches Land; katholische Kirche; historische Entwicklung; Bundesrepublik Deutschland

## A Ermittelte Entitäten der Kookkurenzanalyse

topic	author	issn	subject
114	Werner, Heinz; Walwei, Ulrich; Wagner, Gert; Kühl, Jürgen	0479-611X; 0722-8333; 0342-300X; 0340-3254	Bundesrepublik Deutschland; neue Bundesländer; historische Entwicklung; Arbeitsmarkt
115	Schmähl, Winfried; Bäcker, Gerhard; Wiesenthal, Helmut; Kohlhaas, Michael	0479-611X; 0342-300X; 0006-4416; 0170-1754	Bundesrepublik Deutschland; Umweltpolitik; Reform; Auswirkung
116	Pekrun, Reinhard	-	Prüfung; Angst; Schüler; Schulleistung
117	Meier, Christian; Dresch, Alfred; Naumann, Astrid; Wiemann, Jürgen	0043-6143; 0479-611X; 0945-2419; 0018-974X	Umweltpolitik; Umweltschutz; Wirtschaftspolitik; Außenhandel
118	Pollack, Detlef; Luhmann, Niklas; Pickel, Gert; Tibi, Bassam	0479-611X; 0340-0425; 0379-3664; 0010-3497	historische Entwicklung; Bundesrepublik Deutschland; Entwicklungsland; Gesellschaft
119	Schuler, Heinz; Hartl, Michaela; Walter, Volker; Pokorny, Matthias	0723-3868; 0932-4089; 0179-6437; 0031-5605	Personaleinstellung; Assessment-Center; Führungskraft; Bundesrepublik Deutschland
120	Hunger, Bernd; Krautzberger, Michael; Meyer, Birgit; Faulenbach, Bernd	0170-9364; 0479-611X; 0003-9209; 0251-3625	Bundesrepublik Deutschland; historische Entwicklung; Stadtentwicklung; Stadtplanung
121	Heller, Stephan; Mutz, Gerd; Wood, Donna J.; Casper, Steven	1025-9473; 1012-6902; 0933-9361; 1012-8050	Corporate Identity; Unternehmenskultur; Unternehmen; Öffentlichkeitsarbeit
122	Schwarz, Karl; Roloff, Juliane; Bedau, Klaus-Dietrich; Wendt, Hartmut	0340-2398; 0479-611X; 0012-1304; 0170-1754	Bundesrepublik Deutschland; Bevölkerung; Entwicklung; historische Entwicklung
123	-	-	-
124	Häder, Sabine; Kunz, Gerhard; Gabler, Siegfried	0723-5607; 0941-1670; 0170-1754	Telefoninterview; Befragung; Stichprobe; Telefon
125	Pletzer, Uta; Wild, Helga de; Findl, Peter; Hofacker, Gabriele	0005-7215; 0029-9960; 0344-5550; 1437-4196	Säugling; Sterblichkeit; Ursache; Lebenserwartung
126	Jesse, Eckhard; Backes, Uwe; Butterwegge, Christoph; Jaschke, Hans-Gerd	0941-5491; 0323-3227; 0479-611X; 0177-6762	Bundesrepublik Deutschland; Kultur; historische Entwicklung; DDR
127	Walwei, Ulrich; Keller, Berndt; Werner, Heinz; Schmid, Josef	0340-1758; 0554-5455; 0479-611X; 0170-0847	Wahlverhalten; Wähler; Bundestagswahl; Parteiensystem
128	Dorbritz, Jürgen; Schwarz, Karl; Onnen-Isemann, Corinna; Nave-Herz, Rosemarie	0948-6704; 0935-218X; 0479-611X; 1013-1469	Zufriedenheit; alter Mensch; Lebensqualität; Lebenssituation

## A.1 Ermittelte Entitäten für GIRT

topic	author	issn	subject
129	Kißler, Leo; Naschold, Frieder; Klages, Helmut; Wollmann, Hellmut	0932-8114; 0722-0189; 0342-9857; 0034-3536	Frau; Bundesrepublik Deutschland; Kind; Jugendlicher
130	Eckert, Roland; Heitmeyer, Wilhelm; Schubarth, Wilfried; Willems, Helmut	-	Wasser; Wasserwirtschaft; Entwicklungsland; Asien
131	Vogel, Dita; Werner, Heinz; Wollenschläger, Michael; Lederer, Harald W.	0030-9273; 0028-3355; 0342-9857; 0044-3247	Schule; Kind; Mehrsprachigkeit; interkulturelle Erziehung
132	Hurrelmann, Klaus; Nauck, Bernhard; Aufenanger, Stefan; Bertram, Hans	0932-8114; 0340-8469; 0937-9614; 0170-0537	Kind; Sexualität; sexueller Missbrauch; Gewalt
133	Schniedewind, Karen	0479-611X; 0170-1754; 0010-3497; 0173-5993	Bundesrepublik Deutschland; Kirche; historische Entwicklung; Staat
134	Fthenakis, Wassilios E.	0479-611X; 0945-2419; 0046-970X; 0030-6428	internationale Wirtschaftsbeziehungen; internationale Beziehungen; bilaterale Beziehungen; Außenhandel
135	Walwei, Ulrich; Schömann, Klaus; Linne, Gudrun; Voswinkel, Stephan	0479-611X; 0012-0618; 0043-6143; 0007-5868	postsozialistisches Land; Entwicklungsland; Bundesrepublik Deutschland; UdSSR-Nachfolgestaat
136	Hauser, Richard; Ostner, Ilona; Schmid, Josef; Leibfried, Stephan	0722-8333; 0479-611X; 0006-4416; 0342-300X	Umweltpolitik; Ökologie; Umweltschutz; Umweltfaktoren
137	Lewin, Karl; Engelbrech, Gerhard; Buttler, Friedrich; Bertram, Barbara	0479-611X; 0863-1808; 0933-9361; 0323-3227	Bundesrepublik Deutschland; historische Entwicklung; Frau; Politik
138	Hurrelmann, Klaus; Aufenanger, Stefan; Krüger, Heinz-Hermann; Mansel, Jürgen	-	-
139	Hurrelmann, Klaus; Wilke, Jürgen; Gleich, Uli; Wagner, Gert	0943-1853	Gesundheitswesen; Ökonomie; Gesundheitspolitik; gesetzliche Krankenversicherung
140	Ostner, Ilona; Engelbrech, Gerhard; Wiesenthal, Helmut; Bartunek, Ewald	0006-4416; 0479-611X; 0945-2419; 0340-1758	Entwicklungsland; postsozialistisches Land; Asien; UdSSR-Nachfolgestaat

A *Ermittelte Entitäten der Kookurenzanalyse*

topic	author	issn	subject
141	Hauser, Richard; Leisering, Lutz; Hanesch, Walter; Bäcker, Gerhard	0341-7069; 0342-9857	Straßenkind; Obdachlosigkeit; Streetwork; Verwahrlosung
142	Reuband, Karl-Heinz; Smaus, Gerlinda; Cremer-Schäfer, Helga; Engelbrech, Gerhard	0941-5491; 0010-3497; 0173-5993; 0723-399X	Fernsehen; Wirkung; Bundesrepublik Deutschland; Massenmedien
143	Bergmann, Werner; Erb, Rainer; Stern, Frank; Benz, Wolfgang	0170-2602; 0943-1853; 0479-611X; 0170-7337	Rauchen; Bundesrepublik Deutschland; Tabakkonsum; DDR
144	Oswald, Hans; Krappmann, Lothar; Krüger, Helga; Kaiser, Astrid	0170-1754; 0173-5993; 0941-5491; 0171-3957	Hörfunk; Nutzung; Rundfunk; Inhalt
145	Jürgens, Ulrich; Naschold, Frieder; Fürstenberg, Friedrich; Ernst, Angelika	0479-611X; 0006-4416; 0340-0425; 0342-300X	soziale Ungleichheit; Einkommen; Sozialpolitik; Bundesrepublik Deutschland
146	Rammert, Werner; Klingler, Walter; Tully, Claus J.; Hörning, Karl H.	0170-2602	chronische Krankheit; Krankheit; Gesundheitsvorsorge; DDR
147	Seifert, Wolfgang; Münz, Rainer; Boos-Nünning, Ursula; Oswald, Hans	1012-6902; 0170-1754; 0173-5993; 0941-5491	Bundesrepublik Deutschland; Sport; historische Entwicklung; Fernsehen
148	Eckert, Roland; Friedrich, Walter; Schubarth, Wilfried; Willems, Helmut	0479-611X; 0933-1883; 0941-5491; 0323-3227	Minderheit; ethnische Gruppe; postsozialistisches Land; Deutscher
149	Hurrelmann, Klaus; Krüger, Heinz-Hermann; Groebel, Jo; Wiedemann, Dieter	0479-611X; 0006-4416; 0323-3227; 0044-2828	UdSSR; postsozialistisches Land; UdSSR-Nachfolgestaat; Drittes Reich
150	Hurrelmann, Klaus; Nauck, Bernhard; Bertram, Hans; Grundmann, Matthias	0170-1754; 0943-4755; 0173-5993; 0936-7780	Medienverhalten; Rezipientenforschung; Nutzung; Fernsehen
151	Kasten, Hartmut	0479-611X; 0340-1758; 0006-4416; 0507-4150	Rechtsradikalismus; Bundesrepublik Deutschland; Partei; historische Entwicklung
152	Schober, Karen; Hurrelmann, Klaus; Kloas, Peter-Werner; Jeschek, Wolfgang	0479-611X; 0342-300X; 0006-4416; 0340-3254	EU; Bundesrepublik Deutschland; Europa; Sozialpolitik
153	Klein, Paul; Timmermann, Heinz; Bertram, Hans; Bock, Hans Manfred	0479-611X; 0340-2398; 0016-5875; 0012-0618	Bundesrepublik Deutschland; Familie; neue Bundesländer; Frau
154	Audretsch, David B.; Albach, Horst; Belitz, Heike; Berthoin Antal, Ariane	0479-611X; 0342-300X; 0003-9209; 0006-4416	Bundesrepublik Deutschland; öffentliche Verwaltung; Reform; historische Entwicklung

## A.1 Ermittelte Entitäten für GIRT

topic	author	issn	subject
155	Meyer, Dagmar; Vaskovics, Laszlo A.; Pinther, Arnold; Rost, Harald	0479-611X; 0170-1754; 0171-3957; 0932-3244	Familie; Frau; Bundesrepublik Deutschland; Kind
156	Senghaas, Dieter; Eckert, Roland; Habermas, Jürgen; Mögelin, Chris	0479-611X; 0722-8333; 0507-4150; 0006-4416	Bundesrepublik Deutschland; Einwanderung; Ausländerrecht; Migration
157	Wingen, Max; Schulz, Erika; Schweitzer, Rosemarie von; Lampert, Heinz	0933-1883; 0943-4755; 0171-3957; 0723-399X	Kind; Schule; Familie; Bundesrepublik Deutschland
158	Engelbrech, Gerhard; Nickel, Hildegard Maria; Ostner, Ilona; Notz, Gisela	1011-386X	Auswanderung; Herkunftsland; Migration; Entwicklungsland
159	Diekmann, Andreas; Preisendörfer, Peter; Spellerberg, Annette; Lüdtke, Hartmut	-	Vater; Eltern-Kind-Beziehung; Familie; Kind
160	Bonß, Wolfgang; Beck, Ulrich; Imhof, Arthur E.; Japp, Klaus P.	0342-300X; 0479-611X; 0943-7525; 0340-7918	Arbeitsverhältnis; befristetes Arbeitsverhältnis; Bundesrepublik Deutschland; Flexibilität
161	Klages, Helmut; Gensicke, Thomas; Sterbling, Anton; Müller, Klaus	0479-611X; 0342-300X; 0006-4416; 0007-5868	Europa; EU; Bundesrepublik Deutschland; EG
162	Nuscheler, Franz; Straubhaar, Thomas	0479-611X; 0340-3254; 0341-4515; 0723-8525	Bundesrepublik Deutschland; Frau; DDR; Familie
163	Mollenkopf, Heidrun; Backes, Gertrud M.; Naegele, Gerhard; Lehr, Ursula	0171-3957; 0723-399X; 0342-9857; 0937-9614	Jugendlicher; Bundesrepublik Deutschland; Kind; Verhalten
164	Ulrich, Gisela; Geier, Wolfgang; Schauer, Heinz; Voß, Peter	0479-611X; 0170-1754; 0006-4416; 0177-6762	Bundesrepublik Deutschland; historische Entwicklung; neue Bundesländer; DDR
165	-	0479-611X; 0937-9614; 0342-300X; 0722-8333	Bundesrepublik Deutschland; Familie; Kind; Frau
166	Timmermann, Heinz; Meier, Christian; Ignatow, Assen; Knabe, Bernd	0479-611X; 0006-4416; 0342-300X; 0490-1606	Bundesrepublik Deutschland; Sozialpolitik; Arbeitslosigkeit; neue Bundesländer
167	Huinink, Johannes; Reißig, Rolf; Woderich, Rudolf; Dorbritz, Jürgen	0479-611X; 0341-1966; 0023-2653; 0023-4834	Frau; Bundesrepublik Deutschland; geschlechtsspezifische Faktoren; Gewalt

A *Ermittelte Entitäten der Kookkurenzanalyse*

topic	author	issn	subject
168	Huinink, Johannes; Nickel, Hildegard Maria; Blossfeld, Hans-Peter; Krüger, Helga	0479-611X; 0006-4416; 0932-8408; 0340-613X	Bundesrepublik Deutschland; Nationalsozialismus; historische Entwicklung; Drittes Reich
169	Menzel, Ulrich; Braun, Gerald; Rüländ, Jürgen; Senghaas, Dieter	0720-4361; 0723-399X; 0342-183X; 0946-5596	Bundesrepublik Deutschland; Schule; geschlechtsspezifische Faktoren; Schüler
170	Schulte, Axel; Schaller, Christian; Otero, Jose Sanchez; Gloeckner, Eduard	0479-611X; 0016-5492; 0941-6838; 0938-152X	Ostasien; Asien; Bundesrepublik Deutschland; Management
171	Wagner, Gert; Jesse, Eckhard; Seifert, Wolfgang; Werner, Heinz	0171-3957; 0170-1754; 0723-399X; 0479-611X	EDV; Nutzung; Kommunikation; Bundesrepublik Deutschland
172	Rosenbrock, Rolf; Lenhardt, Uwe; Trojan, Alf; Hurrelmann, Klaus	0479-611X; 0170-1754; 0023-2653; 0177-6762	Bundesrepublik Deutschland; Kind; Schule; Schüler
173	Bäcker, Gerhard; Bertram, Hans; Eggen, Bernd; Vaskovics, Laszlo A.	0479-611X; 0171-3957; 0723-399X; 0937-9614	Jugendlicher; Gewalt; Rechtsradikalismus; Jugend
174	Simonis, Udo E.; Walwei, Ulrich; Werner, Heinz; Weidenfeld, Werner	0170-1754; 0173-5993; 0479-611X; 0171-3957	Kind; Jugendlicher; Bundesrepublik Deutschland; Familie
175	Möller, Bärbel; Reißig, Rolf; Köhl, Jürgen; Brinkmann, Christian	0479-611X; 0171-3957; 0723-399X; 0937-9614	Bundesrepublik Deutschland; Kind; Jugendlicher; Schule
176	Rosenbrock, Rolf; Elkeles, Thomas; Bellmann, Lutz; Schober, Karen	-	Geschwister; Geschwisterreihe; Familie; Freundschaft
177	Schmid, Josef; Münch, Richard; Weidenfeld, Werner; Leggewie, Claus	0340-3254; 0723-8525; 0341-4515; 0479-611X	Jugendlicher; Bundesrepublik Deutschland; Berufsbildung; Arbeitslosigkeit
178	Beyme, Klaus von; Wiesenthal, Helmut; Merkel, Wolfgang; Wollmann, Hellmut	0479-611X; 0042-5702; 0006-4416; 0940-3566	Frankreich; internationale Beziehungen; bilaterale Beziehungen; Bundesrepublik Deutschland
179	Schneider, Eberhard; Timmermann, Heinz; Halbach, Uwe; Simon, Gerhard	0479-611X; 0722-7485; 0170-1754; 0938-152X	multinationales Unternehmen; Bundesrepublik Deutschland; Management; Globalisierung
180	Mattusch, Katrin; Baringhorst, Sigrid; Winkler, Beate; Heckmann, Friedrich	0479-611X; 1437-2940; 0340-2398; 0932-8114	Ehe; Familie; Familienplanung; Frau



## A.1 Ermittelte Entitäten für GIRT

topic	author	issn	subject
181	Diekmann, Andreas; Preisendörfer, Peter; Kösters, Walther; Bauer, Siegfried	0479-611X; 0023-4834; 0177-7521; 0863-4890	Menschenrechte; Bundesrepublik Deutschland; Entwicklungsland; Staat
182	Klauder, Wolfgang; Werner, Heinz; Kohler, Hans; Bach, Hans-Uwe	0479-611X; 0342-300X; 0012-1304; 0043-6143	Bundesrepublik Deutschland; Sozialpolitik; DDR; Entwicklung
183	Preisendörfer, Peter	0946-5596; 0722-0189; 0723-5186; 0004-1157	Bundesrepublik Deutschland; Frau; neue Bundesländer; geschlechtsspezifische Faktoren
184	Bette, Karl-Heinrich; Schimank, Uwe; Spitzer, Giselher; Brinkhoff, Klaus-Peter	0379-3664; 0587-5234; 0479-611X; 1011-0070	Umweltbewusstsein; Verhalten; Umwelt; Einstellung
185	Wießner, Frank; Goldberg, Andreas; Zapf, Wolfgang; Pollack, Detlef	0479-611X; 0006-4416; 0342-300X; 0038-6073	Risiko; Bundesrepublik Deutschland; Sicherheit; neue Bundesländer
186	Häussermann, Hartmut; Strohmeier, Klaus Peter; Dangschat, Jens S.; Friedrichs, Jürgen	0479-611X; 0863-1808; 0006-4416; 0863-4564	Wertwandel; Transformation; Bundesrepublik Deutschland; sozialer Wandel
187	Werner, Heinz; Münz, Rainer; Straubhaar, Thomas; Seifert, Wolfgang	0479-611X; 0945-2419; 0020-7985	Migration; Einwanderungspolitik; internationale Wanderung; Migrationspolitik
188	-	0479-611X; 0490-1606; 0948-6704; 1013-1469	alter Mensch; Bundesrepublik Deutschland; neue Bundesländer; Entwicklung
189	Wetzels, Peter; Pfeiffer, Christian; Naegele, Gerhard; Rosenmayr, Leopold	0171-3957; 0173-5993; 0170-1754; 0943-4755	Freizeit; Verhalten; Jugendlicher; schichtspezifische Faktoren
190	-	1941386	Sterblichkeit; Quantität; Bevölkerung; Fruchtbarkeit
191	Halbach, Uwe; Tibi, Bassam; Stauth, Georg; Wohlrab-Sahr, Monika	0479-611X; 0006-4416; 0030-6428; 0863-4890	postsozialistisches Land; Russland; UdSSR-Nachfolgestaat; Transformation
192	Dangschat, Jens S.; Kronauer, Martin; Bäcker, Gerhard; Hanesch, Walter	0479-611X; 0863-4564; 0012-1304; 0340-2398	neue Bundesländer; Bundesrepublik Deutschland; sozialer Wandel; Transformation
193	Werner, Andreas; Sandbothe, Mike; Breinker, Carsten; Klingler, Walter	0023-2653; 0946-5596; 0723-5186; 0029-9960	Frau; geschlechtsspezifische Faktoren; Bundesrepublik Deutschland; Karriere
194	-	0072-9566; 0938-152X; 0340-0425; 0006-4416	Schwellenland; Entwicklungsland; Asien; Taiwan

A *Ermittelte Entitäten der Kookkurenzanalyse*

topic	author	issn	subject
195	Girtler, Roland; Velten, Doris; Stallberg, Friedrich W.; Kleiber, Dieter	0479-611X; 0006-4416; 1615-5548; 0072-9566	Integration; Ausländerpolitik; Europapolitik; Europa
196	Schubarth, Wilfried; Hurrelmann, Klaus; Eckert, Roland; Tillmann, Klaus-Jürgen	0479-611X; 0170-1754; 0006-4416; 0722-8333	Bundesrepublik Deutschland; historische Entwicklung; Tourismus; neue Bundesländer
197	Ott, Erich; Wagner, Gert; Magvas, Emil; Schupp, Jürgen	0170-2602; 0342-300X; 0943-1853; 0724-3464	Betrieb; Gesundheitsvorsorge; Gesundheit; Bundesrepublik Deutschland
198	Charlton, Michael; Gleich, Uli; Groebel, Jo; Kübler, Hans-Dieter	0479-611X; 0342-300X; 0170-1754; 0722-8333	Bundesrepublik Deutschland; Familie; Frau; DDR
199	Walwei, Ulrich; Egle, Franz; Deeke, Axel; Bach, Heinz Willi	0479-611X; 0006-4416; 0945-2419; 0177-6762	Umweltpolitik; EU; internationale Politik; Klima
200	Hurrelmann, Klaus; Gerhardt, Uta; Schmidt, Bernd; Schaeffer, Doris	0479-611X; 0940-4678; 0342-300X; 0303-2493	neue Bundesländer; Bundesrepublik Deutschland; regionale Wirtschaftsförderung; Wirtschaftspolitik
201	Lindner, Bernhard; Stiehler, Hans-Jörg; Wiedemann, Dieter; Goeschel, Albrecht	0342-300X; 0340-3254; 0341-4515; 0031-5605	Bundesrepublik Deutschland; Gesundheit; Arbeitsbedingungen; DDR
202	Brettschneider, Frank; Gabriel, Oscar W.; Schultze, Rainer-Olaf; Plasser, Fritz	0479-611X; 0006-4416; 0170-1754; 1615-5548	Europa; Bundesrepublik Deutschland; historische Entwicklung; EU
203	Habich, Roland; Meulemann, Heiner; Klesse, Rosemarie; Papastefanou, Georgios	0479-611X; 0863-4890; 0342-8176; 0863-1808	Osteuropa; postsozialistisches Land; Demokratisierung; politischer Wandel
204	Schmidt, Bernd; Lautmann, Rüdiger; Starke, Kurt; Keune, Saskia	0479-611X; 0030-6428; 0012-0812; 0342-9857	UdSSR-Nachfolgestaat; Russland; postsozialistisches Land; Transformation
205	-	0012-0812; 0479-611X; 0014-2476	postsozialistisches Land; Lettland; Minderheit; Estland
206	Luchtenberg, Sigrid; Rauschenbach, Thomas; Gerth, Werner; Meier, Artur	0722-8333; 0943-1780; 0720-2946; 0340-1804	Umwelterziehung; Umweltbewusstsein; Umweltverträglichkeit; Umweltpolitik
207	Wetzels, Peter; Heiliger, Anita; Neubauer, Georg; Schetsche, Michael	0479-611X; 0340-3254; 0945-2419; 0940-9211	Umweltpolitik; Ökologie; Umweltschaden; Umweltschutz

## A.1 Ermittelte Entitäten für GIRT

topic	author	issn	subject
208	Pollack, Detlef; Ebertz, Michael N.; Gabriel, Karl; Altvater, Elmar	-	horizontale Mobilität; Freizeit; Kraftfahrzeug; Freizeitpolitik
209	Meier, Christian; Höhmann, Hans-Hermann; Jacobsen, Hanns-Dieter; Surubovic, Aleksej	1012-6902; 0170-1754; 0173-5993; 0342-2380	Leistungssport; Sportler; Sportpolitik; Sportverband
210	Elsenhans, Hartmut; Walwei, Ulrich; Werner, Heinz; Schmähl, Winfried	0479-611X; 0340-3254; 0006-4416; 0171-645X	neue Bundesländer; Bundesrepublik Deutschland; alte Bundesländer; Unternehmensgründung
211	Weizsäcker, Ernst Ulrich von; Simonis, Udo E.; Beck, Ulrich; Jänicke, Martin	0479-611X; 0170-9364; 0003-9209; 0034-0111	Bundesrepublik Deutschland; Stadt; Entwicklung; Bevölkerung
212	Vogt, Ludgera; Zingerle, Arnold; Frevert, Ute; Luhmann, Niklas	0479-611X; 0340-3254; 0021-4027; 0006-4416	Bundesrepublik Deutschland; Entwicklung; Einwanderung; ausländischer Arbeitnehmer
213	-	-	-
214	-	0479-611X; 0170-1754; 0490-1606; 0948-6704	Bundesrepublik Deutschland; Jugendlicher; alter Mensch; Frau
215	Halbach, Uwe; Götz, Roland; Beyme, Klaus von; Beck, Ulrich	-	Tabak; Werbung; Rauchen; Wirkung
216	Dücker, Uwe von; Zinnecker, Jürgen; Walger, Christian; Specht, Walther	0479-611X; 0006-4416; 0863-4890; 0004-8194	Islam; Asien; Europa; Entwicklungsland
217	Gleich, Uli; Groebel, Jo; Neumann-Braun, Klaus; Ulrich, Peter	0479-611X; 0006-4416; 0342-9857; 0490-1606	Bundesrepublik Deutschland; Sozialpolitik; soziale Ungleichheit; Arbeitslosigkeit
218	Reißig, Monika; Troschke, Jürgen von; Stiehr, Karin; Riemann, Klaus	0005-9455; 0177-6762; 0173-5993; 0170-1754	Bundesrepublik Deutschland; interaktive Medien; Online-Dienst; Computer
219	Klingler, Walter; Oehmichen, Ekkehardt; Eimeren, Birgit van; Krekeler, Michael	0342-183X	Begabung; Begabtenförderung; Intelligenz; Schüler
220	Huster, Ernst-Ulrich; Dangschat, Jens S.; Hauser, Richard; Krause, Peter	0341-2059; 0932-8114; 0172-6404	Prostitution; Bundesrepublik Deutschland; Prophylaxe; AIDS
221	-	0479-611X; 0171-3957; 0723-399X; 0937-9614	Jugendlicher; Schüler; Bundesrepublik Deutschland; Kind

## A Ermittelte Entitäten der Kookkurenzanalyse

topic	author	issn	subject
222	Luhmann, Niklas; Beck, Ulrich; Pollack, Detlef; Zapf, Wolfgang	0012-1304	Pendler; Nahwanderung; Bundesrepublik Deutschland; Berufsmobilität
223	Dietz, Barbara; Hilkes, Peter; Jäger, Siegfried; Bourdieu, Pierre	0171-3957; 0943-4755; 0170-1754; 0723-399X	Kind; Fernsehen; Nutzung; Bundesrepublik Deutschland
224	Bergmann, Werner; Erb, Rainer; Halbach, Uwe; Timmermann, Heinz	0343-1886; 0340-3254; 0723-8525; 0007-585X	Arbeitsverwaltung; Monopol; privater Sektor; Arbeitsmarktpolitik
225	Klingler, Walter; Schmidt, Claudia; Zöllner, Oliver; Grajczyk, Andreas	0948-6704; 0943-1853; 0170-2602; 0479-611X	Krankheit; chronische Krankheit; Gesundheit; Prophylaxe

## A.2 Ermittelte Entitäten für iSearch

topic	author	source
1	Mostepanenko, V. M.; Steinacker, Harold; Herrera, L.; Esposito, Giampiero	Phys Rev Lett; Phys Rev E; Phys Rev B; Phys Rev A
2	Esposito, Giampiero; de Mello, E. R. Bezerra; Herrera, L.; Jr.	Phys Rev E; Phys Rev Lett; Phys Rev B; Phys Rev D
3	Zhang, Y.; Visser, Matt; Bleicher, Marcus; Helbing, Dirk	Nucl Instrum Meth A; Phys Rev Lett; Phys Rev A; Phys Rev D
4	Dietrich, S.; Thirumalai, D.; Mostepanenko, V. M.; Shukla, P. K.	Phys Rev B; Phys Rev Lett; Phys Rev E; Phys Rev D
5	-	-
6	Shukla, P. K.; Vanderbilt, David; Marklund, M.; Berret, Jean-Francois	Phys Rev B; Phys Rev Lett; Phys Rev D; Phys Rev E
7	Kevrekidis, P. G.; Visser, Matt; Esposito, Giampiero; Jr.	Phys Rev D; Phys Lett B; Nucl Phys B; Phys Rev Lett
8	Iorio, Lorenzo; Zoller, P.; Jr.; Dadhich, Naresh	Phys Rev Lett; Phys Rev A; Phys Rev B; Phys Rev D
9	Bazant, Martin Z.; Loidl, A.; Vanderbilt, David; Prellier, W.	Phys Rev E; Phys Rev B; Phys Rev Lett; Phys Rev A
10	Dai, Hongjie; Bazant, Martin Z.; Delogu, P.; Fantacci, M. E.	Phys Rev E; Phys Rev Lett; Phys Rev A; Phys Rev B
11	Bazant, Martin Z.; Bleicher, M.; Prellier, W.; Dai, Hongjie	Phys Rev E; Phys Rev A; Phys Rev B; Phys Rev Lett
12	Thirumalai, D.; Oset, E.; Jr.; Ben-Naim, E.	Phys Rev D; Phys Rev Lett; Phys Rev C; Phys Rev E
13	Zhang, Xinmin; Gracey, J. A.; Neto, A. H. Castro; Riotto, Antonio	Phys Rev D; Phys Lett B; Phys Rev Lett; Nucl Phys B

topic	author	source
14	Bazant, Martin Z.; Sachdev, Subir; Peeters, F. M.; Oset, E.	Phys Rev B; Phys Rev Lett; Phys Rev D; Phys Lett B
15	Dai, Hongjie; Liu, Y.; Tokura, Y.; Jonson, M.	Phys Rev B; Phys Rev Lett; Phys Rev E; Appl Phys Lett
16	Schierholz, G.; Laine, M.; Sachrajda, C. T.; Gäckeler, M.	Phys Rev D; Phys Rev B; Phys Rev Lett; Phys Lett B
17	Bauer, Gerrit E. W.; Brataas, Arne; Abdalla, M. C. B.; Vancea, I. V.	Phys Rev B; Phys Rev D; Phys Rev Lett; Phys Lett B
18	Fisher, I. R.; Ebert, D.; Dai, Hongjie; Chen, X. H.	Phys Rev B; Phys Rev D; Phys Rev Lett; Phys Rev A
19	Vanderbilt, David; Kozłowski, Mirosław; Shukla, P. K.; Lewenstein, M.	Phys Rev Lett; Phys Rev B; Phys Rev A; Phys Rev D
20	Kivshar, Yuri S.; Yamamoto, Yoshihisa; Morigi, Giovanna; Hofmann, Holger F.	Phys Rev Lett; Phys Rev B; Phys Rev A; Phys Rev D
21	Kivshar, Yuri S.; Jr.; Yamamoto, Yoshihisa; Zoller, P.	Phys Rev Lett; Phys Rev A; Phys Rev D; Phys Rev B
22	Flambaum, V. V.; Jr.; Zoller, P.; Yamamoto, Yoshihisa	Phys Rev Lett; Phys Rev B; Phys Rev A; Phys Rev D
23	Kivshar, Yuri S.; Milburn, G. J.; Morigi, Giovanna; Zoller, P.	Phys Rev B; Phys Rev D; Phys Rev Lett; Phys Rev A
24	Taniguchi, Y.; Morris, Tim R.; Schierholz, G.; Jr.	Phys Rev D; Phys Rev Lett; Nucl Phys B; Phys Rev B
25	Rosu, H. C.; Schutz, Bernard F.; Sahay, S. K.; Soshnikov, V. N.	Phys Rev D; Phys Rev A; Phys Rev Lett; Phys Rev B
26	Vanderbilt, David; Glinchuk, M. D.; Eliseev, E. A.; Jonson, M.	Phys Rev B; Phys Rev Lett; Physical Review B; Phys Rev E
27	Loidl, A.; Ye, Jun; Chen, Yanbei; Oset, E.	Phys Rev Lett; Phys Rev D; Phys Rev B; Phys Lett B
28	-	-
29	Loss, Daniel; Wegscheider, W.; Ritchie, D. A.; Altshuler, B. L.	Phys Rev A; Phys Rev Lett; Phys Rev B; Phys Rev D
30	Canfield, P. C.; Nojiri, Shin'ichi; Deser, S.; Barrow, John D.	Phys Rev B; Phys Rev Lett; Phys Rev D; Class Quant Grav
31	Vanderbilt, David; Ritchie, D. A.; Maude, D. K.; Kotthaus, J. P.	Phys Rev B; Phys Rev Lett; Appl Phys Lett; Physical Review B
32	Reynaud, Serge; Pinard, M.; Yau, Shing-Tung; Heidmann, A.	Phys Rev A; Phys Rev Lett; Phys Rev B; Phys Rev D
33	Aldaya, V.; Marmo, G.; de Boer, Jan; Landsman, N. P.	Phys Rev D; Nucl Phys B; JHEP; Phys Lett B
34	Freedman, Daniel Z.; Ellis, John; Hu, B. L.; Elizalde, E.	Phys Rev D; Nucl Phys B; JHEP; Phys Lett B

## A Ermittelte Entitäten der Kookkurenzanalyse

topic	author	source
35	Witten, Edward; Sen, Ashoke; Nojiri, Shin'ichi; de Boer, Jan	JHEP; Nucl Phys B; Phys Rev D; Phys Lett B
36	Petcov, S. T.; Grimus, W.; Ellis, John; Kim, C. S.	Phys Rev D; Phys Lett B; Phys Rev Lett; Nucl Phys B
37	Iorio, Lorenzo; Ritchie, D. A.; Wegscheider, W.; Kidonakis, Nikolaos	Phys Rev Lett; Phys Rev D; Phys Rev B; Phys Lett B
38	Anlage, Steven M.; Peskov, V.; Lin, C. D.; Davis, J. C.	Nucl Instrum Meth A; Phys Rev Lett; Phys Rev A; Phys Rev B
39	Melnikov, V. N.; Ellis, John; Ivashchuk, V. D.; Ellis, J.	Phys Rev D; Phys Lett B; Nucl Phys B; JHEP
40	Boorchand, P.; Yoshikawa, Kenichi; Jr.; Sachdev, Subir	Phys Rev B; Phys Rev Lett; Phys Rev E; Phys Rev D
41	Ellis, John; Kim, C. S.; King, S. F.; Nanopoulos, D. V.	Phys Rev D; Phys Rev Lett; Phys Lett B; JHEP
42	Libbrecht, Kenneth G.; Cava, R. J.; Ben-Naim, E.; Peeters, F. M.	Phys Rev Lett; Phys Rev B; Phys Rev D; Phys Lett B
43	Ellis, John; Hooper, Dan; Liu, Y.; Moroi, Takeo	Phys Rev D; Phys Lett B; Phys Rev Lett; JHEP
44	Schierholz, G.; Volovik, G. E.; Suzuki, T.; Helbing, Dirk	Phys Rev D; Phys Rev Lett; Phys Rev E; Phys Lett B
45	Rapp, R.; Blaschke, D.; Laine, M.; Bleicher, M.	Phys Rev C; Nucl Phys A; Phys Rev D; Phys Lett B
46	Royon, C.; Levin, E.; Szymanowski, L.; Maor, U.	Phys Rev D; Phys Rev Lett; Phys Lett B; Phys Rev A
47	Tokura, Y.; Chen, X. H.; Cava, R. J.; Eisaki, H.	Phys Rev B; Phys Rev Lett; Appl Phys Lett; Physical Review B
48	Canfield, P. C.; Anisimov, V. I.; Sachdev, Subir; Felser, C.	Phys Rev B; Phys Rev Lett; Phys Rev A; Physical Review B
49	Canfield, P. C.; Chen, X. H.; Mavromatos, N. E.; Sachdev, Subir	Phys Rev B; Phys Rev Lett; Phys Rev D; Phys Rev A
50	Chen, X. H.; Canfield, P. C.; Sachdev, Subir; Jr.	Phys Rev B; Phys Rev Lett; Phys Rev D; Phys Rev A
51	Canfield, P. C.; Awana, V. P. S.; Dai, Hongjie; Kishan, H.	Phys Rev B; Phys Rev D; Appl Phys Lett; Phys Rev Lett
52	Canfield, P. C.; Chen, X. H.; Cava, R. J.; Jr.	Phys Rev B; Phys Rev Lett; Phys Rev D; Phys Lett B
53	Peskov, V.; Iorio, Lorenzo; Turyshev, Slava G.; Dai, Hongjie	Phys Rev Lett; Nucl Instrum Meth A; Phys Rev D; Phys Rev B
54	Moroi, Takeo; Kiselev, V. V.; Rischke, Dirk H.; Mohanty, Subhendra	Phys Rev B; Phys Rev Lett; Phys Rev D; Nucl Instrum Meth A
55	Miller, G. A.; Lyon, S. A.; Oset, E.; Chen, X. H.	Phys Rev B; Phys Rev Lett; Phys Rev A; Phys Rev C

## A.2 Ermittelte Entitäten für iSearch

topic	author	source
56	Oset, E.; Yamamoto, Yoshihisa; Zoller, P.; Lewenstein, M.	Phys Rev B; Phys Rev Lett; Phys Rev A; Appl Phys Lett
57	Avouris, Phaedon; Dai, Hongjie; Perebeinos, Vasili; Louie, Steven G.	Phys Rev B; Phys Rev Lett; Physical Review B; J Phys Condens Matter
58	Kim, Sang Pyo; Laine, M.; Scherer, S.; Esposito, Giampiero	Phys Rev D; Phys Rev B; Phys Rev Lett; Phys Lett B
59	Loidl, A.; Loss, Daniel; Sachdev, Subir; Peeters, F. M.	Phys Rev B; Phys Rev Lett; Phys Rev A; Phys Rev D
60	Ambjorn, J.; Loll, R.; Slater, Paul B.; Pfeiffer, Hendryk	Nucl Phys B; J Math Phys; Phys Rev D; Phys Lett B
61	Martinez, R.; Soni, Amarjit; Cheng, Hai-Yang; He, Xiao-Gang	Nucl Phys B; Phys Rev D; Phys Lett B; JHEP
62	Ambjorn, J.; Loll, R.; Smolin, Lee; Carfora, M.	Nucl Phys B; Phys Rev D; Phys Lett B; Class Quant Grav
63	Oset, E.; Kim, C. S.; He, Xiao-Gang; Mavromatos, N. E.	Phys Rev D; Phys Lett B; Nucl Phys B; Phys Rev Lett
64	Dai, Hongjie; Ritchie, D. A.; Bagraev, N. T.; Peeters, F. M.	Phys Rev B; Phys Rev Lett; Phys Rev D; Phys Lett B
65	-	-





## **B Ermittelte Werte der Power-Law-Analyse**

Die folgenden Plots und tabellarisch dargestellten Werte entstanden auf Grundlage der in den Abschnitten 6.1.9 und 8.2 beschriebenen Methoden zur Bestimmung einer Power-Law-Verteilung.

## B.1 Produktivität der Autoren in GIRT

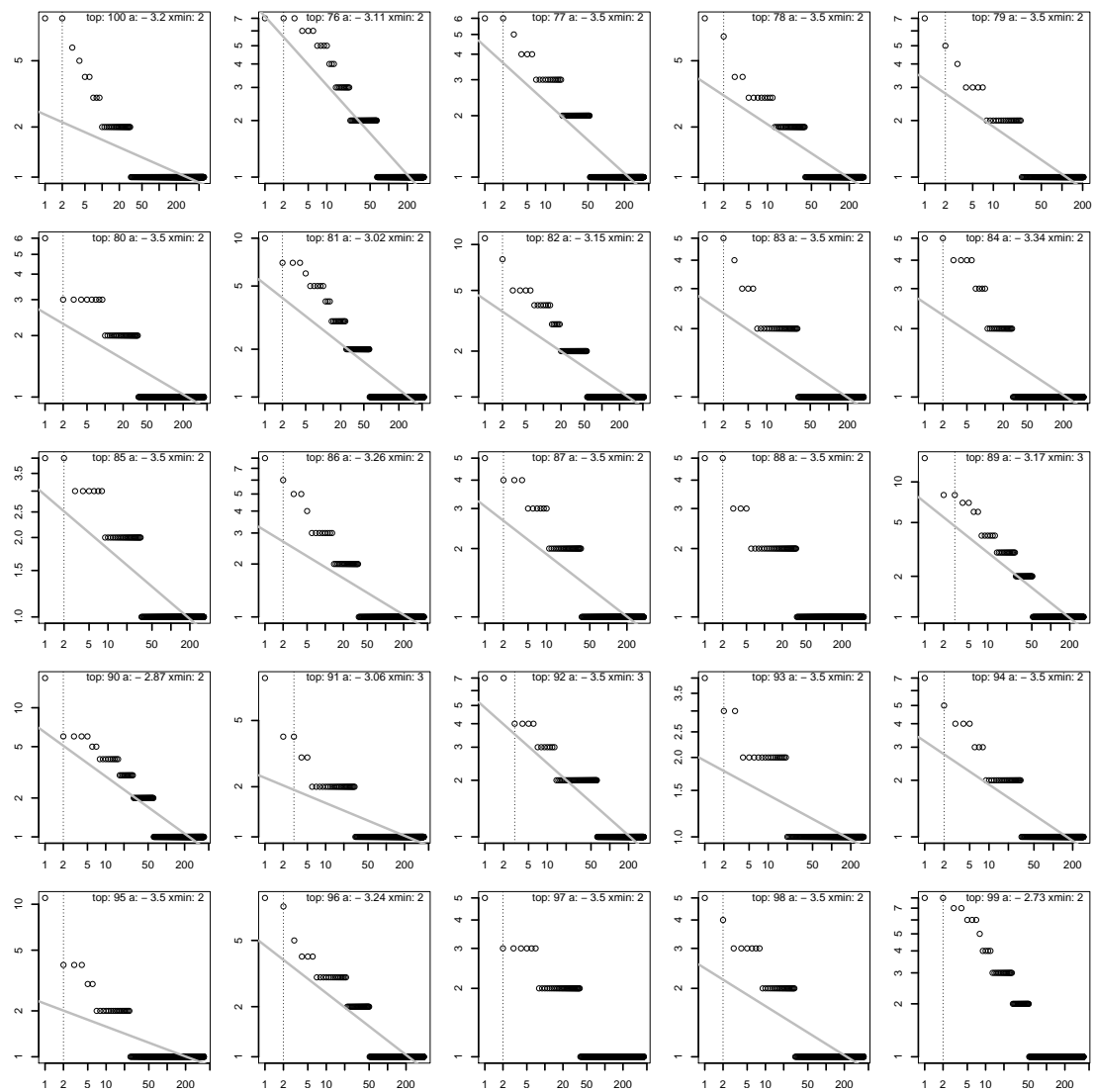


Abbildung B.1: Doppelt logarithmierter Plot der Produktivität der Autoren für die Topics 76-100 (CLEF Jahrgang 2003).

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
76	author	3,11	0,048	2	<b>0,334</b>	0,048
77	author	3,5	0,037	2	<b>0,678</b>	0,037
78	author	3,5	0,031	2	<b>0,89</b>	0,031
79	author	3,5	0,023	2	<b>0,999</b>	0,023
80	author	3,5	0,105	2	<b>0,116</b>	0,105
81	author	3,02	0,043	2	<b>0,494</b>	0,043
82	author	3,15	0,06	2	<b>0,251</b>	0,06
83	author	3,5	0,109	2	<b>0,139</b>	0,109
84	author	3,34	0,063	2	<b>0,556</b>	0,063
85	author	3,5	0,073	2	<b>0,33</b>	0,073
86	author	3,26	0,033	2	<b>0,893</b>	0,033
87	author	3,5	0,044	2	<b>0,716</b>	0,044
88	author	3,5	0,156	2	0,022	0,156
89	author	3,17	0,072	3	<b>0,35</b>	0,072
90	author	2,87	0,054	2	<b>0,317</b>	0,054
91	author	3,06	0,117	3	<b>0,681</b>	0,117
92	author	3,5	0,085	3	<b>0,535</b>	0,085
93	author	3,5	0,145	2	<b>0,124</b>	0,145
94	author	3,5	0,045	2	<b>0,764</b>	0,045
95	author	3,5	0,053	2	<b>0,771</b>	0,053
96	author	3,24	0,079	2	<b>0,115</b>	0,079
97	author	3,5	0,118	2	0,08	0,118
98	author	3,5	0,067	2	<b>0,487</b>	0,067
99	author	2,73	0,097	2	0,056	0,097
100	author	3,2	0,057	2	<b>0,579</b>	0,057

Tabelle B.1: Einzelwerte der Produktivität der Autoren für die Topics 76-100 (CLEF Jahrgang 2003). Statistisch signifikante Werte sind **fett** gedruckt.

B *Ermittelte Werte der Power-Law-Analyse*

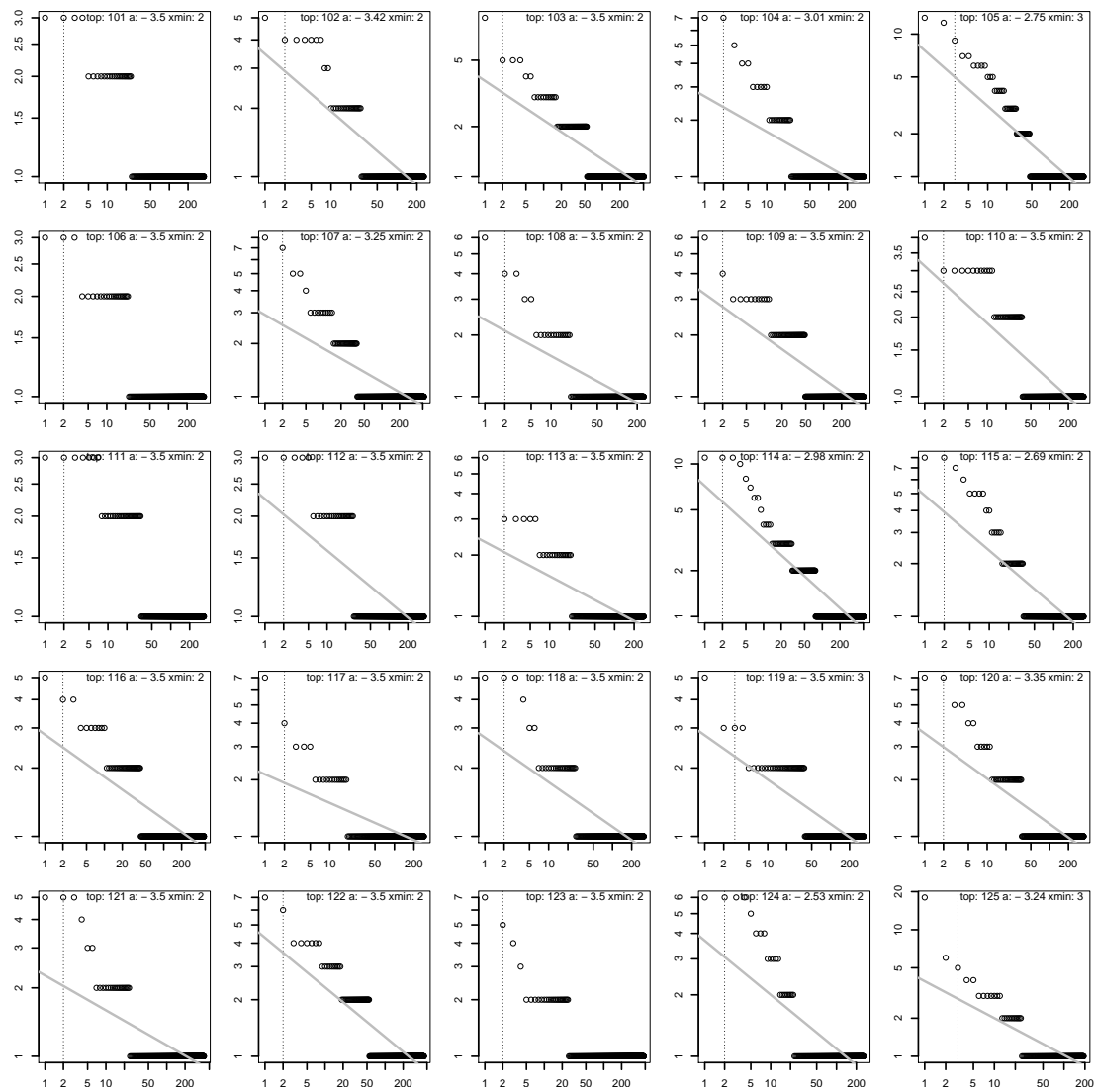


Abbildung B.2: Doppelt logarithmierter Plot der Produktivität der Autoren für die Topics 101-125 (CLEF Jahrgang 2004).

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
101	author	3,5	0,136	2	0,09	0,136
102	author	3,42	0,108	2	<b>0,16</b>	0,108
103	author	3,5	0,025	2	<b>0,904</b>	0,025
104	author	3,01	0,043	2	<b>0,896</b>	0,043
105	author	2,75	0,08	3	<b>0,307</b>	0,08
106	author	3,5	0,166	2	0,042	0,166
107	author	3,25	0,039	2	<b>0,815</b>	0,039
108	author	3,5	0,039	2	<b>0,972</b>	0,039
109	author	3,5	0,092	2	<b>0,125</b>	0,092
110	author	3,5	0,106	2	<b>0,122</b>	0,106
111	author	3,5	0,134	2	0,06	0,134
112	author	3,5	0,134	2	<b>0,122</b>	0,134
113	author	3,5	0,086	2	<b>0,433</b>	0,086
114	author	2,98	0,03	2	<b>0,684</b>	0,03
115	author	2,69	0,076	2	<b>0,35</b>	0,076
116	author	3,5	0,059	2	<b>0,446</b>	0,059
117	author	3,5	0,027	2	<b>1</b>	0,027
118	author	3,5	0,072	2	<b>0,522</b>	0,072
119	author	3,5	0,192	3	<b>0,521</b>	0,192
120	author	3,35	0,03	2	<b>0,942</b>	0,03
121	author	3,5	0,053	2	<b>0,772</b>	0,053
122	author	3,5	0,035	2	<b>0,716</b>	0,035
123	author	3,5	0,143	2	0,084	0,143
124	author	2,53	0,115	2	<b>0,253</b>	0,115
125	author	3,24	0,07	3	<b>0,663</b>	0,07

Tabelle B.2: Einzelwerte der Produktivität der Autoren für die Topics 101-125 (CLEF Jahrgang 2004). Statistisch signifikante Werte sind **fett** gedruckt.

B Ermittelte Werte der Power-Law-Analyse

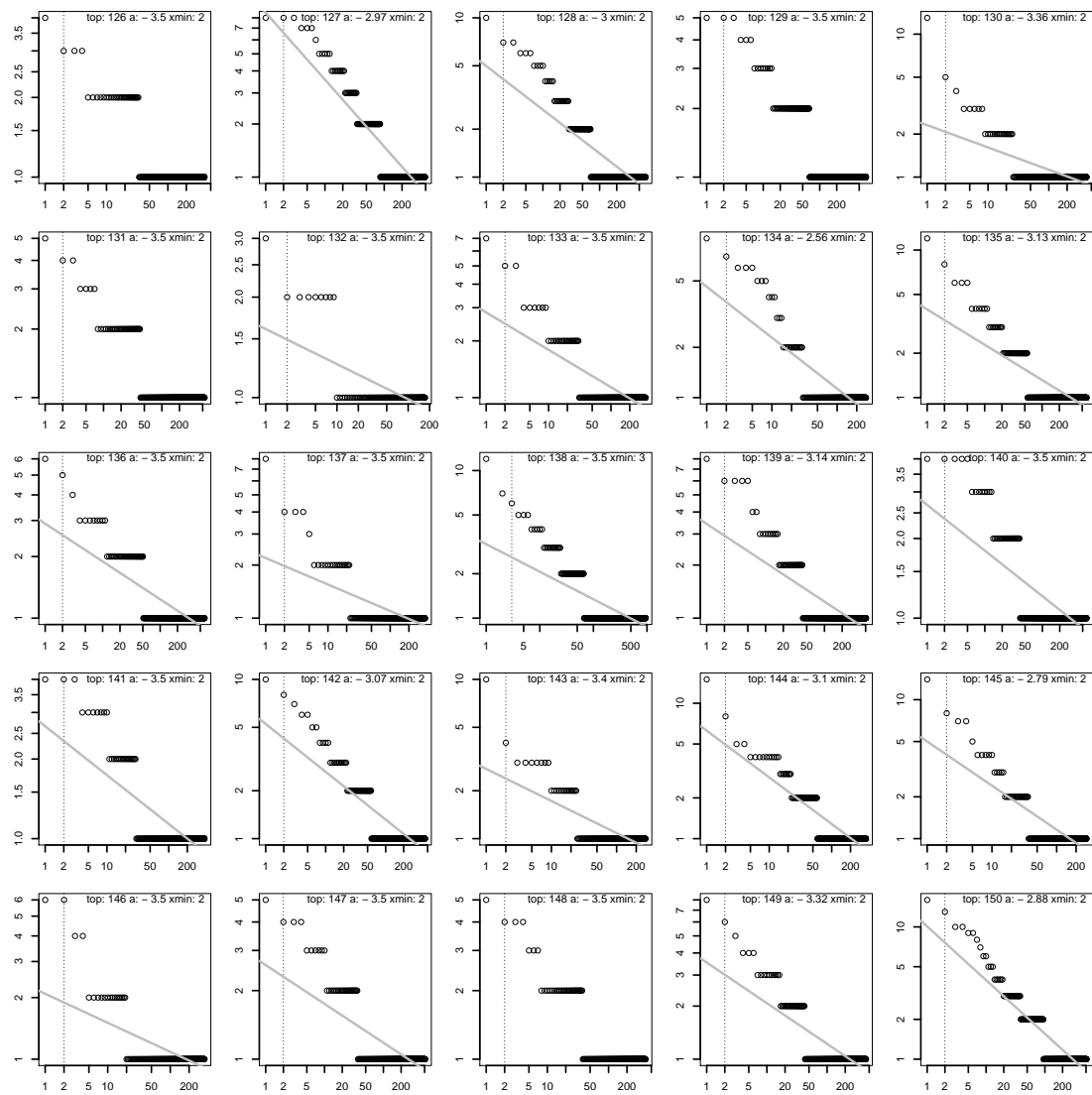


Abbildung B.3: Doppelt logarithmierter Plot der Produktivität der Autoren für die Topics 126-150 (CLEF Jahrgang 2005).

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
126	author	3,5	0,181	2	0,01	0,181
127	author	2,97	0,044	2	<b>0,301</b>	0,044
128	author	3	0,029	2	<b>0,766</b>	0,029
129	author	3,5	0,087	2	0,07	0,087
130	author	3,36	0,032	2	<b>0,964</b>	0,032
131	author	3,5	0,136	2	0,03	0,136
132	author	3,5	0,191	2	<b>0,205</b>	0,191
133	author	3,5	0,034	2	<b>0,909</b>	0,034
134	author	2,56	0,099	2	<b>0,239</b>	0,099
135	author	3,13	0,022	2	<b>0,957</b>	0,022
136	author	3,5	0,083	2	<b>0,162</b>	0,083
137	author	3,5	0,075	2	<b>0,511</b>	0,075
138	author	3,5	0,026	3	<b>0,845</b>	0,026
139	author	3,14	0,051	2	<b>0,508</b>	0,051
140	author	3,5	0,072	2	<b>0,314</b>	0,072
141	author	3,5	0,072	2	<b>0,446</b>	0,072
142	author	3,07	0,02	2	<b>0,979</b>	0,02
143	author	3,4	0,061	2	<b>0,643</b>	0,061
144	author	3,1	0,058	2	<b>0,229</b>	0,058
145	author	2,79	0,044	2	<b>0,796</b>	0,044
146	author	3,5	0,092	2	<b>0,423</b>	0,092
147	author	3,5	0,044	2	<b>0,722</b>	0,044
148	author	3,5	0,113	2	0,092	0,113
149	author	3,32	0,052	2	<b>0,484</b>	0,052
150	author	2,88	0,022	2	<b>0,888</b>	0,022

Tabelle B.3: Einzelwerte der Produktivität der Autoren für die Topics 126-151 (CLEF Jahrgang 2005). Statistisch signifikante Werte sind **fett** gedruckt.

B Ermittelte Werte der Power-Law-Analyse

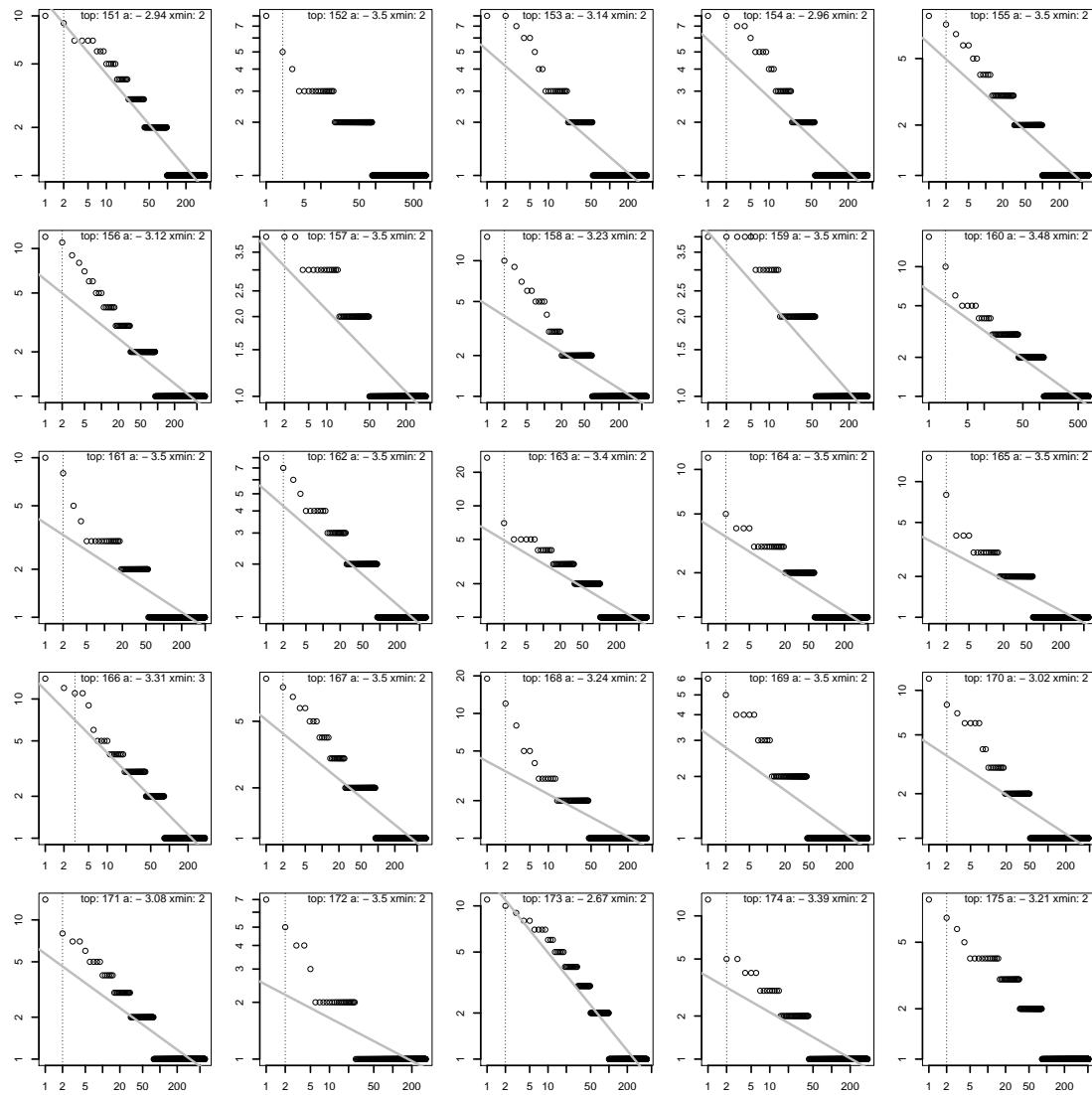


Abbildung B.4: Doppelt logarithmierter Plot der Produktivität der Autoren für die Topics 151-175 (CLEF Jahrgang 2006).



Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
151	author	2,94	0,03	2	<b>0,635</b>	0,03
152	author	3,5	0,105	2	0,01	0,105
153	author	3,14	0,028	2	<b>0,89</b>	0,028
154	author	2,96	0,036	2	<b>0,709</b>	0,036
155	author	3,5	0,014	2	<b>0,97</b>	0,014
156	author	3,12	0,012	2	<b>0,996</b>	0,012
157	author	3,5	0,073	2	<b>0,227</b>	0,073
158	author	3,23	0,059	2	<b>0,189</b>	0,059
159	author	3,5	0,072	2	<b>0,19</b>	0,072
160	author	3,48	0,039	2	<b>0,32</b>	0,039
161	author	3,5	0,06	2	<b>0,304</b>	0,06
162	author	3,5	0,027	2	<b>0,721</b>	0,027
163	author	3,4	0,03	2	<b>0,553</b>	0,03
164	author	3,5	0,053	2	<b>0,359</b>	0,053
165	author	3,5	0,056	2	<b>0,333</b>	0,056
166	author	3,31	0,06	3	<b>0,362</b>	0,06
167	author	3,5	0,022	2	<b>0,834</b>	0,022
168	author	3,24	0,052	2	<b>0,467</b>	0,052
169	author	3,5	0,063	2	<b>0,344</b>	0,063
170	author	3,02	0,064	2	<b>0,238</b>	0,064
171	author	3,08	0,022	2	<b>0,884</b>	0,022
172	author	3,5	0,103	2	<b>0,239</b>	0,103
173	author	2,67	0,05	2	<b>0,231</b>	0,05
174	author	3,39	0,026	2	<b>0,945</b>	0,026
175	author	3,21	0,069	2	<b>0,056</b>	0,069

Tabelle B.4: Einzelwerte der Produktivität der Autoren für die Topics 151-175 (CLEF Jahrgang 2006). Statistisch signifikante Werte sind **fett** gedruckt.

B Ermittelte Werte der Power-Law-Analyse

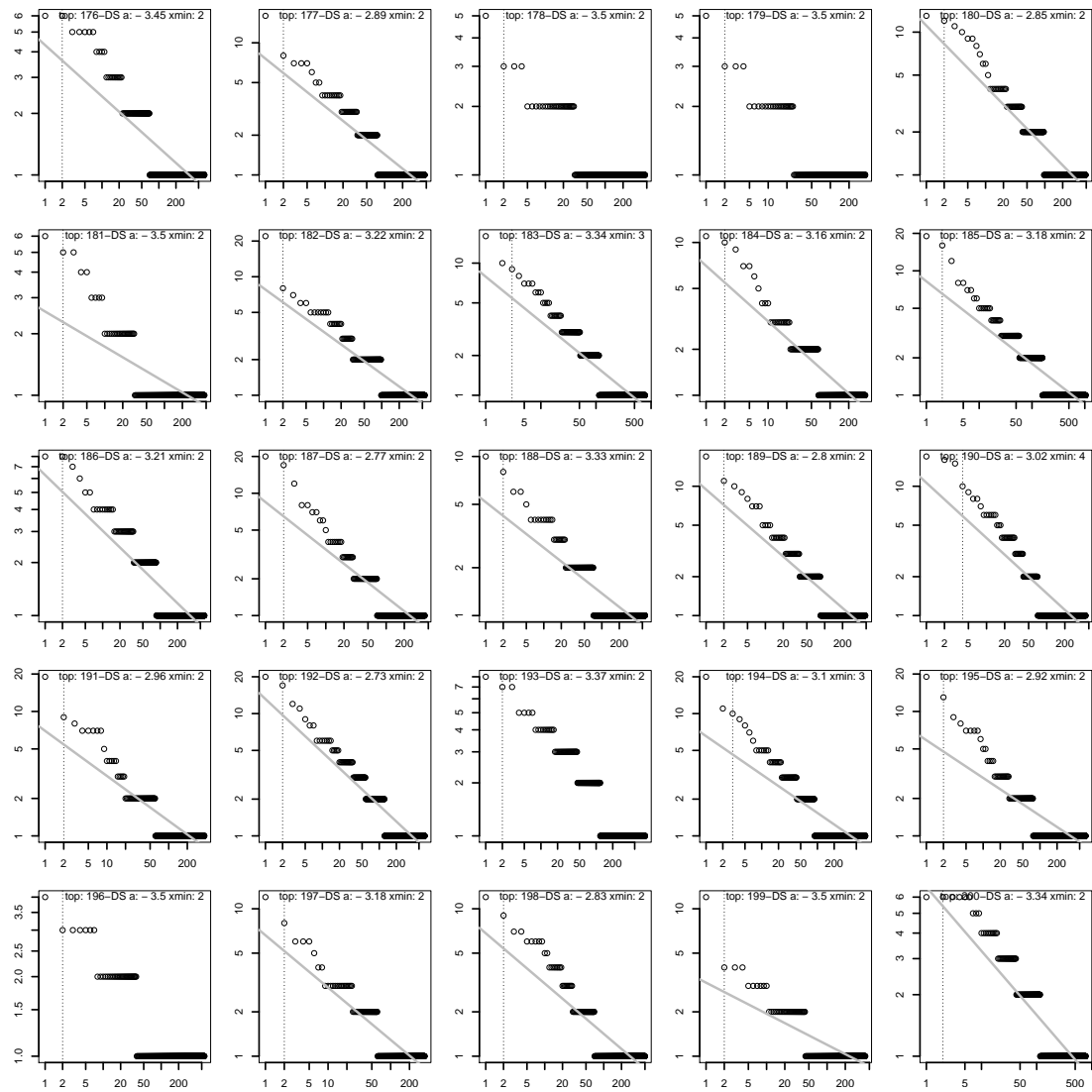


Abbildung B.5: Doppelt logarithmierter Plot der Produktivität der Autoren für die Topics 176-200 (CLEF Jahrgang 2007).

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
10.2452/176-DS	author	3,45	0,031	2	<b>0,671</b>	0,031
10.2452/177-DS	author	2,89	0,052	2	<b>0,268</b>	0,052
10.2452/178-DS	author	3,5	0,169	2	0,026	0,169
10.2452/179-DS	author	3,5	0,143	2	0,077	0,143
10.2452/180-DS	author	2,85	0,026	2	<b>0,75</b>	0,026
10.2452/181-DS	author	3,5	0,029	2	<b>0,967</b>	0,029
10.2452/182-DS	author	3,22	0,04	2	<b>0,303</b>	0,04
10.2452/183-DS	author	3,34	0,032	3	<b>0,761</b>	0,032
10.2452/184-DS	author	3,16	0,032	2	<b>0,702</b>	0,032
10.2452/185-DS	author	3,18	0,013	2	<b>0,934</b>	0,013
10.2452/186-DS	author	3,21	0,054	2	<b>0,148</b>	0,054
10.2452/187-DS	author	2,77	0,025	2	<b>0,916</b>	0,025
10.2452/188-DS	author	3,33	0,044	2	<b>0,398</b>	0,044
10.2452/189-DS	author	2,8	0,022	2	<b>0,911</b>	0,022
10.2452/190-DS	author	3,02	0,077	4	<b>0,307</b>	0,077
10.2452/191-DS	author	2,96	0,072	2	<b>0,102</b>	0,072
10.2452/192-DS	author	2,73	0,031	2	<b>0,553</b>	0,031
10.2452/193-DS	author	3,37	0,06	2	0,045	0,06
10.2452/194-DS	author	3,1	0,03	3	<b>0,826</b>	0,03
10.2452/195-DS	author	2,92	0,042	2	<b>0,476</b>	0,042
10.2452/196-DS	author	3,5	0,113	2	0,093	0,113
10.2452/197-DS	author	3,18	0,043	2	<b>0,471</b>	0,043
10.2452/198-DS	author	2,83	0,049	2	<b>0,355</b>	0,049
10.2452/199-DS	author	3,5	0,07	2	<b>0,318</b>	0,07
10.2452/200-DS	author	3,34	0,04	2	<b>0,234</b>	0,04

Tabelle B.5: Einzelwerte der Produktivität der Autoren für die Topics 176-200 (CLEF Jahrgang 2007). Statistisch signifikante Werte sind **fett** gedruckt.

B Ermittelte Werte der Power-Law-Analyse

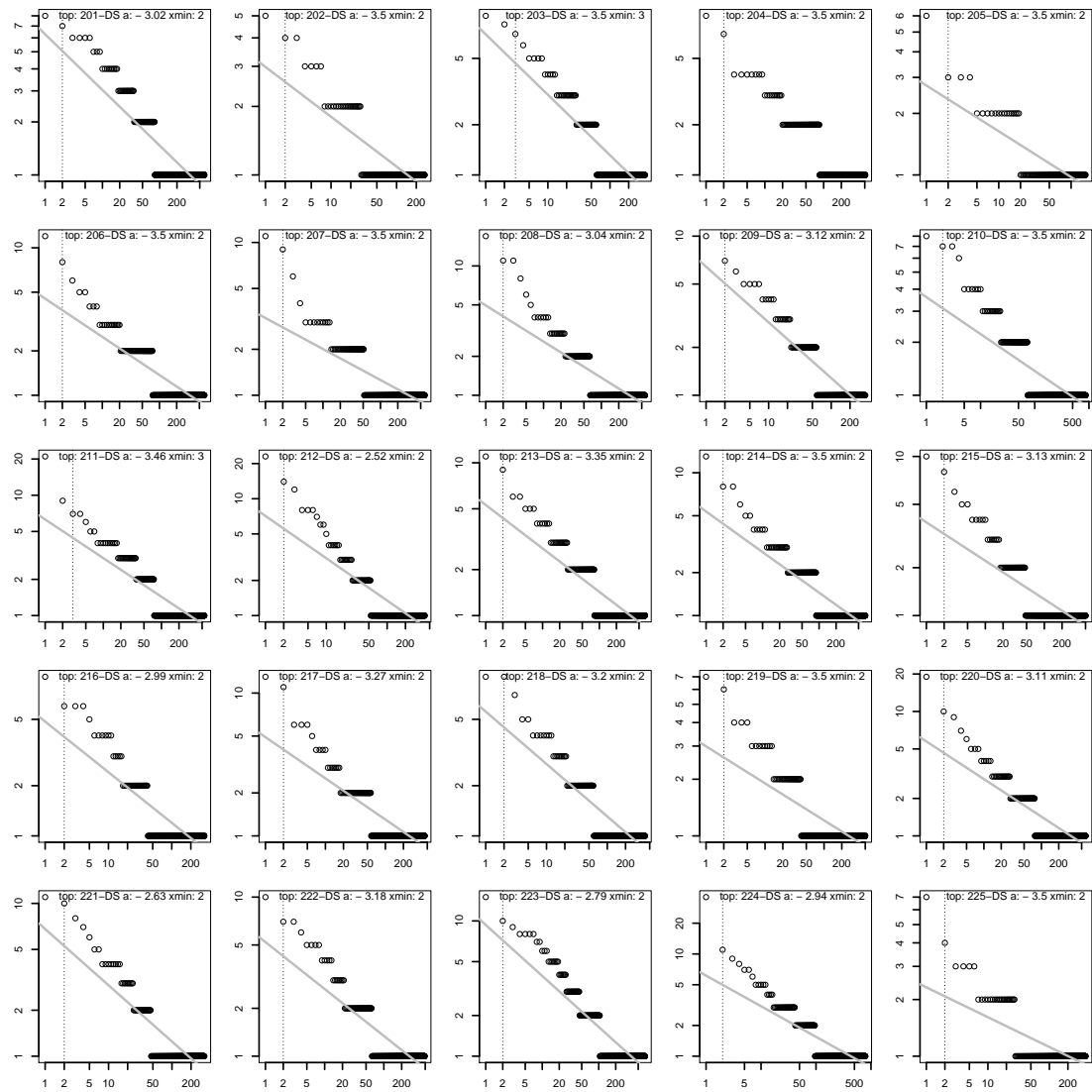


Abbildung B.6: Doppelt logarithmierter Plot der Produktivität der Autoren für die Topics 201-225 (CLEF Jahrgang 2008).

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
10.2452/201-DS	author	3,02	0,054	2	<b>0,188</b>	0,054
10.2452/202-DS	author	3,5	0,053	2	<b>0,709</b>	0,053
10.2452/203-DS	author	3,5	0,047	3	<b>0,671</b>	0,047
10.2452/204-DS	author	3,5	0,076	2	0,058	0,076
10.2452/205-DS	author	3,5	0,092	2	<b>0,423</b>	0,092
10.2452/206-DS	author	3,5	0,036	2	<b>0,591</b>	0,036
10.2452/207-DS	author	3,5	0,055	2	<b>0,422</b>	0,055
10.2452/208-DS	author	3,04	0,027	2	<b>0,83</b>	0,027
10.2452/209-DS	author	3,12	0,027	2	<b>0,843</b>	0,027
10.2452/210-DS	author	3,5	0,017	2	<b>0,968</b>	0,017
10.2452/211-DS	author	3,46	0,05	3	<b>0,617</b>	0,05
10.2452/212-DS	author	2,52	0,023	2	<b>0,988</b>	0,023
10.2452/213-DS	author	3,35	0,011	2	<b>0,999</b>	0,011
10.2452/214-DS	author	3,5	0,014	2	<b>0,982</b>	0,014
10.2452/215-DS	author	3,13	0,034	2	<b>0,797</b>	0,034
10.2452/216-DS	author	2,99	0,069	2	<b>0,286</b>	0,069
10.2452/217-DS	author	3,27	0,038	2	<b>0,619</b>	0,038
10.2452/218-DS	author	3,2	0,034	2	<b>0,694</b>	0,034
10.2452/219-DS	author	3,5	0,022	2	<b>0,981</b>	0,022
10.2452/220-DS	author	3,11	0,011	2	<b>0,998</b>	0,011
10.2452/221-DS	author	2,63	0,068	2	<b>0,329</b>	0,068
10.2452/222-DS	author	3,18	0,038	2	<b>0,57</b>	0,038
10.2452/223-DS	author	2,79	0,033	2	<b>0,556</b>	0,033
10.2452/224-DS	author	2,94	0,039	2	<b>0,373</b>	0,039
10.2452/225-DS	author	3,5	0,08	2	<b>0,372</b>	0,08

Tabelle B.6: Einzelwerte der Produktivität der Autoren für die Topics 201-225 (CLEF Jahrgang 2008). Statistisch signifikante Werte sind **fett** gedruckt.

## B.2 Produktivität der Klassifikationen in GIRT

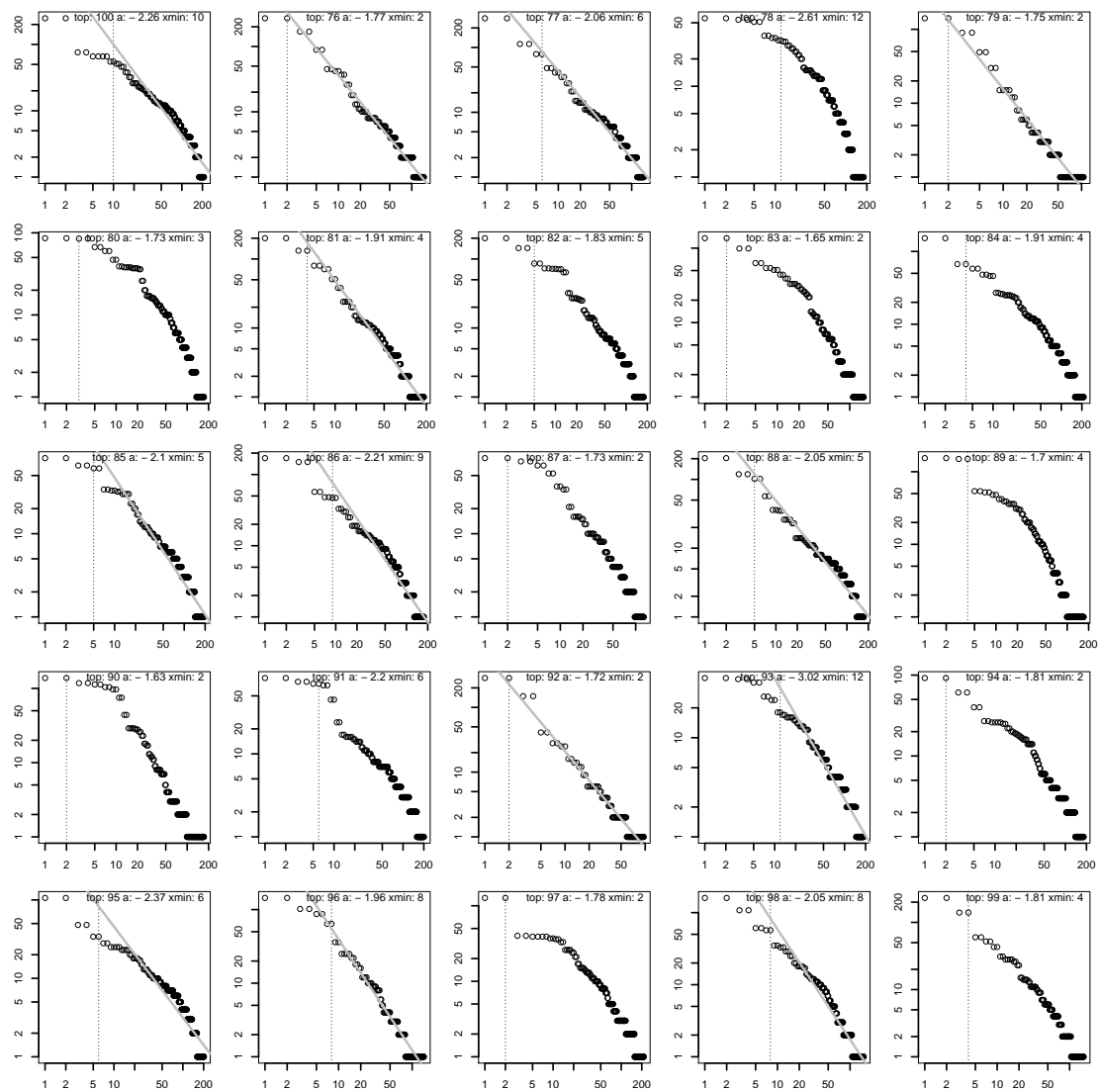


Abbildung B.7: Doppelt logarithmierter Plot der Produktivität der Klassifikationen für die Topics 76-100 (CLEF Jahrgang 2003).

B.2 Produktivität der Klassifikationen in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
76	classification	1,77	0,048	2	<b>0,581</b>	0,048
77	classification	2,06	0,079	6	<b>0,203</b>	0,079
78	classification	2,61	0,118	12	0,044	0,118
79	classification	1,75	0,047	2	<b>0,881</b>	0,047
80	classification	1,73	0,111	3	0	0,111
81	classification	1,91	0,052	4	<b>0,677</b>	0,052
82	classification	1,83	0,074	5	0,065	0,074
83	classification	1,65	0,09	2	0,011	0,09
84	classification	1,91	0,092	4	0,006	0,092
85	classification	2,1	0,073	5	<b>0,141</b>	0,073
86	classification	2,21	0,064	9	<b>0,599</b>	0,064
87	classification	1,73	0,098	2	0,006	0,098
88	classification	2,05	0,069	5	<b>0,169</b>	0,069
89	classification	1,7	0,123	4	0,001	0,123
90	classification	1,63	0,086	2	0,023	0,086
91	classification	2,2	0,09	6	0,008	0,09
92	classification	1,72	0,048	2	<b>0,875</b>	0,048
93	classification	3,02	0,097	12	<b>0,158</b>	0,097
94	classification	1,81	0,074	2	0,009	0,074
95	classification	2,37	0,056	6	<b>0,474</b>	0,056
96	classification	1,96	0,093	8	<b>0,238</b>	0,093
97	classification	1,78	0,08	2	0,004	0,08
98	classification	2,05	0,062	8	<b>0,759</b>	0,062
99	classification	1,81	0,082	4	0,093	0,082
100	classification	2,26	0,061	10	<b>0,495</b>	0,061

Tabelle B.7: Einzelwerte der Produktivität der Klassifikationen für die Topics 76-100 (CLEF Jahrgang 2003). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

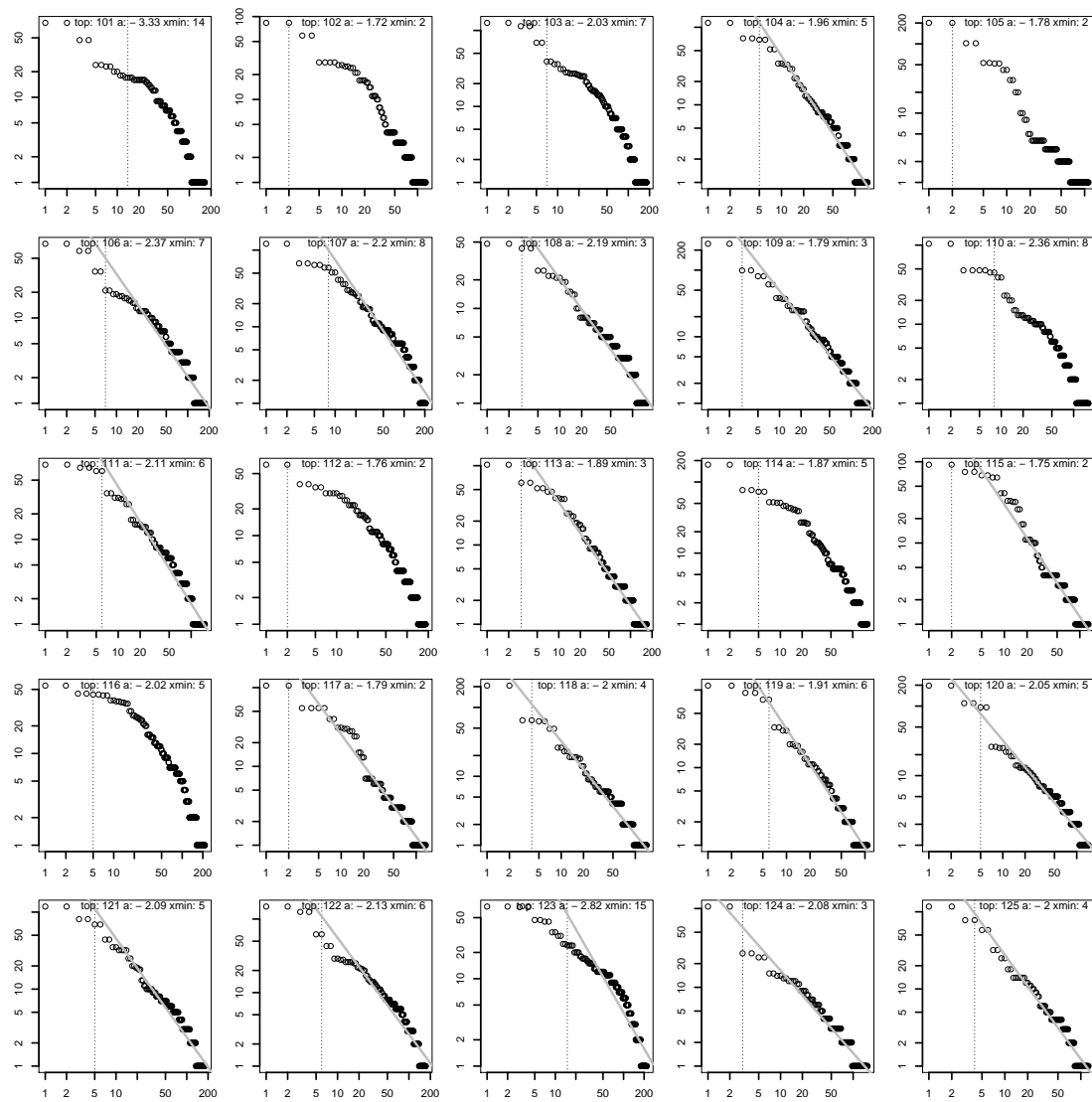


Abbildung B.8: Doppelt logarithmierter Plot der Produktivität der Klassifikationen für die Topics 101-125 (CLEF Jahrgang 2004).



Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
101	classification	3,33	0,132	14	0,03	0,132
102	classification	1,72	0,087	2	0,042	0,087
103	classification	2,03	0,116	7	0,004	0,116
104	classification	1,96	0,071	5	<b>0,401</b>	0,071
105	classification	1,78	0,121	2	0,006	0,121
106	classification	2,37	0,083	7	<b>0,189</b>	0,083
107	classification	2,2	0,082	8	<b>0,147</b>	0,082
108	classification	2,19	0,053	3	<b>0,425</b>	0,053
109	classification	1,79	0,07	3	<b>0,114</b>	0,07
110	classification	2,36	0,109	8	0,034	0,109
111	classification	2,11	0,077	6	<b>0,254</b>	0,077
112	classification	1,76	0,101	2	0	0,101
113	classification	1,89	0,063	3	<b>0,226</b>	0,063
114	classification	1,87	0,094	5	0,019	0,094
115	classification	1,75	0,075	2	<b>0,125</b>	0,075
116	classification	2,02	0,081	5	0,042	0,081
117	classification	1,79	0,072	2	<b>0,147</b>	0,072
118	classification	2	0,071	4	<b>0,37</b>	0,071
119	classification	1,91	0,073	6	<b>0,511</b>	0,073
120	classification	2,05	0,067	5	<b>0,363</b>	0,067
121	classification	2,09	0,064	5	<b>0,344</b>	0,064
122	classification	2,13	0,068	6	<b>0,323</b>	0,068
123	classification	2,82	0,092	15	<b>0,205</b>	0,092
124	classification	2,08	0,068	3	<b>0,202</b>	0,068
125	classification	2	0,066	4	<b>0,472</b>	0,066

Tabelle B.8: Einzelwerte der Produktivität der Klassifikationen für die Topics 101-125 (CLEF Jahrgang 2004). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

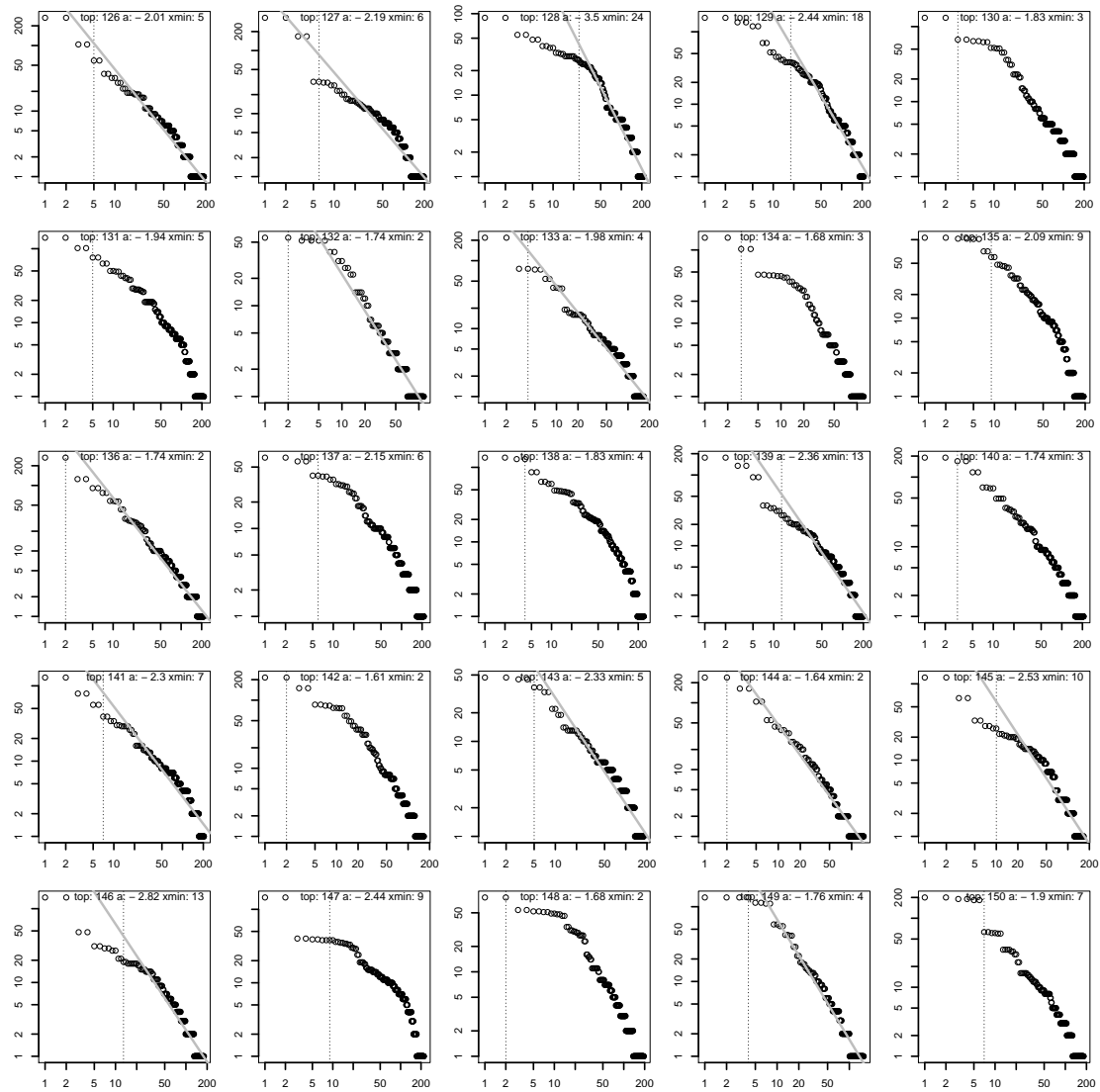


Abbildung B.9: Doppelt logarithmierter Plot der Produktivität der Klassifikationen für die Topics 126-150 (CLEF Jahrgang 2005).

## B.2 Produktivität der Klassifikationen in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
126	classification	2,01	0,084	5	<b>0,107</b>	0,084
127	classification	2,19	0,066	6	<b>0,349</b>	0,066
128	classification	3,5	0,1	24	<b>0,261</b>	0,1
129	classification	2,44	0,088	18	<b>0,257</b>	0,088
130	classification	1,83	0,07	3	0,034	0,07
131	classification	1,94	0,084	5	0,02	0,084
132	classification	1,74	0,07	2	<b>0,274</b>	0,07
133	classification	1,98	0,05	4	<b>0,653</b>	0,05
134	classification	1,68	0,097	3	0,019	0,097
135	classification	2,09	0,088	9	0,053	0,088
136	classification	1,74	0,055	2	<b>0,106</b>	0,055
137	classification	2,15	0,099	6	0,009	0,099
138	classification	1,83	0,093	4	0,001	0,093
139	classification	2,36	0,084	13	<b>0,195</b>	0,084
140	classification	1,74	0,075	3	0,017	0,075
141	classification	2,3	0,074	7	<b>0,177</b>	0,074
142	classification	1,61	0,101	2	0,001	0,101
143	classification	2,33	0,052	5	<b>0,603</b>	0,052
144	classification	1,64	0,075	2	<b>0,113</b>	0,075
145	classification	2,53	0,082	10	<b>0,302</b>	0,082
146	classification	2,82	0,093	13	<b>0,238</b>	0,093
147	classification	2,44	0,093	9	0,01	0,093
148	classification	1,68	0,08	2	0,005	0,08
149	classification	1,76	0,077	4	<b>0,191</b>	0,077
150	classification	1,9	0,088	7	0,043	0,088

Tabelle B.9: Einzelwerte der Produktivität der Klassifikationen für die Topics 126-150 (CLEF Jahrgang 2005). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

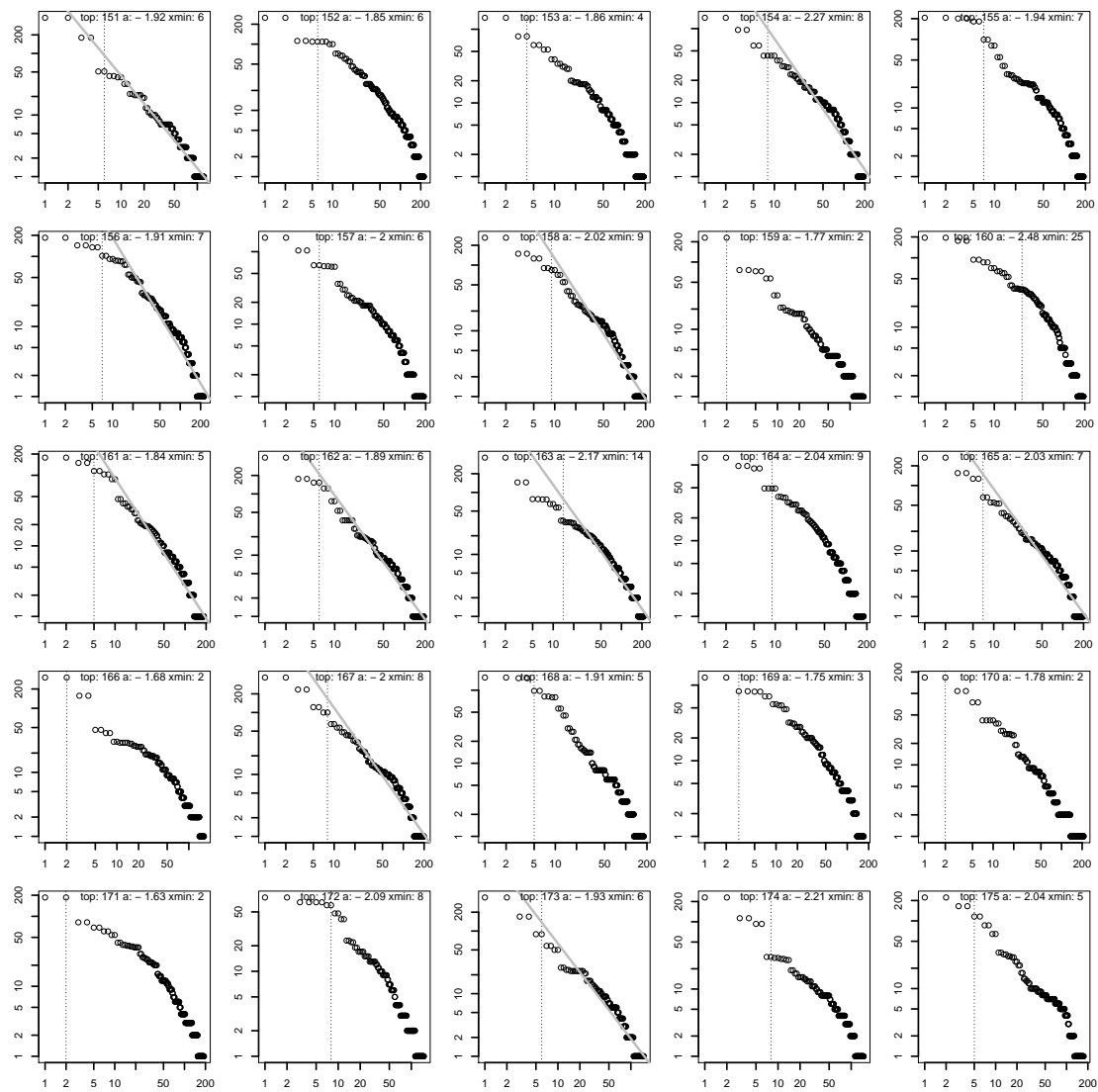


Abbildung B.10: Doppelt logarithmierter Plot der Produktivität der Klassifikationen für die Topics 151-175 (CLEF Jahrgang 2006).

B.2 Produktivität der Klassifikationen in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
151	classification	1,92	0,088	6	<b>0,149</b>	0,088
152	classification	1,85	0,079	6	0,012	0,079
153	classification	1,86	0,088	4	0,014	0,088
154	classification	2,27	0,054	8	<b>0,68</b>	0,054
155	classification	1,94	0,093	7	0,013	0,093
156	classification	1,91	0,07	7	<b>0,105</b>	0,07
157	classification	2	0,085	6	0,088	0,085
158	classification	2,02	0,082	9	<b>0,168</b>	0,082
159	classification	1,77	0,088	2	0,016	0,088
160	classification	2,48	0,104	25	0,099	0,104
161	classification	1,84	0,067	5	<b>0,164</b>	0,067
162	classification	1,89	0,063	6	<b>0,38</b>	0,063
163	classification	2,17	0,083	14	<b>0,116</b>	0,083
164	classification	2,04	0,101	9	0,018	0,101
165	classification	2,03	0,05	7	<b>0,66</b>	0,05
166	classification	1,68	0,11	2	0	0,11
167	classification	2	0,07	8	<b>0,276</b>	0,07
168	classification	1,91	0,088	5	0,03	0,088
169	classification	1,75	0,087	3	0,004	0,087
170	classification	1,78	0,077	2	0,018	0,077
171	classification	1,63	0,099	2	0	0,099
172	classification	2,09	0,093	8	0,079	0,093
173	classification	1,93	0,08	6	<b>0,19</b>	0,08
174	classification	2,21	0,1	8	0,049	0,1
175	classification	2,04	0,097	5	0,013	0,097

Tabelle B.10: Einzelwerte der Produktivität der Klassifikationen für die Topics 151-175 (CLEF Jahrgang 2006). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

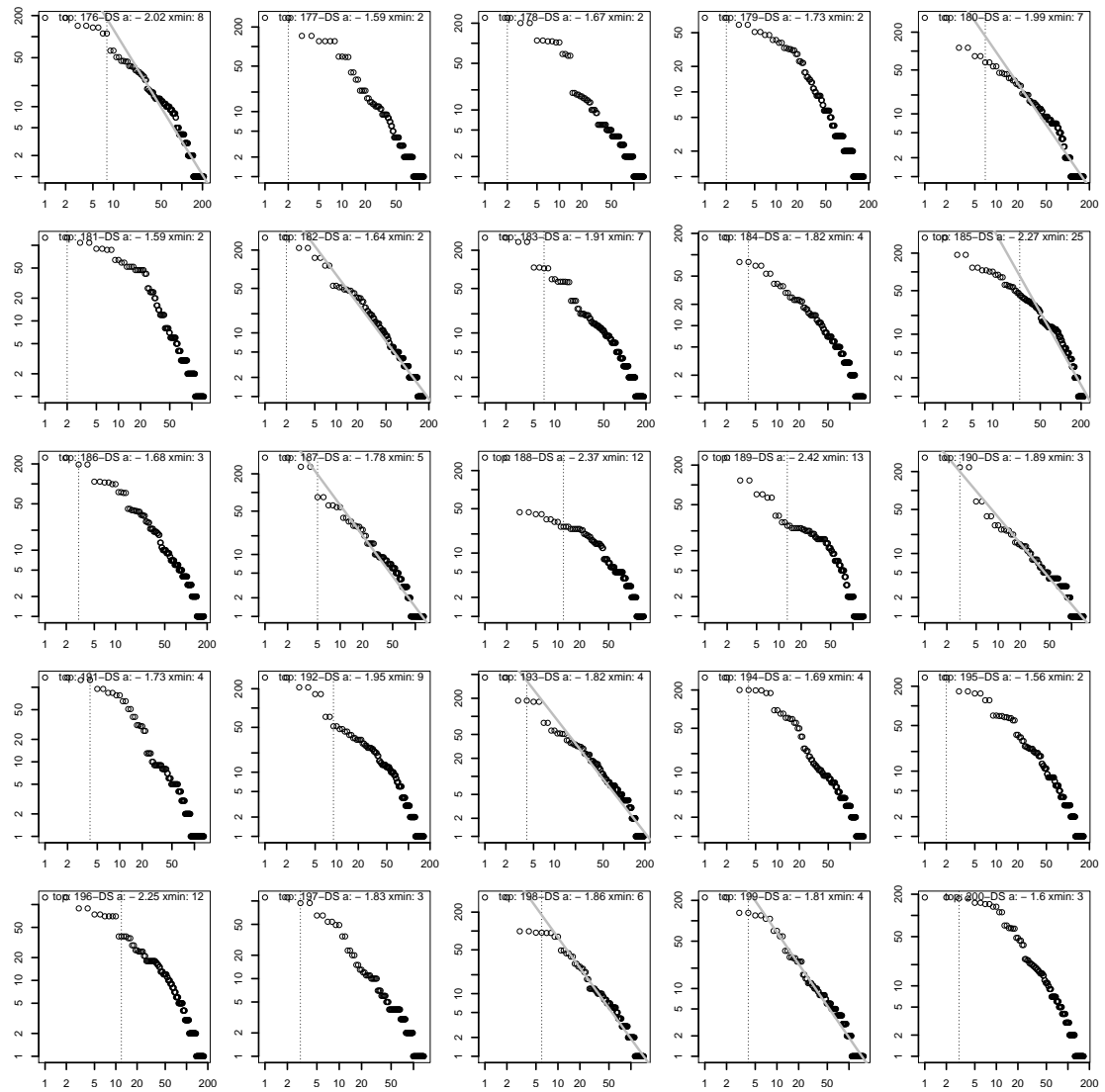


Abbildung B.11: Doppelt logarithmierter Plot der Produktivität der Klassifikationen für die Topics 176-200 (CLEF Jahrgang 2007).

B.2 Produktivität der Klassifikationen in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
10.2452/176-DS	classification	2,02	0,074	8	<b>0,207</b>	0,074
10.2452/177-DS	classification	1,59	0,099	2	0,028	0,099
10.2452/178-DS	classification	1,67	0,078	2	0,084	0,078
10.2452/179-DS	classification	1,73	0,072	2	0,04	0,072
10.2452/180-DS	classification	1,99	0,074	7	<b>0,276</b>	0,074
10.2452/181-DS	classification	1,59	0,076	2	0,049	0,076
10.2452/182-DS	classification	1,64	0,048	2	<b>0,361</b>	0,048
10.2452/183-DS	classification	1,91	0,089	7	0,077	0,089
10.2452/184-DS	classification	1,82	0,087	4	0,019	0,087
10.2452/185-DS	classification	2,27	0,074	25	<b>0,549</b>	0,074
10.2452/186-DS	classification	1,68	0,071	3	0,064	0,071
10.2452/187-DS	classification	1,78	0,066	5	<b>0,441</b>	0,066
10.2452/188-DS	classification	2,37	0,125	12	0,014	0,125
10.2452/189-DS	classification	2,42	0,13	13	0,018	0,13
10.2452/190-DS	classification	1,89	0,039	3	<b>0,89</b>	0,039
10.2452/191-DS	classification	1,73	0,098	4	0,033	0,098
10.2452/192-DS	classification	1,95	0,091	9	0,088	0,091
10.2452/193-DS	classification	1,82	0,066	4	<b>0,141</b>	0,066
10.2452/194-DS	classification	1,69	0,071	4	0,095	0,071
10.2452/195-DS	classification	1,56	0,099	2	0,001	0,099
10.2452/196-DS	classification	2,25	0,1	12	0,055	0,1
10.2452/197-DS	classification	1,83	0,086	3	0,05	0,086
10.2452/198-DS	classification	1,86	0,078	6	<b>0,197</b>	0,078
10.2452/199-DS	classification	1,81	0,067	4	<b>0,251</b>	0,067
10.2452/200-DS	classification	1,6	0,087	3	0,007	0,087

Tabelle B.11: Einzelwerte der Produktivität der Klassifikationen für die Topics 176-200 (CLEF Jahrgang 2007). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

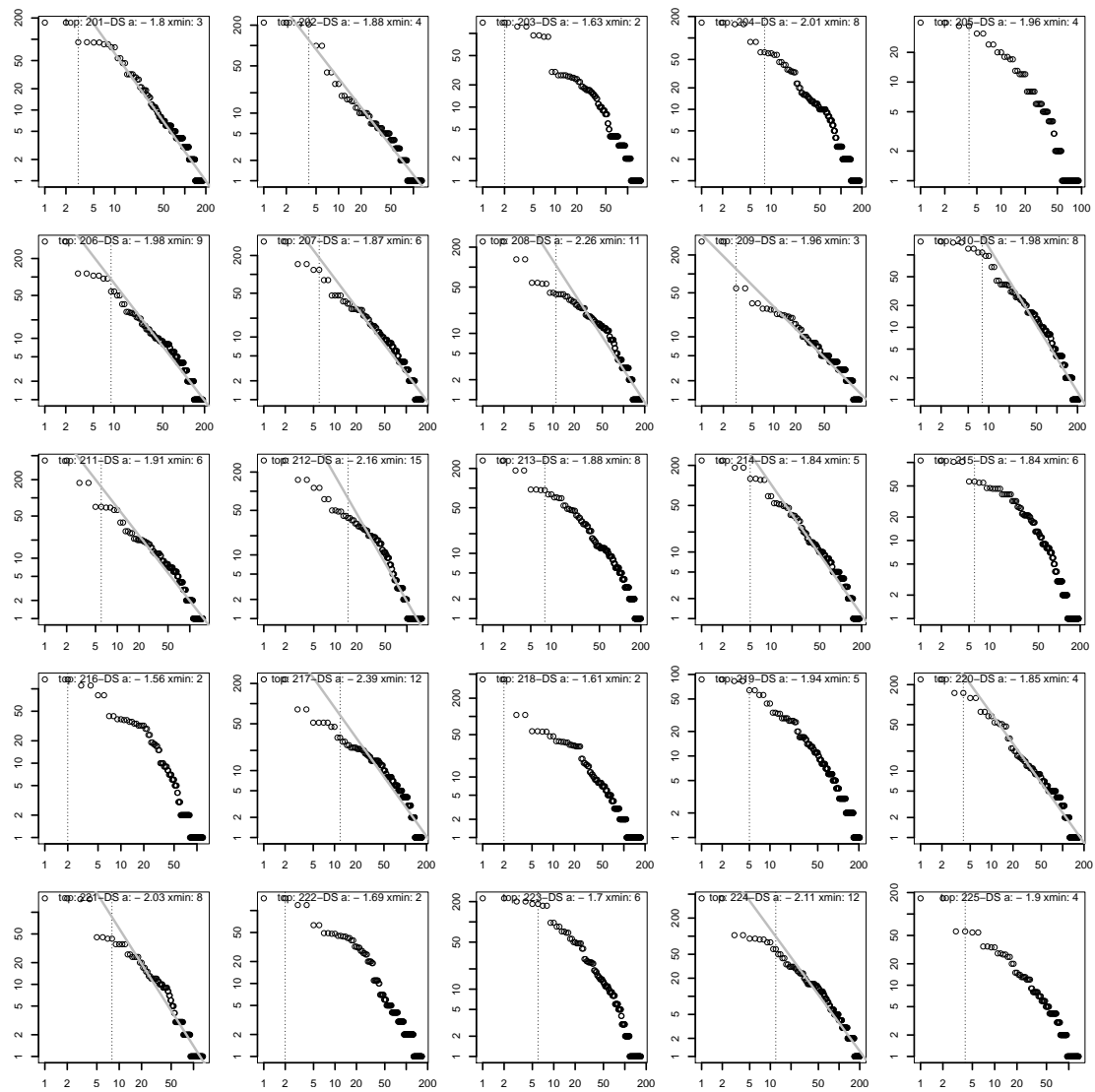


Abbildung B.12: Doppelt logarithmierter Plot der Produktivität der Klassifikationen für die Topics 201-225 (CLEF Jahrgang 2008).



## B.2 Produktivität der Klassifikationen in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
10.2452/201-DS	classification	1,8	0,057	3	<b>0,244</b>	0,057
10.2452/202-DS	classification	1,88	0,061	4	<b>0,589</b>	0,061
10.2452/203-DS	classification	1,63	0,104	2	0,002	0,104
10.2452/204-DS	classification	2,01	0,096	8	0,037	0,096
10.2452/205-DS	classification	1,96	0,134	4	0,015	0,134
10.2452/206-DS	classification	1,98	0,065	9	<b>0,492</b>	0,065
10.2452/207-DS	classification	1,87	0,074	6	<b>0,154</b>	0,074
10.2452/208-DS	classification	2,26	0,089	11	<b>0,165</b>	0,089
10.2452/209-DS	classification	1,96	0,047	3	<b>0,615</b>	0,047
10.2452/210-DS	classification	1,98	0,075	8	<b>0,173</b>	0,075
10.2452/211-DS	classification	1,91	0,064	6	<b>0,513</b>	0,064
10.2452/212-DS	classification	2,16	0,065	15	<b>0,821</b>	0,065
10.2452/213-DS	classification	1,88	0,084	8	0,021	0,084
10.2452/214-DS	classification	1,84	0,058	5	<b>0,286</b>	0,058
10.2452/215-DS	classification	1,84	0,138	6	0	0,138
10.2452/216-DS	classification	1,56	0,135	2	0,003	0,135
10.2452/217-DS	classification	2,39	0,075	12	<b>0,457</b>	0,075
10.2452/218-DS	classification	1,61	0,092	2	0,014	0,092
10.2452/219-DS	classification	1,94	0,091	5	0,003	0,091
10.2452/220-DS	classification	1,85	0,059	4	<b>0,392</b>	0,059
10.2452/221-DS	classification	2,03	0,082	8	<b>0,41</b>	0,082
10.2452/222-DS	classification	1,69	0,09	2	0,003	0,09
10.2452/223-DS	classification	1,7	0,095	6	0,012	0,095
10.2452/224-DS	classification	2,11	0,084	12	<b>0,126</b>	0,084
10.2452/225-DS	classification	1,9	0,097	4	0,05	0,097

Tabelle B.12: Einzelwerte der Produktivität der Klassifikationen für die Topics 201-225 (CLEF Jahrgang 2008). Statistisch signifikante Werte sind **fett** gedruckt.

### B.3 Produktivität der Zeitschriften in GIRT

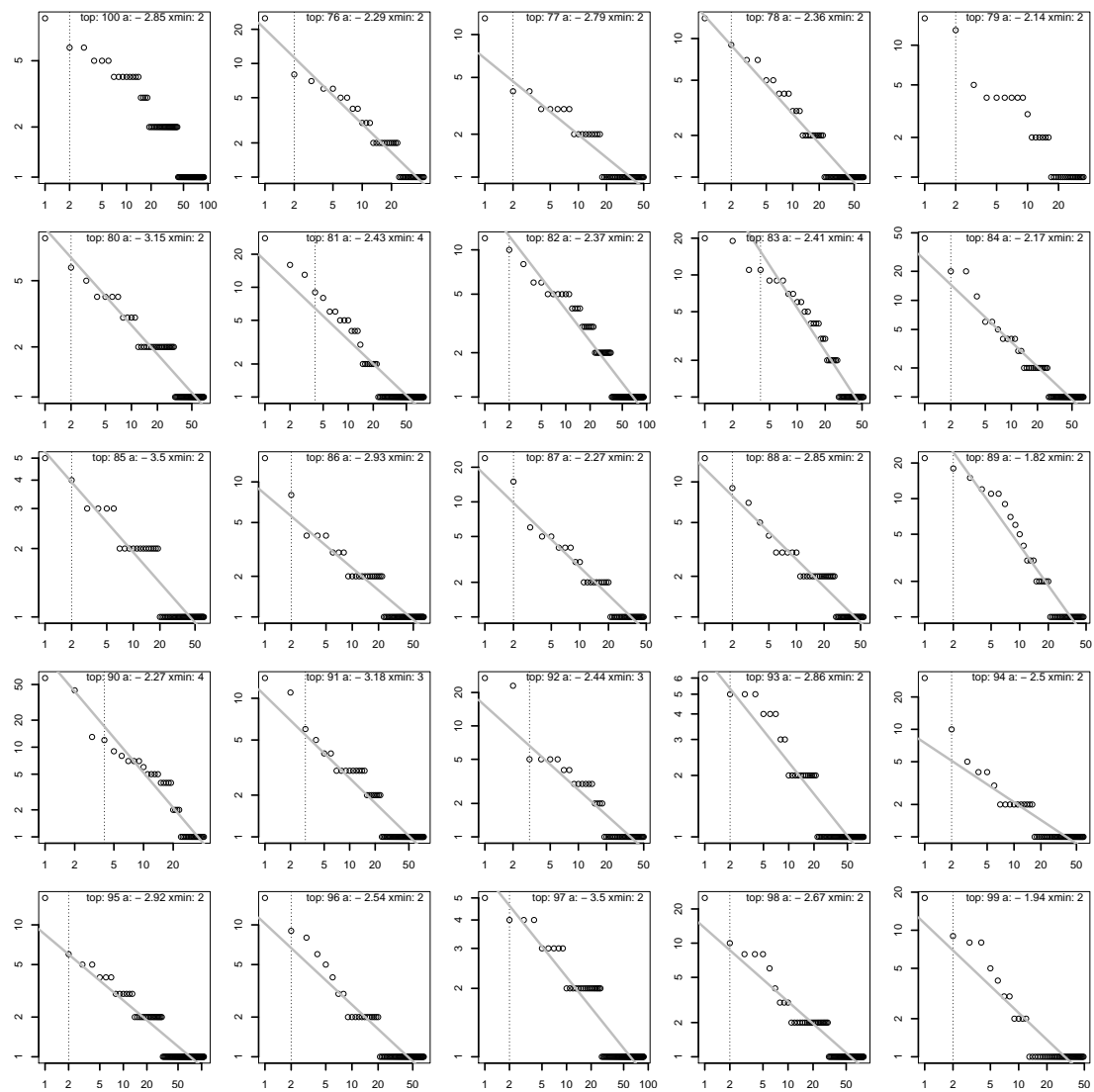


Abbildung B.13: Doppelt logarithmierter Plot der Produktivität der Zeitschriften für die Topics 76-100 (CLEF Jahrgang 2003).

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
76	issn	2,291	0,067	2	<b>0,779</b>	0,067
77	issn	2,788	0,075	2	<b>0,663</b>	0,075
78	issn	2,36	0,08	2	<b>0,648</b>	0,08
79	issn	2,144	0,194	2	0,055	0,194
80	issn	3,15	0,05	2	<b>0,678</b>	0,05
81	issn	2,43	0,068	4	<b>0,973</b>	0,068
82	issn	2,37	0,098	2	<b>0,165</b>	0,098
83	issn	2,41	0,123	4	<b>0,268</b>	0,123
84	issn	2,17	0,064	2	<b>0,819</b>	0,064
85	issn	3,5	0,044	2	<b>0,934</b>	0,044
86	issn	2,93	0,042	2	<b>0,94</b>	0,042
87	issn	2,273	0,065	2	<b>0,872</b>	0,065
88	issn	2,85	0,044	2	<b>0,903</b>	0,044
89	issn	1,817	0,119	2	<b>0,418</b>	0,119
90	issn	2,269	0,07	4	<b>0,955</b>	0,07
91	issn	3,18	0,09	3	<b>0,579</b>	0,09
92	issn	2,439	0,158	3	<b>0,125</b>	0,158
93	issn	2,86	0,112	2	<b>0,209</b>	0,112
94	issn	2,5	0,107	2	<b>0,452</b>	0,107
95	issn	2,92	0,033	2	<b>0,945</b>	0,033
96	issn	2,54	0,073	2	<b>0,703</b>	0,073
97	issn	3,5	0,044	2	<b>0,879</b>	0,044
98	issn	2,67	0,112	2	<b>0,1</b>	0,112
99	issn	1,935	0,133	2	<b>0,577</b>	0,133
100	issn	2,85	0,11	2	0,051	0,11

Tabelle B.13: Einzelwerte der Produktivität der Zeitschriften für die Topics 76-100 (CLEF Jahrgang 2003). Statistisch signifikante Werte sind **fett** gedruckt.

B Ermittelte Werte der Power-Law-Analyse

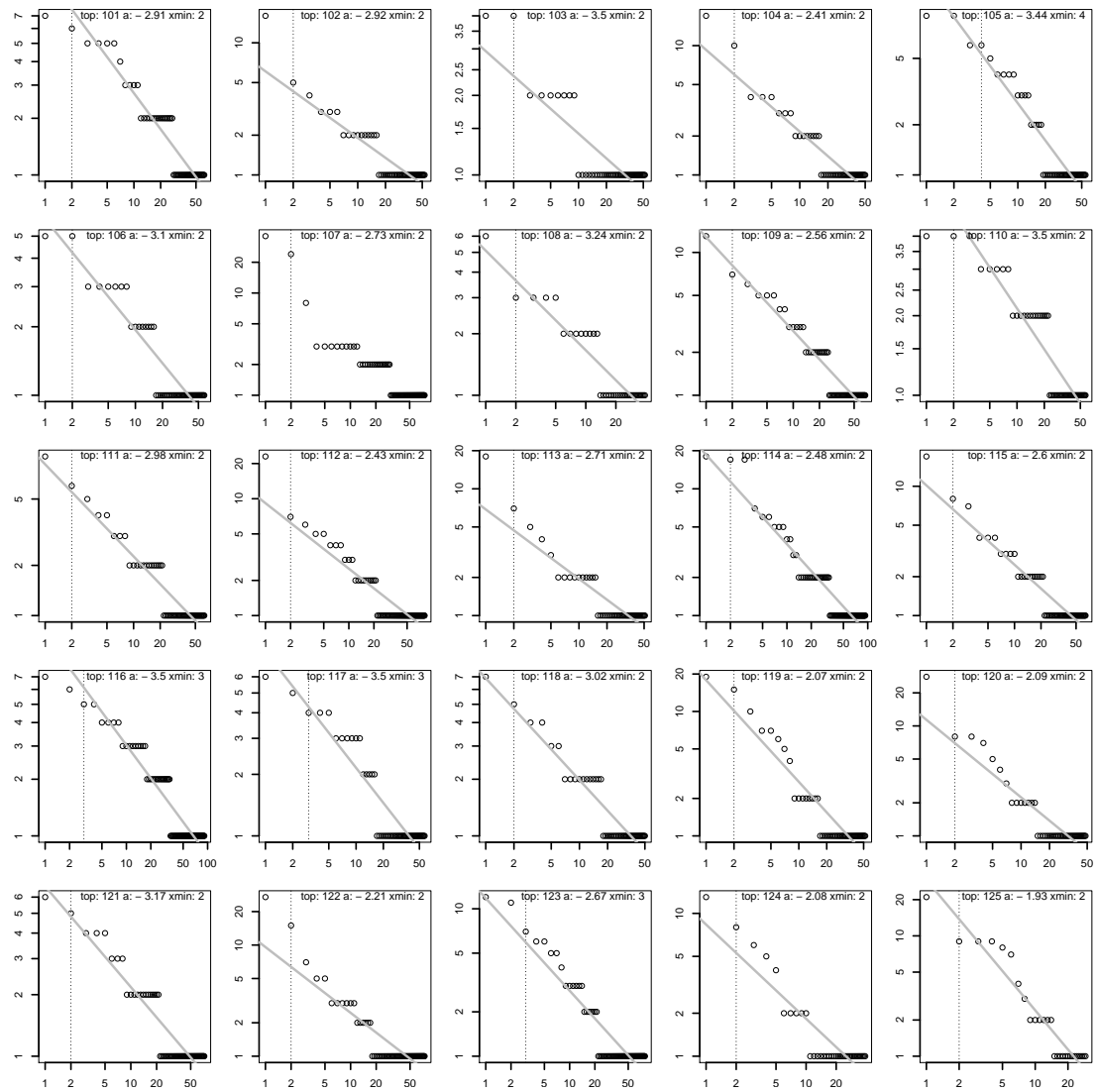


Abbildung B.14: Doppelt logarithmierter Plot der Produktivität der Zeitschriften für die Topics 101-125 (CLEF Jahrgang 2004).

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
101	issn	2,91	0,09	2	<b>0,284</b>	0,09
102	issn	2,92	0,05	2	<b>0,935</b>	0,05
103	issn	3,5	0,088	2	<b>0,785</b>	0,088
104	issn	2,406	0,078	2	<b>0,787</b>	0,078
105	issn	3,44	0,107	4	<b>0,602</b>	0,107
106	issn	3,1	0,136	2	<b>0,164</b>	0,136
107	issn	2,73	0,142	2	0,048	0,142
108	issn	3,243	0,072	2	<b>0,837</b>	0,072
109	issn	2,56	0,051	2	<b>0,863</b>	0,051
110	issn	3,5	0,072	2	<b>0,609</b>	0,072
111	issn	2,98	0,037	2	<b>0,979</b>	0,037
112	issn	2,43	0,07	2	<b>0,768</b>	0,07
113	issn	2,71	0,103	2	<b>0,488</b>	0,103
114	issn	2,48	0,093	2	<b>0,244</b>	0,093
115	issn	2,6	0,041	2	<b>0,979</b>	0,041
116	issn	3,5	0,067	3	<b>0,628</b>	0,067
117	issn	3,5	0,096	3	<b>0,77</b>	0,096
118	issn	3,024	0,059	2	<b>0,837</b>	0,059
119	issn	2,07	0,119	2	<b>0,477</b>	0,119
120	issn	2,086	0,099	2	<b>0,761</b>	0,099
121	issn	3,17	0,065	2	<b>0,704</b>	0,065
122	issn	2,21	0,093	2	<b>0,68</b>	0,093
123	issn	2,67	0,118	3	<b>0,277</b>	0,118
124	issn	2,08	0,125	2	<b>0,715</b>	0,125
125	issn	1,929	0,19	2	<b>0,108</b>	0,19

Tabelle B.14: Einzelwerte der Produktivität der Zeitschriften für die Topics 101-125 (CLEF Jahrgang 2004). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

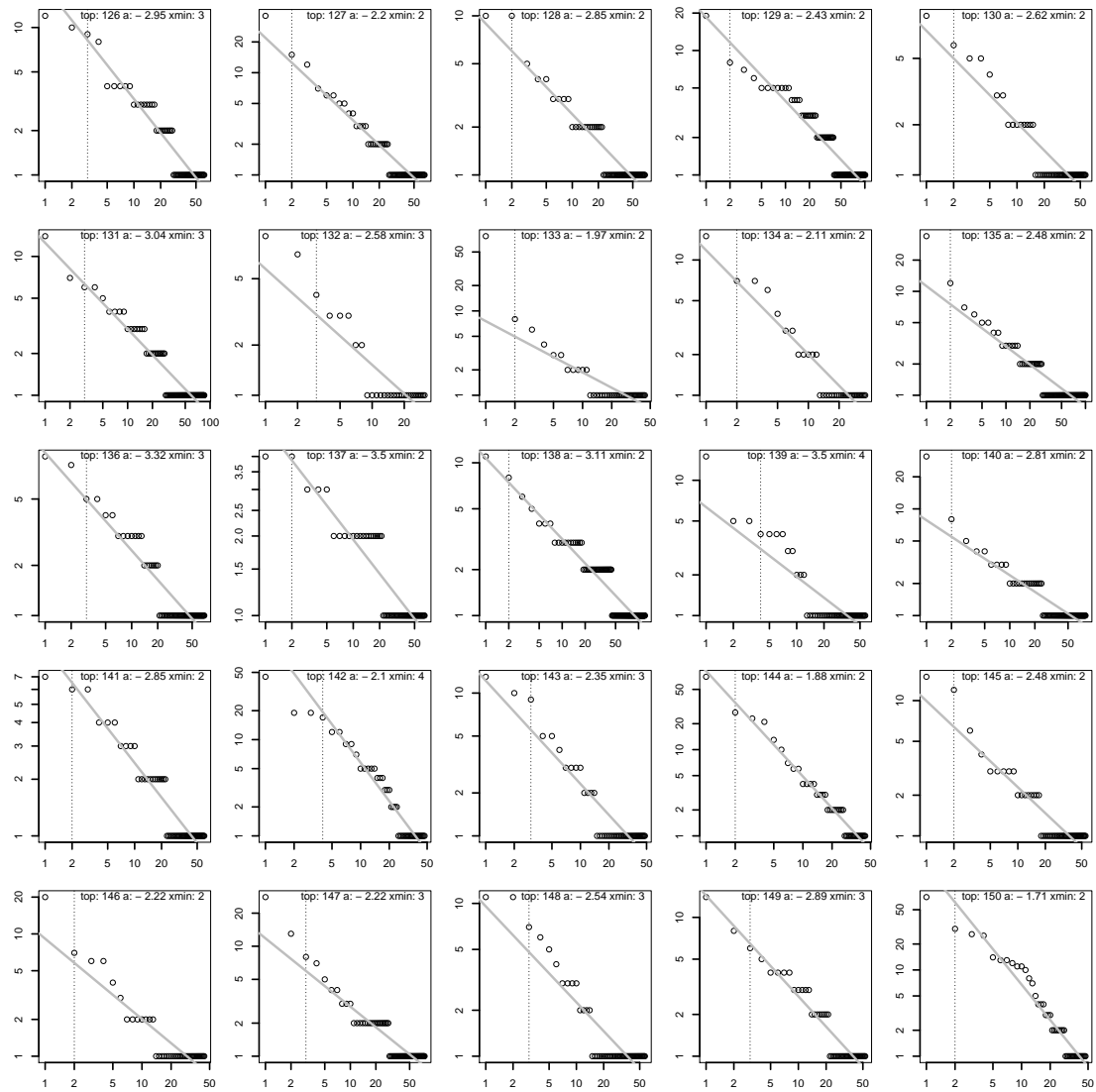


Abbildung B.15: Doppelt logarithmierter Plot der Produktivität der Zeitschriften für die Topics 126-150 (CLEF Jahrgang 2005).

B.3 Produktivität der Zeitschriften in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
126	issn	2,95	0,114	3	<b>0,199</b>	0,114
127	issn	2,2	0,042	2	<b>0,989</b>	0,042
128	issn	2,85	0,055	2	<b>0,837</b>	0,055
129	issn	2,43	0,101	2	<b>0,121</b>	0,101
130	issn	2,62	0,088	2	<b>0,687</b>	0,088
131	issn	3,04	0,05	3	<b>0,917</b>	0,05
132	issn	2,583	0,166	3	<b>0,537</b>	0,166
133	issn	1,973	0,075	2	<b>0,973</b>	0,075
134	issn	2,109	0,117	2	<b>0,641</b>	0,117
135	issn	2,48	0,026	2	<b>1</b>	0,026
136	issn	3,32	0,072	3	<b>0,858</b>	0,072
137	issn	3,5	0,072	2	<b>0,609</b>	0,072
138	issn	3,11	0,047	2	<b>0,566</b>	0,047
139	issn	3,5	0,186	4	<b>0,28</b>	0,186
140	issn	2,81	0,042	2	<b>0,955</b>	0,042
141	issn	2,85	0,055	2	<b>0,803</b>	0,055
142	issn	2,101	0,114	4	<b>0,41</b>	0,114
143	issn	2,35	0,137	3	<b>0,386</b>	0,137
144	issn	1,875	0,059	2	<b>0,945</b>	0,059
145	issn	2,48	0,066	2	<b>0,904</b>	0,066
146	issn	2,218	0,119	2	<b>0,582</b>	0,119
147	issn	2,22	0,084	3	<b>0,704</b>	0,084
148	issn	2,54	0,098	3	<b>0,814</b>	0,098
149	issn	2,89	0,077	3	<b>0,773</b>	0,077
150	issn	1,714	0,132	2	<b>0,15</b>	0,132

Tabelle B.15: Einzelwerte der Produktivität der Zeitschriften für die Topics 126-150 (CLEF Jahrgang 2005). Statistisch signifikante Werte sind **fett** gedruckt.

B Ermittelte Werte der Power-Law-Analyse

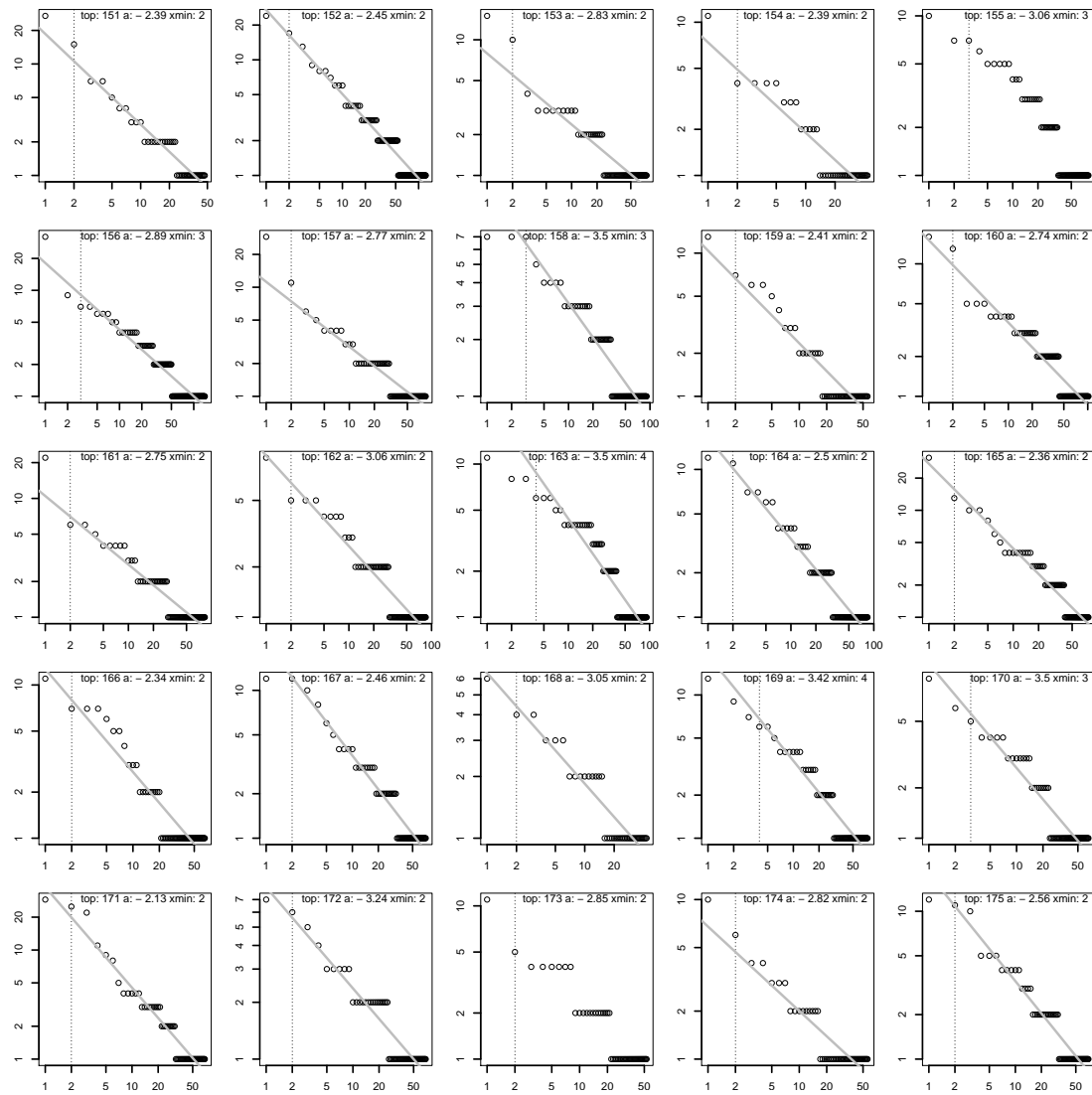


Abbildung B.16: Doppelt logarithmierter Plot der Produktivität der Zeitschriften für die Topics 151-175 (CLEF Jahrgang 2006).



Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
151	issn	2,387	0,059	2	<b>0,852</b>	0,059
152	issn	2,45	0,034	2	<b>0,896</b>	0,034
153	issn	2,83	0,096	2	<b>0,279</b>	0,096
154	issn	2,394	0,135	2	<b>0,329</b>	0,135
155	issn	3,06	0,123	3	0,099	0,123
156	issn	2,89	0,069	3	<b>0,372</b>	0,069
157	issn	2,77	0,058	2	<b>0,631</b>	0,058
158	issn	3,5	0,071	3	<b>0,614</b>	0,071
159	issn	2,41	0,071	2	<b>0,863</b>	0,071
160	issn	2,74	0,081	2	<b>0,173</b>	0,081
161	issn	2,75	0,069	2	<b>0,497</b>	0,069
162	issn	3,06	0,078	2	<b>0,306</b>	0,078
163	issn	3,5	0,118	4	<b>0,237</b>	0,118
164	issn	2,5	0,06	2	<b>0,675</b>	0,06
165	issn	2,36	0,071	2	<b>0,409</b>	0,071
166	issn	2,34	0,11	2	<b>0,343</b>	0,11
167	issn	2,46	0,071	2	<b>0,512</b>	0,071
168	issn	3,053	0,052	2	<b>0,925</b>	0,052
169	issn	3,42	0,076	4	<b>0,707</b>	0,076
170	issn	3,5	0,058	3	<b>0,937</b>	0,058
171	issn	2,13	0,102	2	<b>0,249</b>	0,102
172	issn	3,24	0,03	2	<b>0,977</b>	0,03
173	issn	2,85	0,158	2	0,049	0,158
174	issn	2,82	0,031	2	<b>0,999</b>	0,031
175	issn	2,56	0,074	2	<b>0,459</b>	0,074

Tabelle B.16: Einzelwerte der Produktivität der Zeitschriften für die Topics 151-175 (CLEF Jahrgang 2006). Statistisch signifikante Werte sind **fett** gedruckt.

B Ermittelte Werte der Power-Law-Analyse

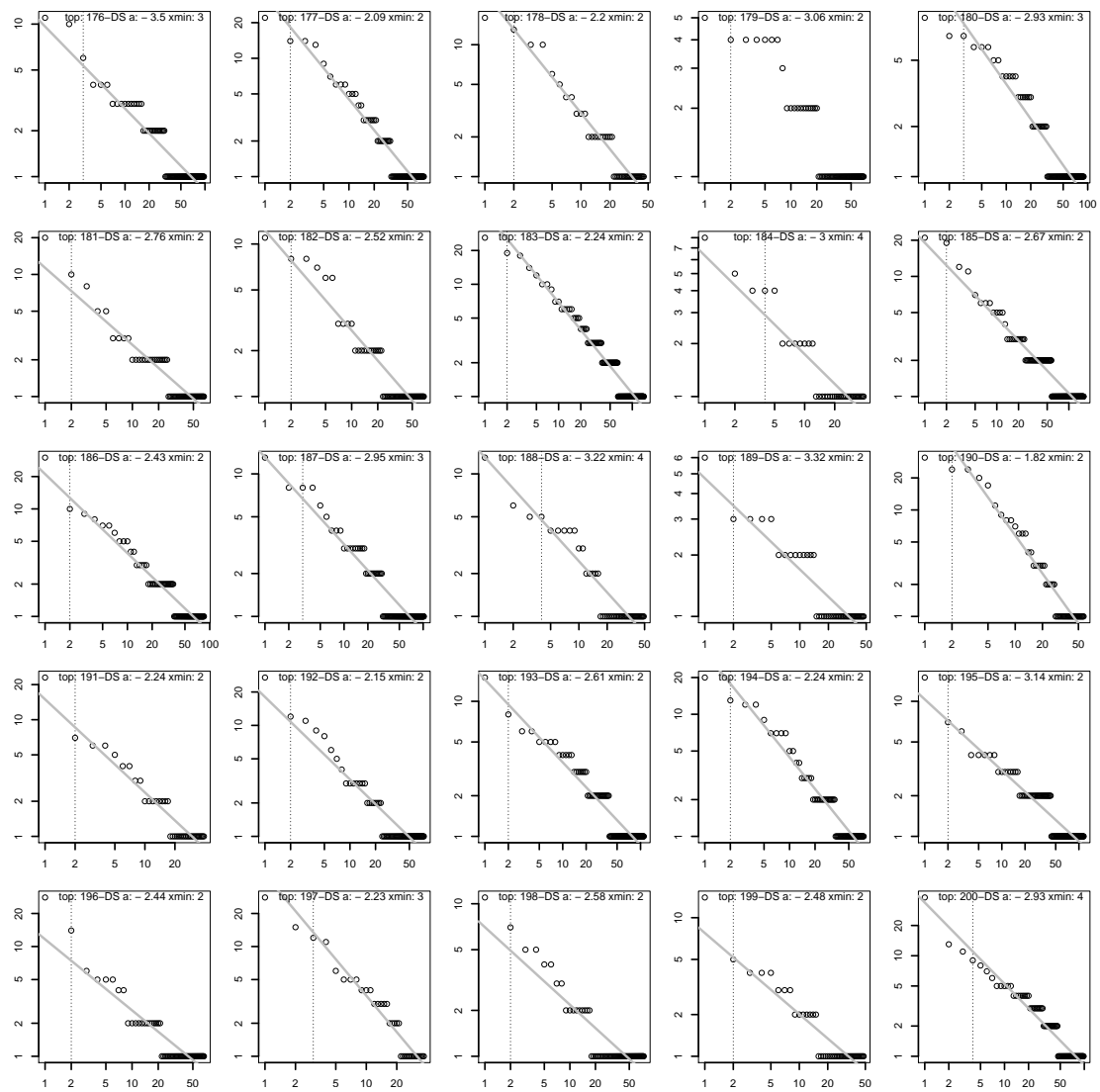


Abbildung B.17: Doppelt logarithmierter Plot der Produktivität der Zeitschriften für die Topics 176-200 (CLEF Jahrgang 2007).

B.3 Produktivität der Zeitschriften in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
10.2452/176-DS	issn	3,5	0,088	3	<b>0,41</b>	0,088
10.2452/177-DS	issn	2,09	0,091	2	<b>0,404</b>	0,091
10.2452/178-DS	issn	2,2	0,088	2	<b>0,603</b>	0,088
10.2452/179-DS	issn	3,06	0,161	2	0,044	0,161
10.2452/180-DS	issn	2,93	0,119	3	<b>0,146</b>	0,119
10.2452/181-DS	issn	2,76	0,067	2	<b>0,598</b>	0,067
10.2452/182-DS	issn	2,52	0,126	2	<b>0,164</b>	0,126
10.2452/183-DS	issn	2,24	0,057	2	<b>0,455</b>	0,057
10.2452/184-DS	issn	3	0,139	4	<b>0,628</b>	0,139
10.2452/185-DS	issn	2,67	0,041	2	<b>0,716</b>	0,041
10.2452/186-DS	issn	2,43	0,059	2	<b>0,646</b>	0,059
10.2452/187-DS	issn	2,95	0,101	3	<b>0,267</b>	0,101
10.2452/188-DS	issn	3,222	0,106	4	<b>0,78</b>	0,106
10.2452/189-DS	issn	3,321	0,062	2	<b>0,871</b>	0,062
10.2452/190-DS	issn	1,82	0,129	2	<b>0,156</b>	0,129
10.2452/191-DS	issn	2,242	0,072	2	<b>0,87</b>	0,072
10.2452/192-DS	issn	2,15	0,082	2	<b>0,669</b>	0,082
10.2452/193-DS	issn	2,61	0,064	2	<b>0,49</b>	0,064
10.2452/194-DS	issn	2,24	0,095	2	<b>0,234</b>	0,095
10.2452/195-DS	issn	3,14	0,035	2	<b>0,804</b>	0,035
10.2452/196-DS	issn	2,44	0,115	2	<b>0,247</b>	0,115
10.2452/197-DS	issn	2,228	0,095	3	<b>0,656</b>	0,095
10.2452/198-DS	issn	2,58	0,077	2	<b>0,725</b>	0,077
10.2452/199-DS	issn	2,475	0,07	2	<b>0,879</b>	0,07
10.2452/200-DS	issn	2,93	0,073	4	<b>0,544</b>	0,073

Tabelle B.17: Einzelwerte der Produktivität der Zeitschriften für die Topics 176-200 (CLEF Jahrgang 2007). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

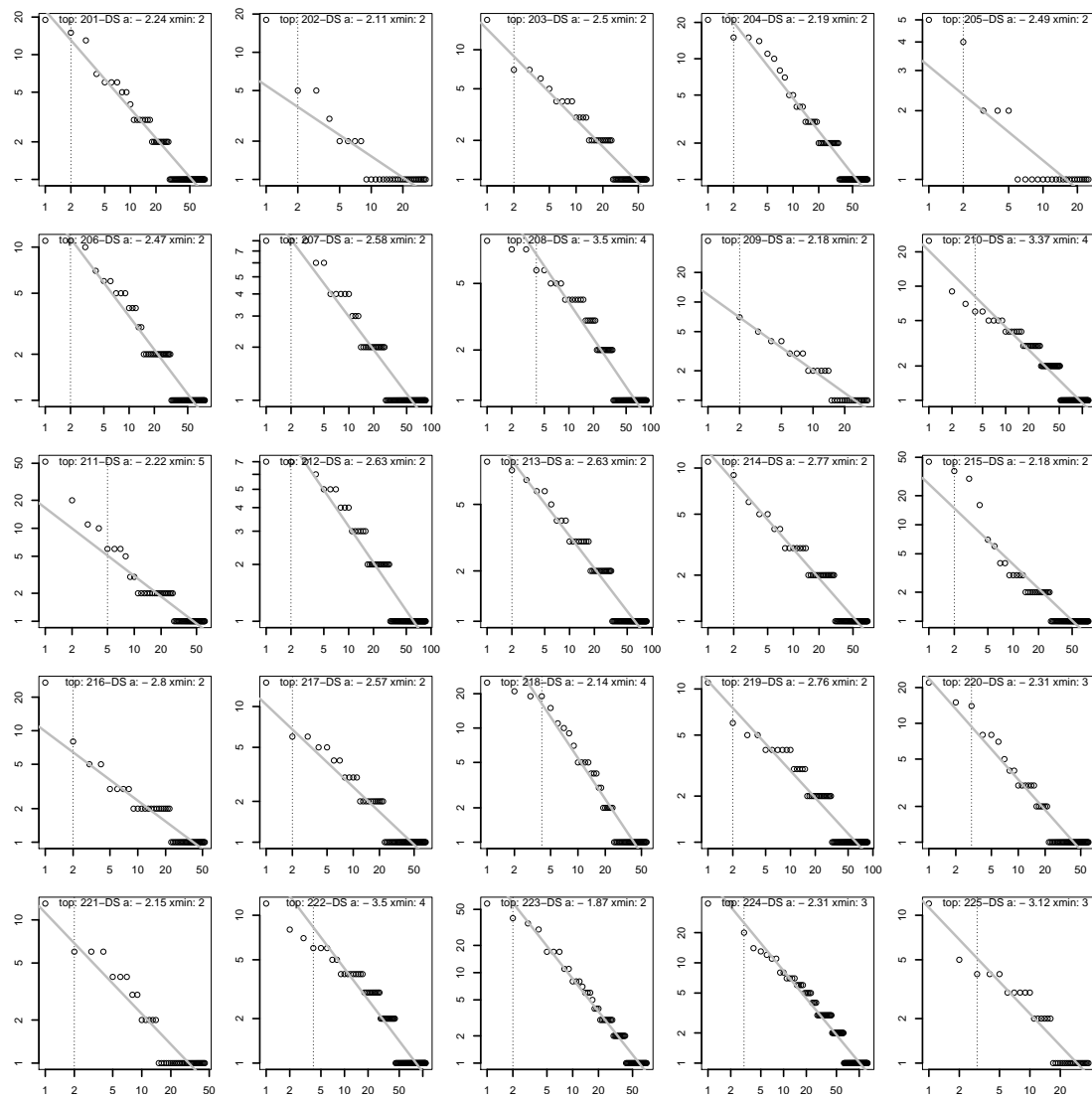


Abbildung B.18: Doppelt logarithmierter Plot der Produktivität der Zeitschriften für die Topics 201-225 (CLEF Jahrgang 2008).

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
10.2452/201-DS	issn	2,24	0,061	2	<b>0,797</b>	0,061
10.2452/202-DS	issn	2,111	0,117	2	<b>0,855</b>	0,117
10.2452/203-DS	issn	2,5	0,066	2	<b>0,708</b>	0,066
10.2452/204-DS	issn	2,19	0,061	2	<b>0,723</b>	0,061
10.2452/205-DS	issn	2,488	0,179	2	<b>0,607</b>	0,179
10.2452/206-DS	issn	2,47	0,085	2	<b>0,338</b>	0,085
10.2452/207-DS	issn	2,58	0,094	2	<b>0,281</b>	0,094
10.2452/208-DS	issn	3,5	0,084	4	<b>0,645</b>	0,084
10.2452/209-DS	issn	2,179	0,064	2	<b>0,966</b>	0,064
10.2452/210-DS	issn	3,37	0,052	4	<b>0,724</b>	0,052
10.2452/211-DS	issn	2,22	0,14	5	<b>0,583</b>	0,14
10.2452/212-DS	issn	2,63	0,078	2	<b>0,388</b>	0,078
10.2452/213-DS	issn	2,63	0,078	2	<b>0,38</b>	0,078
10.2452/214-DS	issn	2,77	0,042	2	<b>0,882</b>	0,042
10.2452/215-DS	issn	2,18	0,084	2	<b>0,561</b>	0,084
10.2452/216-DS	issn	2,8	0,055	2	<b>0,834</b>	0,055
10.2452/217-DS	issn	2,57	0,061	2	<b>0,834</b>	0,061
10.2452/218-DS	issn	2,14	0,134	4	<b>0,231</b>	0,134
10.2452/219-DS	issn	2,76	0,083	2	<b>0,298</b>	0,083
10.2452/220-DS	issn	2,31	0,109	3	<b>0,407</b>	0,109
10.2452/221-DS	issn	2,151	0,137	2	<b>0,37</b>	0,137
10.2452/222-DS	issn	3,5	0,069	4	<b>0,638</b>	0,069
10.2452/223-DS	issn	1,87	0,059	2	<b>0,784</b>	0,059
10.2452/224-DS	issn	2,31	0,066	3	<b>0,515</b>	0,066
10.2452/225-DS	issn	3,124	0,071	3	<b>0,91</b>	0,071

Tabelle B.18: Einzelwerte der Produktivität der Zeitschriften für die Topics 201-225 (CLEF Jahrgang 2008). Statistisch signifikante Werte sind **fett** gedruckt.

## B.4 Produktivität des Erscheinunglandes in GIRT

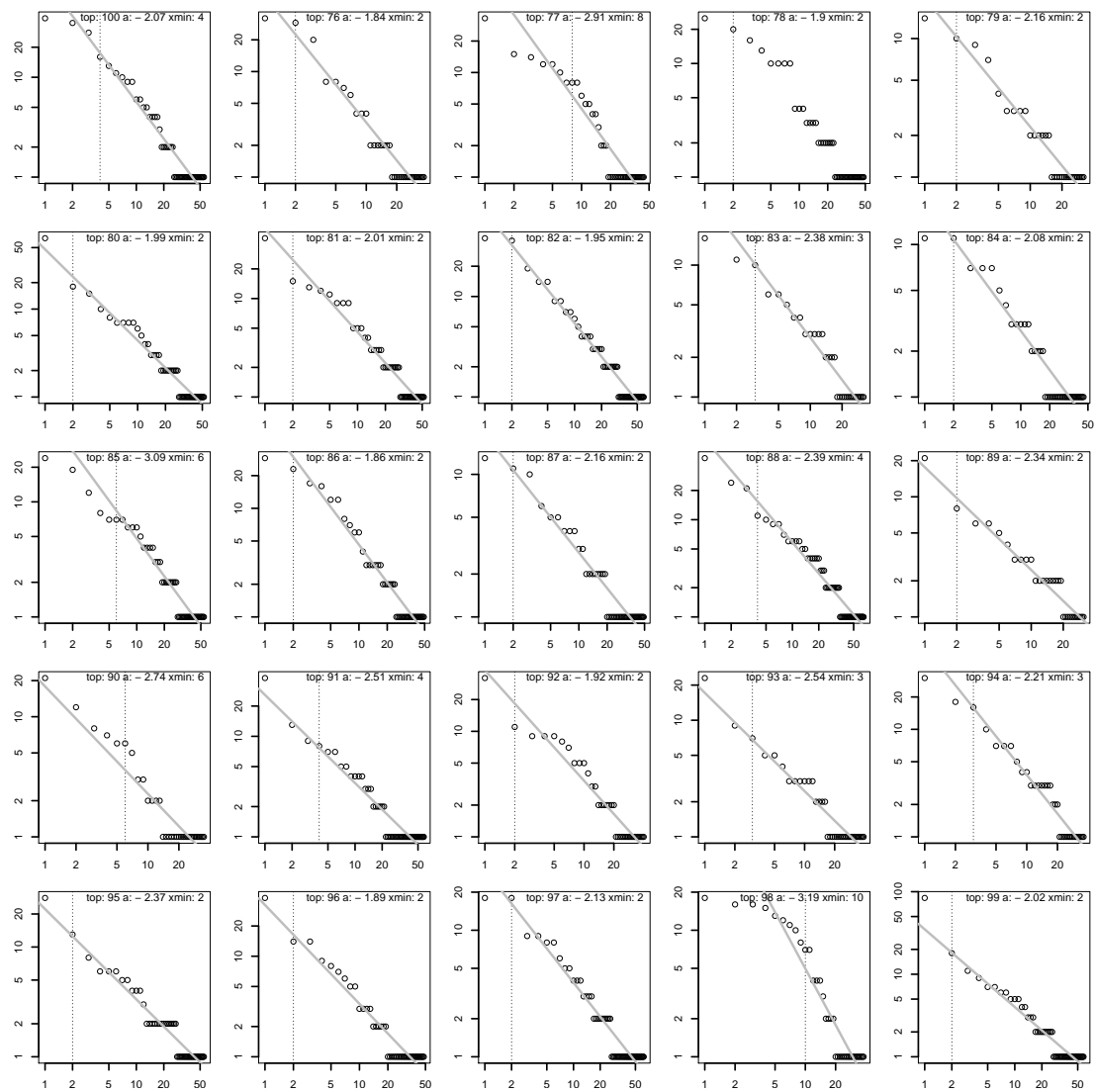


Abbildung B.19: Doppelt logarithmierter Plot der Produktivität des Erscheinunglandes für die Topics 76-100 (CLEF Jahrgang 2003).

B.4 Produktivität des Erscheinungslandes in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
76	location	1,838	0,107	2	<b>0,65</b>	0,107
77	location	2,911	0,156	8	<b>0,446</b>	0,156
78	location	1,899	0,166	2	0,062	0,166
79	location	2,157	0,097	2	<b>0,72</b>	0,097
80	location	1,99	0,092	2	<b>0,495</b>	0,092
81	location	2,01	0,107	2	<b>0,294</b>	0,107
82	location	1,95	0,047	2	<b>0,974</b>	0,047
83	location	2,375	0,091	3	<b>0,807</b>	0,091
84	location	2,082	0,136	2	<b>0,291</b>	0,136
85	location	3,09	0,125	6	<b>0,325</b>	0,125
86	location	1,862	0,098	2	<b>0,514</b>	0,098
87	location	2,165	0,108	2	<b>0,408</b>	0,108
88	location	2,39	0,063	4	<b>0,916</b>	0,063
89	location	2,336	0,042	2	<b>0,991</b>	0,042
90	location	2,742	0,118	6	<b>0,884</b>	0,118
91	location	2,51	0,104	4	<b>0,583</b>	0,104
92	location	1,922	0,146	2	<b>0,174</b>	0,146
93	location	2,54	0,076	3	<b>0,934</b>	0,076
94	location	2,205	0,113	3	<b>0,47</b>	0,113
95	location	2,37	0,081	2	<b>0,491</b>	0,081
96	location	1,891	0,108	2	<b>0,553</b>	0,108
97	location	2,13	0,065	2	<b>0,815</b>	0,065
98	location	3,188	0,19	10	<b>0,253</b>	0,19
99	location	2,02	0,087	2	<b>0,573</b>	0,087
100	location	2,07	0,14	4	<b>0,169</b>	0,14

Tabelle B.19: Einzelwerte der Produktivität des Erscheinungslandes für die Topics 76-100 (CLEF Jahrgang 2003). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

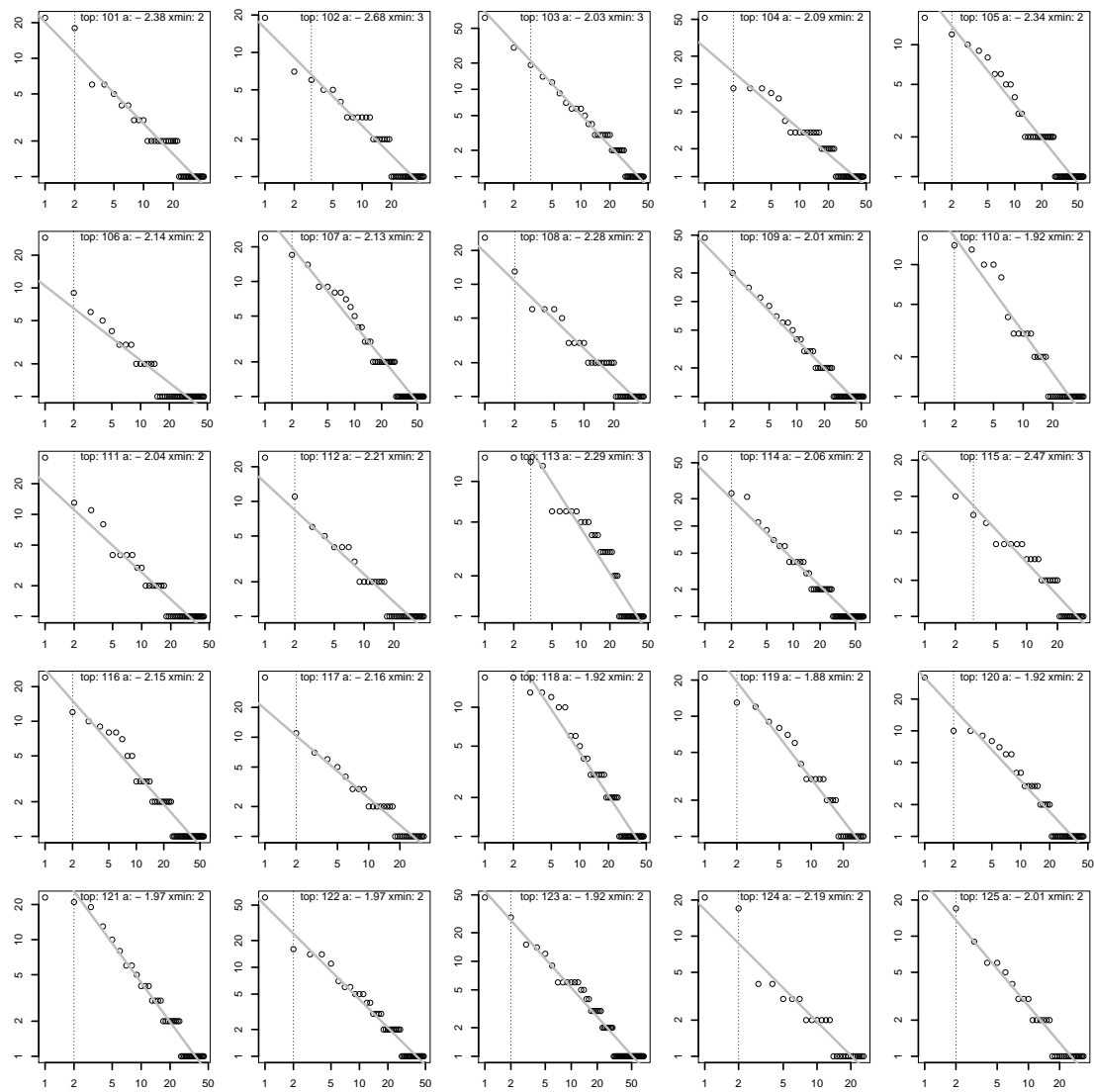


Abbildung B.20: Doppelt logarithmierter Plot der Produktivität des Erscheinungslandes für die Topics 101-125 (CLEF Jahrgang 2004).



B.4 Produktivität des Erscheinungslandes in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
101	location	2,375	0,06	2	<b>0,871</b>	0,06
102	location	2,678	0,068	3	<b>0,912</b>	0,068
103	location	2,026	0,068	3	<b>0,85</b>	0,068
104	location	2,09	0,123	2	<b>0,217</b>	0,123
105	location	2,34	0,093	2	<b>0,374</b>	0,093
106	location	2,142	0,043	2	<b>1</b>	0,043
107	location	2,13	0,088	2	<b>0,475</b>	0,088
108	location	2,282	0,069	2	<b>0,842</b>	0,069
109	location	2,01	0,056	2	<b>0,965</b>	0,056
110	location	1,922	0,14	2	<b>0,285</b>	0,14
111	location	2,035	0,074	2	<b>0,907</b>	0,074
112	location	2,209	0,089	2	<b>0,728</b>	0,089
113	location	2,295	0,116	3	<b>0,344</b>	0,116
114	location	2,06	0,062	2	<b>0,868</b>	0,062
115	location	2,468	0,099	3	<b>0,521</b>	0,099
116	location	2,15	0,099	2	<b>0,431</b>	0,099
117	location	2,158	0,04	2	<b>1</b>	0,04
118	location	1,922	0,105	2	<b>0,384</b>	0,105
119	location	1,885	0,134	2	<b>0,342</b>	0,134
120	location	1,922	0,129	2	<b>0,279</b>	0,129
121	location	1,97	0,064	2	<b>0,877</b>	0,064
122	location	1,973	0,068	2	<b>0,827</b>	0,068
123	location	1,92	0,097	2	<b>0,398</b>	0,097
124	location	2,191	0,106	2	<b>0,669</b>	0,106
125	location	2,012	0,06	2	<b>0,987</b>	0,06

Tabelle B.20: Einzelwerte der Produktivität des Erscheinungslandes für die Topics 101-125 (CLEF Jahrgang 2004). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

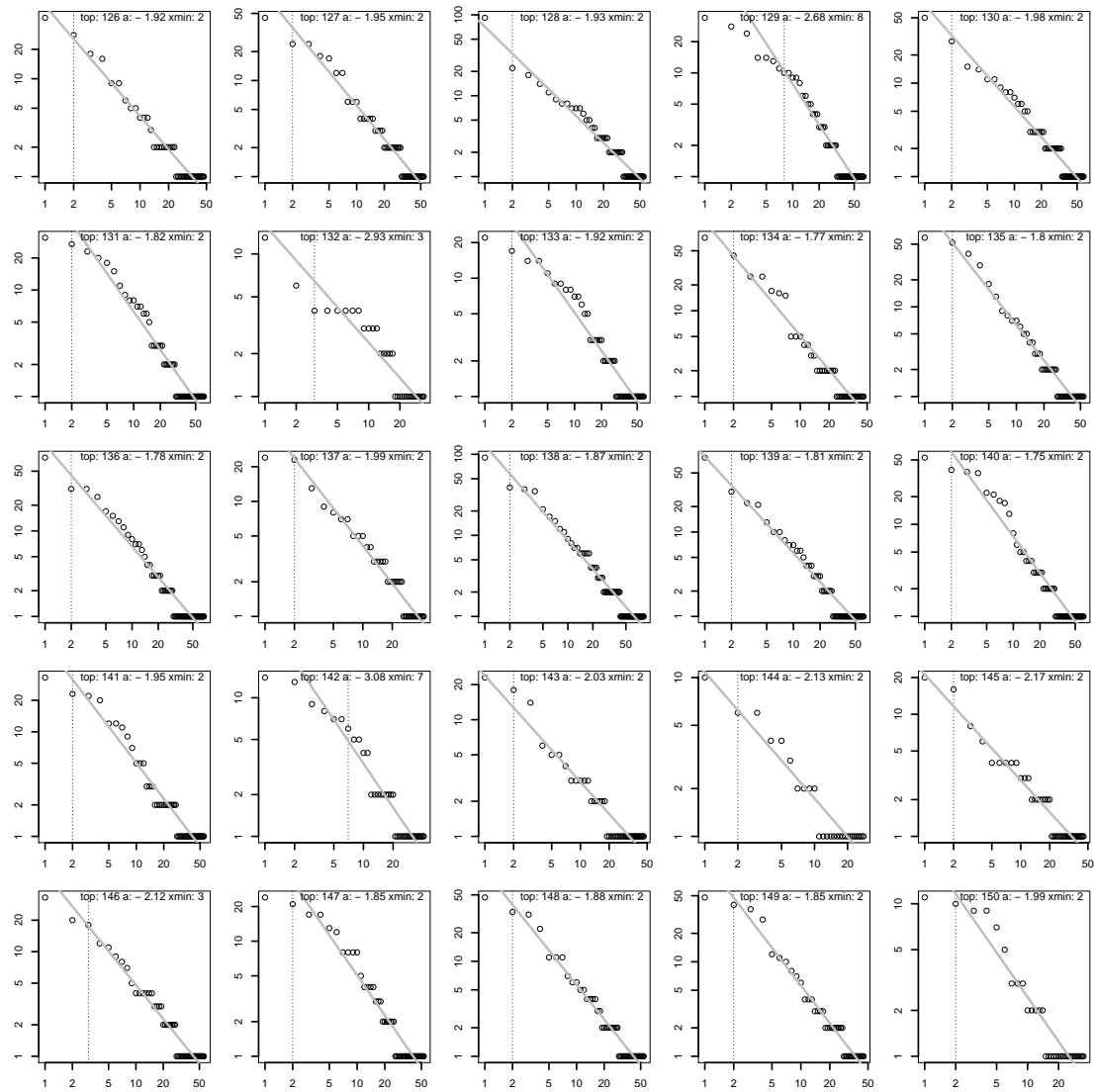


Abbildung B.21: Doppelt logarithmierter Plot der Produktivität des Erscheinunglandes für die Topics 126-150 (CLEF Jahrgang 2005).

B.4 Produktivität des Erscheinungslandes in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
126	location	1,918	0,067	2	<b>0,889</b>	0,067
127	location	1,95	0,084	2	<b>0,567</b>	0,084
128	location	1,93	0,081	2	<b>0,572</b>	0,081
129	location	2,68	0,106	8	<b>0,601</b>	0,106
130	location	1,98	0,075	2	<b>0,66</b>	0,075
131	location	1,82	0,13	2	<b>0,121</b>	0,13
132	location	2,934	0,165	3	<b>0,121</b>	0,165
133	location	1,92	0,127	2	<b>0,137</b>	0,127
134	location	1,775	0,141	2	<b>0,172</b>	0,141
135	location	1,8	0,057	2	<b>0,96</b>	0,057
136	location	1,78	0,088	2	<b>0,596</b>	0,088
137	location	1,989	0,076	2	<b>0,736</b>	0,076
138	location	1,87	0,098	2	<b>0,195</b>	0,098
139	location	1,81	0,087	2	<b>0,612</b>	0,087
140	location	1,75	0,107	2	<b>0,325</b>	0,107
141	location	1,95	0,098	2	<b>0,428</b>	0,098
142	location	3,083	0,137	7	<b>0,529</b>	0,137
143	location	2,029	0,08	2	<b>0,852</b>	0,08
144	location	2,134	0,143	2	<b>0,534</b>	0,143
145	location	2,168	0,084	2	<b>0,693</b>	0,084
146	location	2,12	0,098	3	<b>0,382</b>	0,098
147	location	1,85	0,139	2	<b>0,14</b>	0,139
148	location	1,88	0,086	2	<b>0,591</b>	0,086
149	location	1,847	0,069	2	<b>0,822</b>	0,069
150	location	1,994	0,142	2	<b>0,354</b>	0,142

Tabelle B.21: Einzelwerte der Produktivität des Erscheinungslandes für die Topics 126-150 (CLEF Jahrgang 2005). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

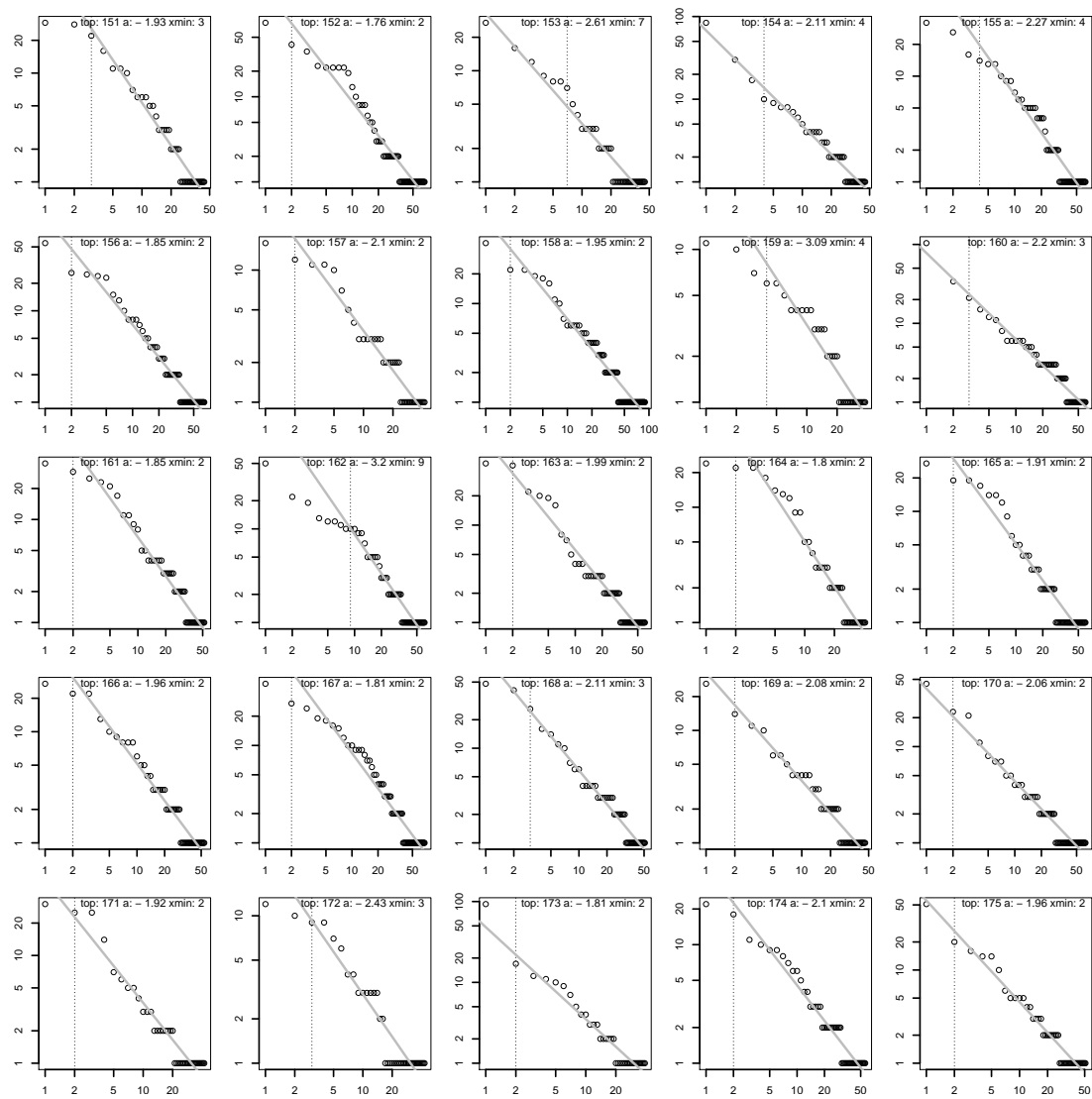


Abbildung B.22: Doppelt logarithmierter Plot der Produktivität des Erscheinunglandes für die Topics 151-175 (CLEF Jahrgang 2006).

B.4 Produktivität des Erscheinungslandes in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
151	location	1,928	0,117	3	<b>0,326</b>	0,117
152	location	1,76	0,113	2	<b>0,184</b>	0,113
153	location	2,611	0,092	7	<b>0,86</b>	0,092
154	location	2,111	0,077	4	<b>0,767</b>	0,077
155	location	2,27	0,102	4	<b>0,505</b>	0,102
156	location	1,85	0,062	2	<b>0,83</b>	0,062
157	location	2,1	0,093	2	<b>0,517</b>	0,093
158	location	1,95	0,103	2	<b>0,145</b>	0,103
159	location	3,091	0,096	4	<b>0,752</b>	0,096
160	location	2,2	0,072	3	<b>0,723</b>	0,072
161	location	1,85	0,092	2	<b>0,437</b>	0,092
162	location	3,2	0,112	9	<b>0,718</b>	0,112
163	location	1,99	0,091	2	<b>0,435</b>	0,091
164	location	1,805	0,136	2	<b>0,149</b>	0,136
165	location	1,91	0,098	2	<b>0,405</b>	0,098
166	location	1,96	0,091	2	<b>0,496</b>	0,091
167	location	1,81	0,109	2	<b>0,159</b>	0,109
168	location	2,11	0,091	3	<b>0,474</b>	0,091
169	location	2,083	0,104	2	<b>0,383</b>	0,104
170	location	2,06	0,048	2	<b>0,972</b>	0,048
171	location	1,922	0,081	2	<b>0,791</b>	0,081
172	location	2,43	0,13	3	<b>0,419</b>	0,13
173	location	1,815	0,085	2	<b>0,84</b>	0,085
174	location	2,1	0,085	2	<b>0,495</b>	0,085
175	location	1,96	0,067	2	<b>0,861</b>	0,067

Tabelle B.22: Einzelwerte der Produktivität des Erscheinungslandes für die Topics 151-175 (CLEF Jahrgang 2006). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

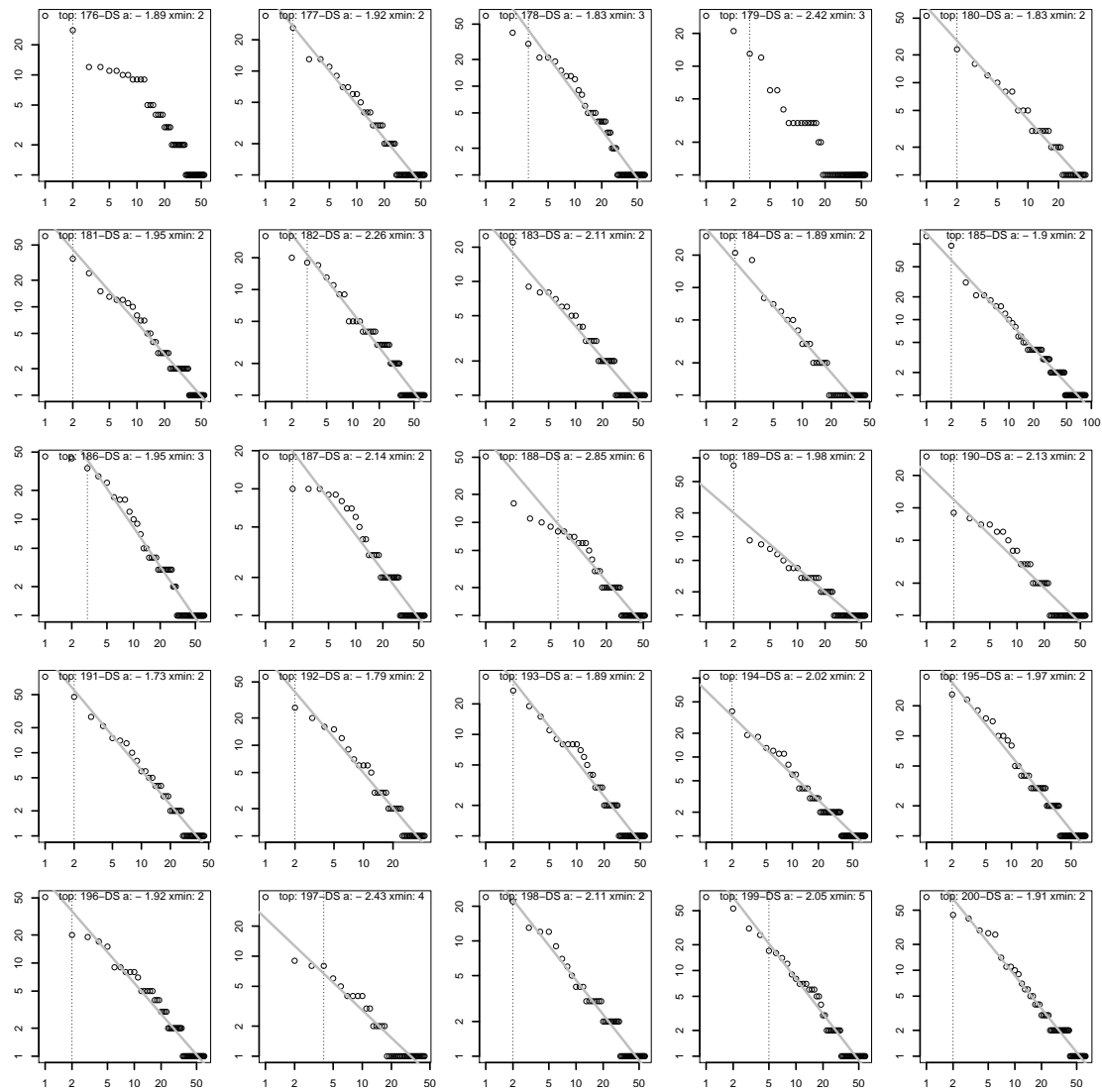


Abbildung B.23: Doppelt logarithmierter Plot der Produktivität des Erscheinungslandes für die Topics 176-200 (CLEF Jahrgang 2007).

B.4 Produktivität des Erscheinungslandes in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
10.2452/176-DS	location	1,89	0,144	2	0,037	0,144
10.2452/177-DS	location	1,92	0,094	2	<b>0,497</b>	0,094
10.2452/178-DS	location	1,83	0,12	3	<b>0,224</b>	0,12
10.2452/179-DS	location	2,42	0,189	3	0,048	0,189
10.2452/180-DS	location	1,829	0,11	2	<b>0,477</b>	0,11
10.2452/181-DS	location	1,95	0,078	2	<b>0,553</b>	0,078
10.2452/182-DS	location	2,26	0,09	3	<b>0,445</b>	0,09
10.2452/183-DS	location	2,11	0,065	2	<b>0,819</b>	0,065
10.2452/184-DS	location	1,888	0,079	2	<b>0,893</b>	0,079
10.2452/185-DS	location	1,9	0,055	2	<b>0,801</b>	0,055
10.2452/186-DS	location	1,95	0,118	3	<b>0,249</b>	0,118
10.2452/187-DS	location	2,14	0,105	2	<b>0,221</b>	0,105
10.2452/188-DS	location	2,85	0,09	6	<b>0,841</b>	0,09
10.2452/189-DS	location	1,98	0,09	2	<b>0,581</b>	0,09
10.2452/190-DS	location	2,13	0,084	2	<b>0,685</b>	0,084
10.2452/191-DS	location	1,731	0,081	2	<b>0,769</b>	0,081
10.2452/192-DS	location	1,795	0,11	2	<b>0,408</b>	0,11
10.2452/193-DS	location	1,89	0,114	2	<b>0,235</b>	0,114
10.2452/194-DS	location	2,02	0,08	2	<b>0,489</b>	0,08
10.2452/195-DS	location	1,97	0,073	2	<b>0,609</b>	0,073
10.2452/196-DS	location	1,92	0,097	2	<b>0,333</b>	0,097
10.2452/197-DS	location	2,431	0,107	4	<b>0,69</b>	0,107
10.2452/198-DS	location	2,11	0,058	2	<b>0,842</b>	0,058
10.2452/199-DS	location	2,05	0,078	5	<b>0,891</b>	0,078
10.2452/200-DS	location	1,91	0,068	2	<b>0,619</b>	0,068

Tabelle B.23: Einzelwerte der Produktivität des Erscheinungslandes für die Topics 176-200 (CLEF Jahrgang 2007). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

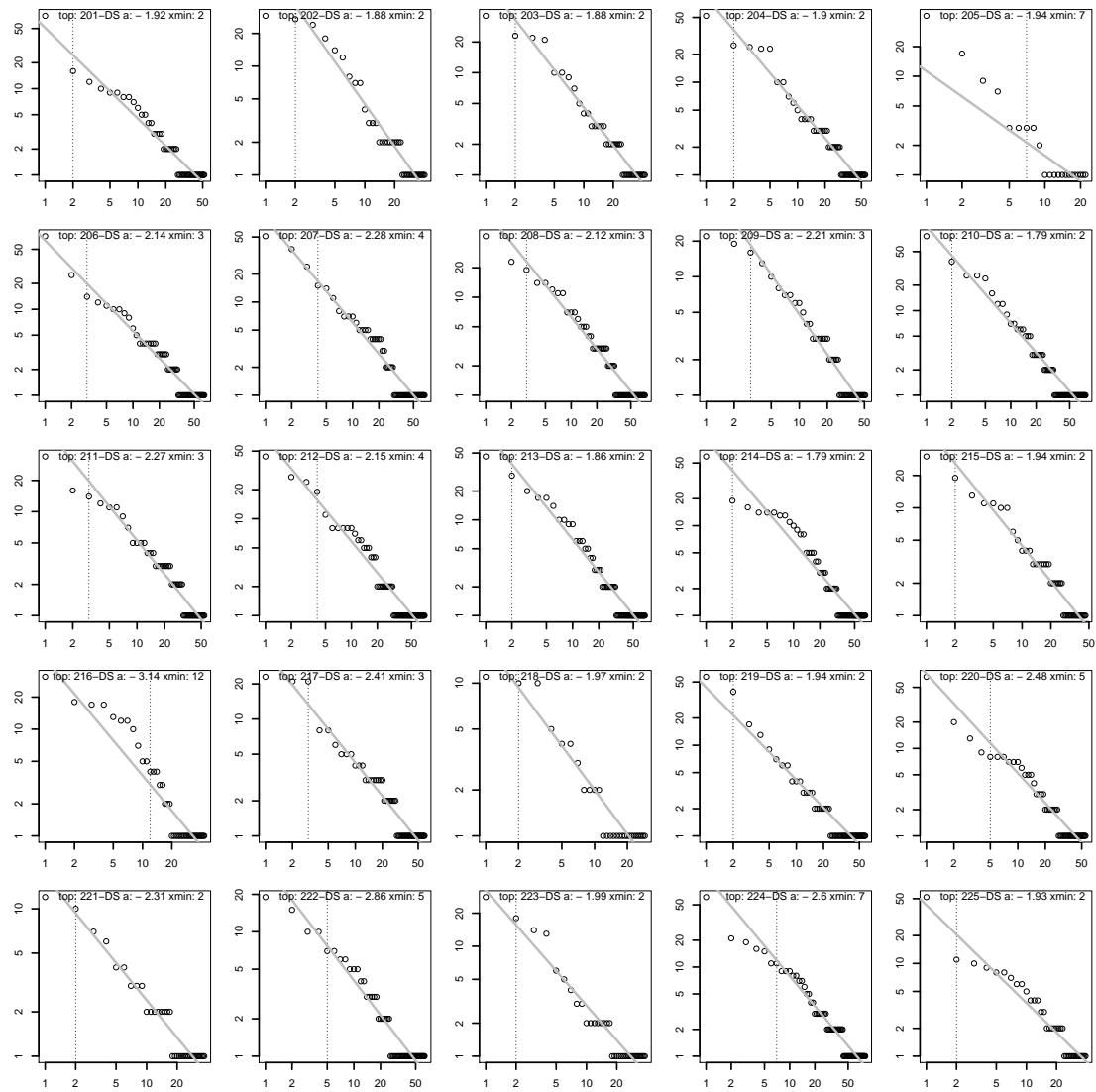


Abbildung B.24: Doppelt logarithmierter Plot der Produktivität des Erscheinungslandes für die Topics 201-225 (CLEF Jahrgang 2008).



B.4 Produktivität des Erscheinungslandes in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
10.2452/201-DS	location	1,92	0,088	2	<b>0,581</b>	0,088
10.2452/202-DS	location	1,88	0,125	2	<b>0,249</b>	0,125
10.2452/203-DS	location	1,882	0,083	2	<b>0,735</b>	0,083
10.2452/204-DS	location	1,9	0,083	2	<b>0,593</b>	0,083
10.2452/205-DS	location	1,938	0,187	7	<b>0,778</b>	0,187
10.2452/206-DS	location	2,14	0,085	3	<b>0,561</b>	0,085
10.2452/207-DS	location	2,28	0,059	4	<b>0,975</b>	0,059
10.2452/208-DS	location	2,12	0,104	3	<b>0,344</b>	0,104
10.2452/209-DS	location	2,21	0,11	3	<b>0,358</b>	0,11
10.2452/210-DS	location	1,79	0,103	2	<b>0,315</b>	0,103
10.2452/211-DS	location	2,27	0,096	3	<b>0,372</b>	0,096
10.2452/212-DS	location	2,15	0,107	4	<b>0,486</b>	0,107
10.2452/213-DS	location	1,86	0,092	2	<b>0,462</b>	0,092
10.2452/214-DS	location	1,79	0,133	2	<b>0,109</b>	0,133
10.2452/215-DS	location	1,942	0,112	2	<b>0,284</b>	0,112
10.2452/216-DS	location	3,143	0,165	12	<b>0,393</b>	0,165
10.2452/217-DS	location	2,41	0,097	3	<b>0,409</b>	0,097
10.2452/218-DS	location	1,973	0,13	2	<b>0,673</b>	0,13
10.2452/219-DS	location	1,94	0,045	2	<b>0,999</b>	0,045
10.2452/220-DS	location	2,48	0,132	5	<b>0,245</b>	0,132
10.2452/221-DS	location	2,308	0,063	2	<b>0,94</b>	0,063
10.2452/222-DS	location	2,86	0,112	5	<b>0,422</b>	0,112
10.2452/223-DS	location	1,988	0,124	2	<b>0,416</b>	0,124
10.2452/224-DS	location	2,6	0,079	7	<b>0,675</b>	0,079
10.2452/225-DS	location	1,928	0,114	2	<b>0,353</b>	0,114

Tabelle B.24: Einzelwerte der Produktivität des Erscheinungslandes für die Topics 201-225 (CLEF Jahrgang 2008). Statistisch signifikante Werte sind **fett** gedruckt.

## B.5 Produktivität der Forschungsmethode in GIRT

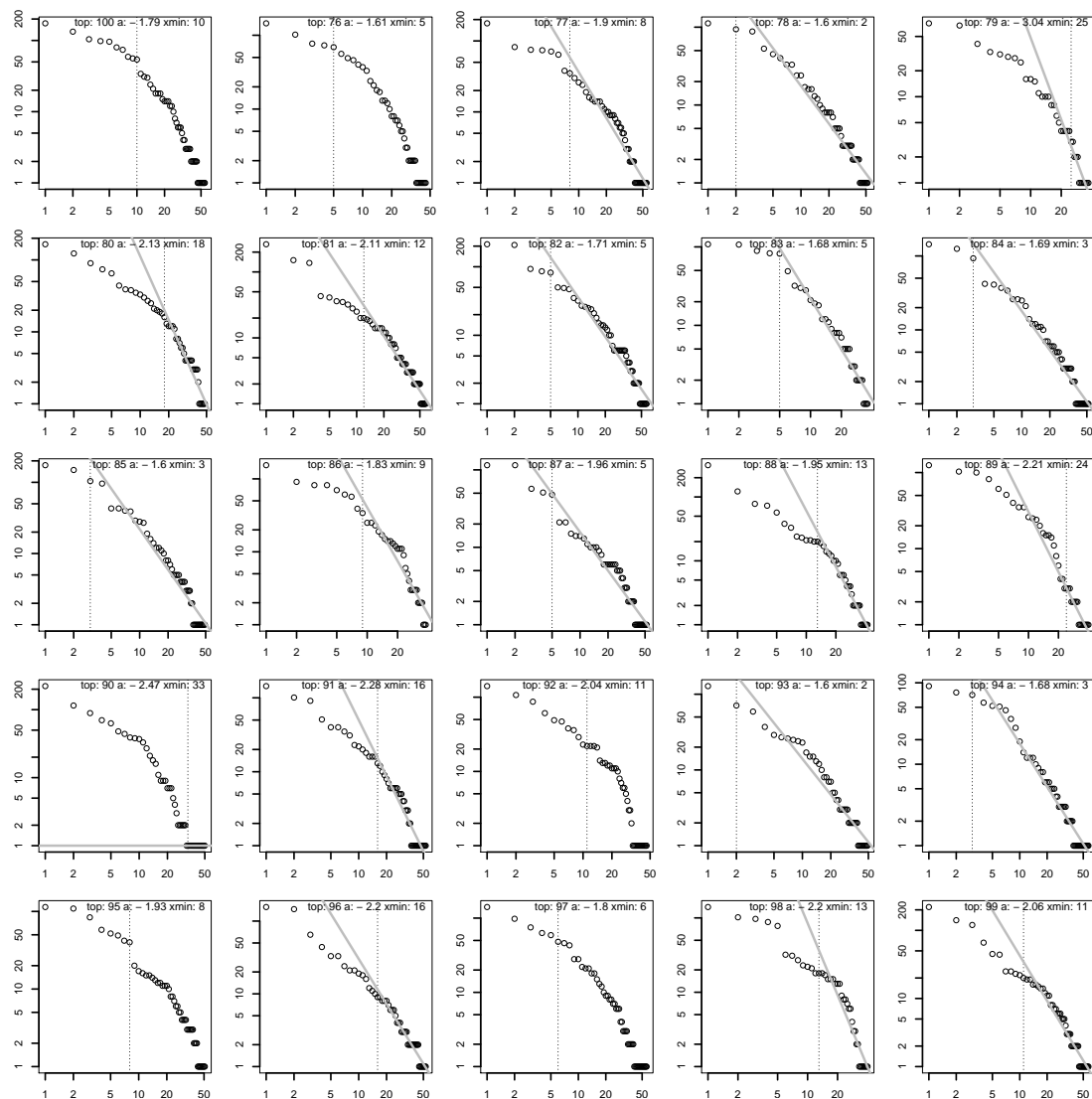


Abbildung B.25: Doppelt logarithmierter Plot der Produktivität der Forschungsmethode für die Topics 76-100 (CLEF Jahrgang 2003).

B.5 Produktivität der Forschungsmethode in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
76	method	1,615	0,143	5	0,074	0,143
77	method	1,9	0,094	8	<b>0,533</b>	0,094
78	method	1,6	0,113	2	<b>0,135</b>	0,113
79	method	3,039	0,152	25	<b>0,521</b>	0,152
80	method	2,129	0,113	18	<b>0,545</b>	0,113
81	method	2,11	0,086	12	<b>0,772</b>	0,086
82	method	1,71	0,103	5	<b>0,212</b>	0,103
83	method	1,68	0,118	5	<b>0,352</b>	0,118
84	method	1,69	0,099	3	<b>0,383</b>	0,099
85	method	1,602	0,091	3	<b>0,551</b>	0,091
86	method	1,832	0,124	9	<b>0,369</b>	0,124
87	method	1,96	0,081	5	<b>0,76</b>	0,081
88	method	1,951	0,15	13	<b>0,139</b>	0,15
89	method	2,21	0,148	24	<b>0,439</b>	0,148
90	method	2,473	0,108	33	<b>0,964</b>	0,108
91	method	2,28	0,123	16	<b>0,384</b>	0,123
92	method	2,04	0,154	11	0,091	0,154
93	method	1,604	0,117	2	<b>0,141</b>	0,117
94	method	1,68	0,087	3	<b>0,487</b>	0,087
95	method	1,93	0,138	8	0,047	0,138
96	method	2,2	0,099	16	<b>0,587</b>	0,099
97	method	1,8	0,139	6	0,092	0,139
98	method	2,2	0,138	13	<b>0,283</b>	0,138
99	method	2,06	0,105	11	<b>0,486</b>	0,105
100	method	1,79	0,145	10	0,048	0,145

Tabelle B.25: Einzelwerte der Produktivität der Forschungsmethode für die Topics 76-100 (CLEF Jahrgang 2003). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

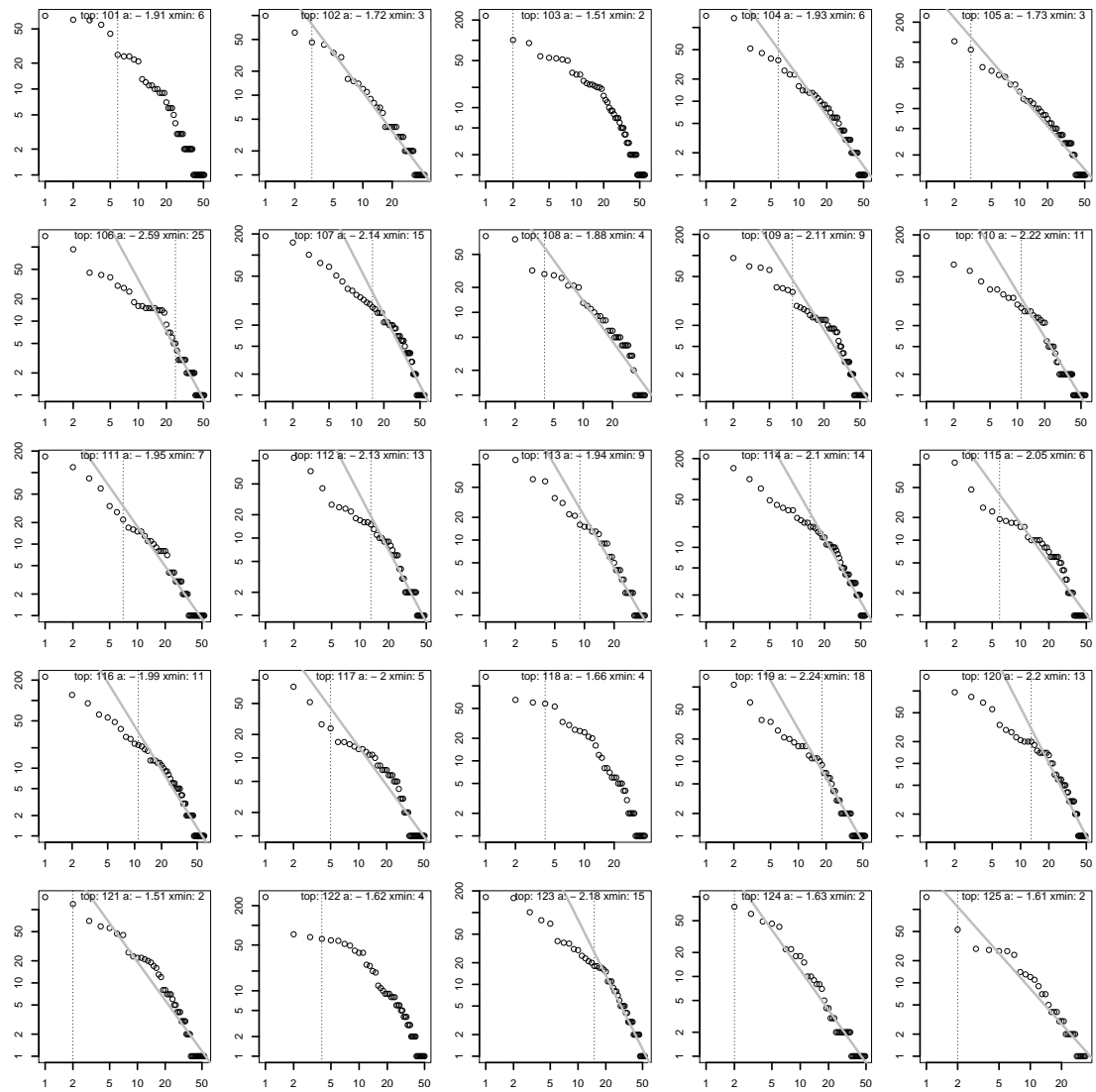


Abbildung B.26: Doppelt logarithmierter Plot der Produktivität der Forschungsmethode für die Topics 101-125 (CLEF Jahrgang 2004).

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
101	method	1,91	0,152	6	0,034	0,152
102	method	1,721	0,085	3	<b>0,578</b>	0,085
103	method	1,51	0,151	2	0,024	0,151
104	method	1,93	0,076	6	<b>0,789</b>	0,076
105	method	1,73	0,076	3	<b>0,703</b>	0,076
106	method	2,59	0,138	25	<b>0,779</b>	0,138
107	method	2,14	0,09	15	<b>0,784</b>	0,09
108	method	1,879	0,1	4	<b>0,406</b>	0,1
109	method	2,11	0,095	9	<b>0,578</b>	0,095
110	method	2,22	0,094	11	<b>0,845</b>	0,094
111	method	1,95	0,079	7	<b>0,914</b>	0,079
112	method	2,133	0,097	13	<b>0,731</b>	0,097
113	method	1,941	0,092	9	<b>0,85</b>	0,092
114	method	2,1	0,094	14	<b>0,683</b>	0,094
115	method	2,046	0,091	6	<b>0,595</b>	0,091
116	method	1,99	0,097	11	<b>0,533</b>	0,097
117	method	2	0,108	5	<b>0,399</b>	0,108
118	method	1,663	0,155	4	0,041	0,155
119	method	2,24	0,123	18	<b>0,52</b>	0,123
120	method	2,197	0,097	13	<b>0,758</b>	0,097
121	method	1,51	0,142	2	<b>0,115</b>	0,142
122	method	1,62	0,134	4	0,046	0,134
123	method	2,18	0,093	15	<b>0,793</b>	0,093
124	method	1,631	0,113	2	<b>0,202</b>	0,113
125	method	1,607	0,115	2	<b>0,342</b>	0,115

Tabelle B.26: Einzelwerte der Produktivität der Forschungsmethode für die Topics 101-125 (CLEF Jahrgang 2004). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

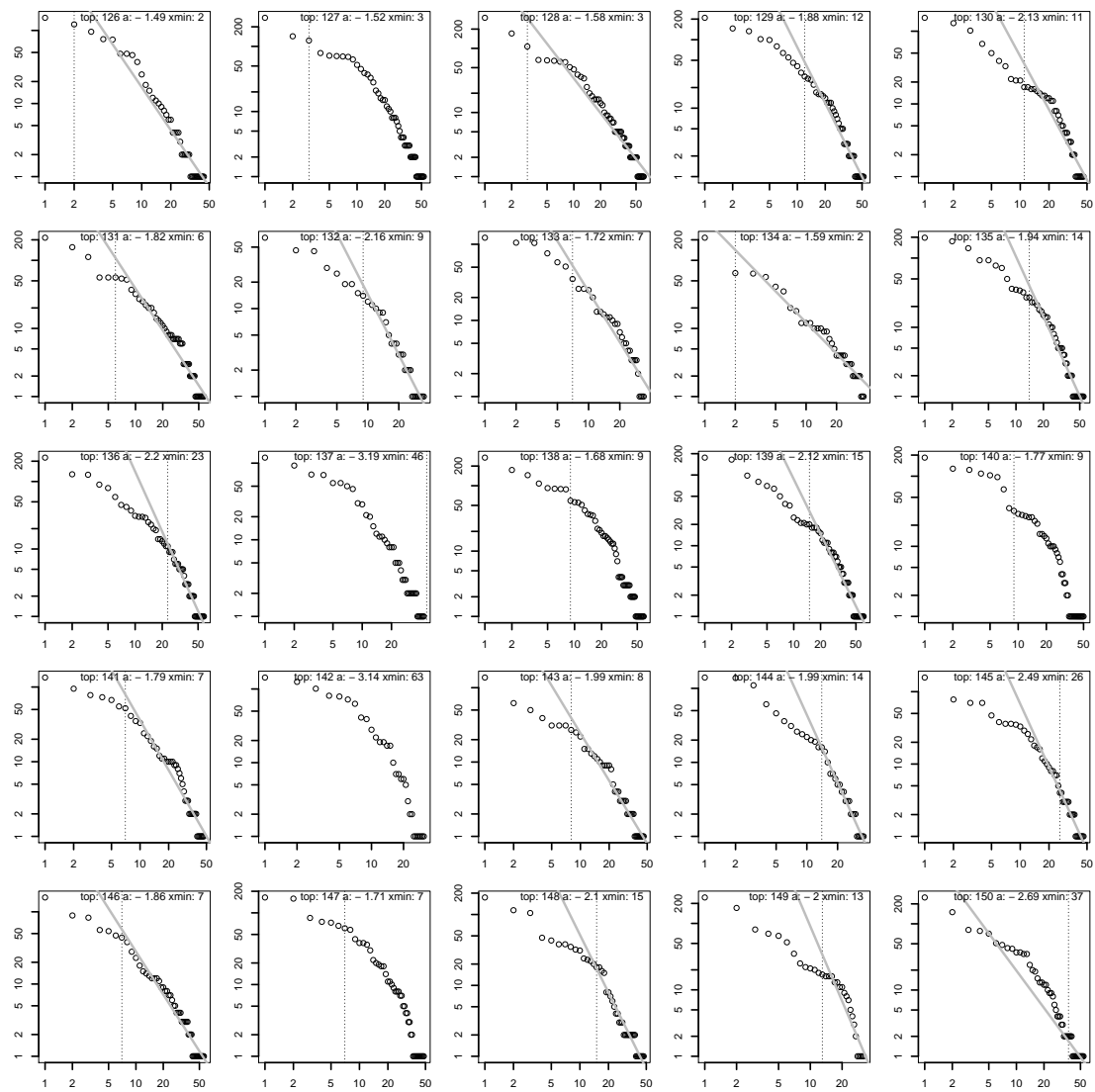


Abbildung B.27: Doppelt logarithmierter Plot der Produktivität der Forschungsmethode für die Topics 126-150 (CLEF Jahrgang 2005).

B.5 Produktivität der Forschungsmethode in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
126	method	1,494	0,123	2	<b>0,32</b>	0,123
127	method	1,52	0,138	3	0,081	0,138
128	method	1,58	0,105	3	<b>0,127</b>	0,105
129	method	1,88	0,112	12	<b>0,423</b>	0,112
130	method	2,126	0,093	11	<b>0,782</b>	0,093
131	method	1,82	0,1	6	<b>0,443</b>	0,1
132	method	2,161	0,121	9	<b>0,593</b>	0,121
133	method	1,722	0,135	7	<b>0,125</b>	0,135
134	method	1,592	0,118	2	<b>0,187</b>	0,118
135	method	1,94	0,109	14	<b>0,501</b>	0,109
136	method	2,2	0,114	23	<b>0,495</b>	0,114
137	method	3,188	0,153	46	<b>0,397</b>	0,153
138	method	1,68	0,166	9	0,009	0,166
139	method	2,12	0,109	15	<b>0,489</b>	0,109
140	method	1,768	0,145	9	0,082	0,145
141	method	1,788	0,12	7	<b>0,223</b>	0,12
142	method	3,143	0,149	63	<b>0,913</b>	0,149
143	method	1,99	0,141	8	<b>0,182</b>	0,141
144	method	1,989	0,091	14	<b>0,923</b>	0,091
145	method	2,494	0,16	26	<b>0,175</b>	0,16
146	method	1,86	0,111	7	<b>0,237</b>	0,111
147	method	1,714	0,162	7	0,017	0,162
148	method	2,105	0,136	15	<b>0,375</b>	0,136
149	method	2,001	0,11	13	<b>0,617</b>	0,11
150	method	2,69	0,137	37	<b>0,586</b>	0,137

Tabelle B.27: Einzelwerte der Produktivität der Forschungsmethode für die Topics 126-150 (CLEF Jahrgang 2005). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

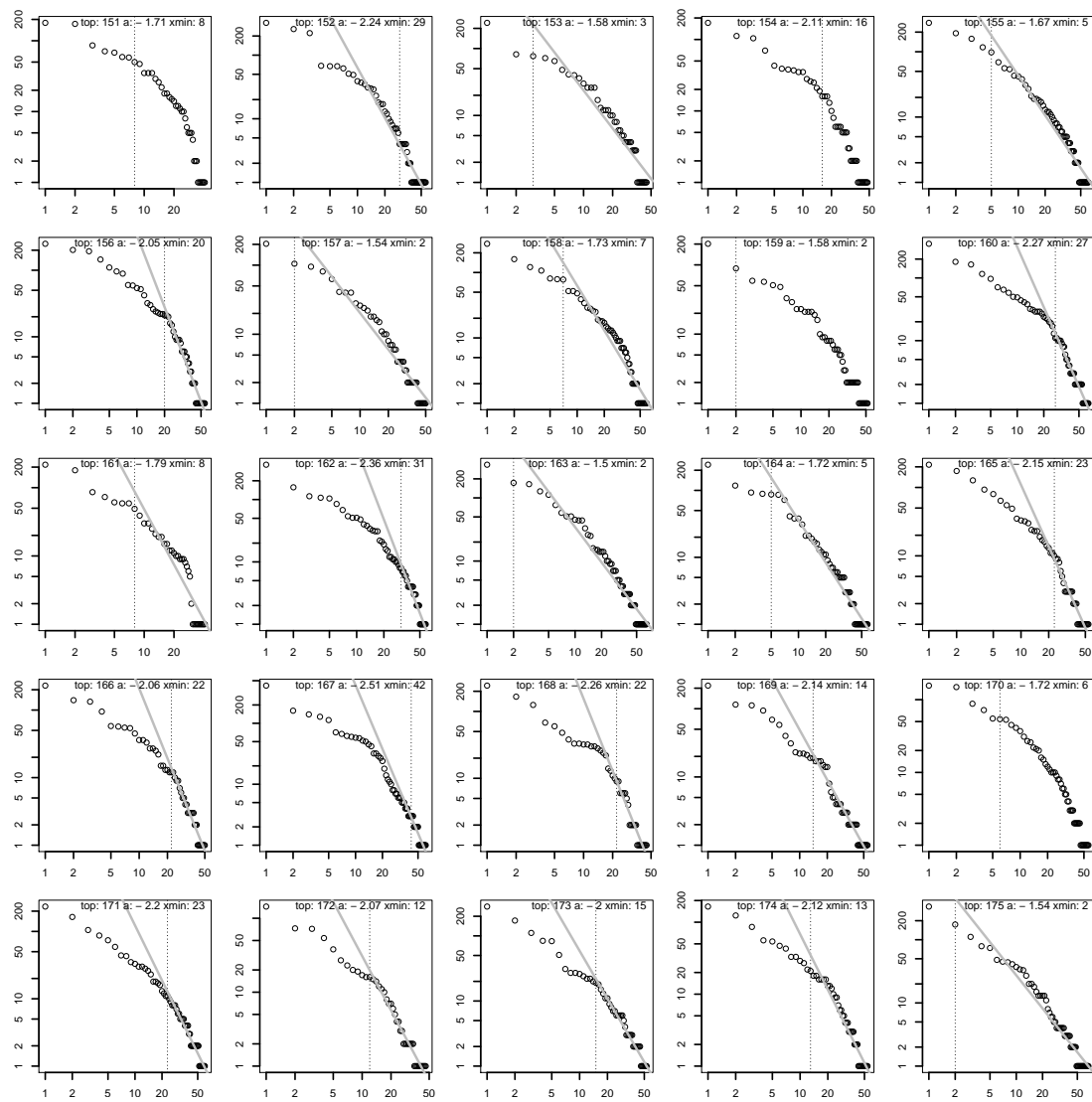


Abbildung B.28: Doppelt logarithmierter Plot der Produktivität der Forschungsmethode für die Topics 151-175 (CLEF Jahrgang 2006).



Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
151	method	1,712	0,138	8	0,097	0,138
152	method	2,24	0,137	29	<b>0,441</b>	0,137
153	method	1,583	0,122	3	<b>0,172</b>	0,122
154	method	2,114	0,166	16	0,072	0,166
155	method	1,67	0,095	5	<b>0,315</b>	0,095
156	method	2,05	0,139	20	<b>0,224</b>	0,139
157	method	1,536	0,109	2	<b>0,247</b>	0,109
158	method	1,73	0,106	7	<b>0,24</b>	0,106
159	method	1,58	0,132	2	0,052	0,132
160	method	2,27	0,088	27	<b>0,86</b>	0,088
161	method	1,788	0,1	8	<b>0,542</b>	0,1
162	method	2,36	0,104	31	<b>0,761</b>	0,104
163	method	1,5	0,118	2	<b>0,139</b>	0,118
164	method	1,72	0,095	5	<b>0,495</b>	0,095
165	method	2,15	0,12	23	<b>0,509</b>	0,12
166	method	2,059	0,143	22	<b>0,207</b>	0,143
167	method	2,51	0,111	42	<b>0,721</b>	0,111
168	method	2,261	0,14	22	<b>0,439</b>	0,14
169	method	2,14	0,132	14	<b>0,347</b>	0,132
170	method	1,72	0,131	6	0,055	0,131
171	method	2,2	0,092	23	<b>0,746</b>	0,092
172	method	2,073	0,102	12	<b>0,724</b>	0,102
173	method	2	0,118	15	<b>0,289</b>	0,118
174	method	2,12	0,119	13	<b>0,36</b>	0,119
175	method	1,54	0,115	2	<b>0,105</b>	0,115

Tabelle B.28: Einzelwerte der Produktivität der Forschungsmethode für die Topics 151-175 (CLEF Jahrgang 2006). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

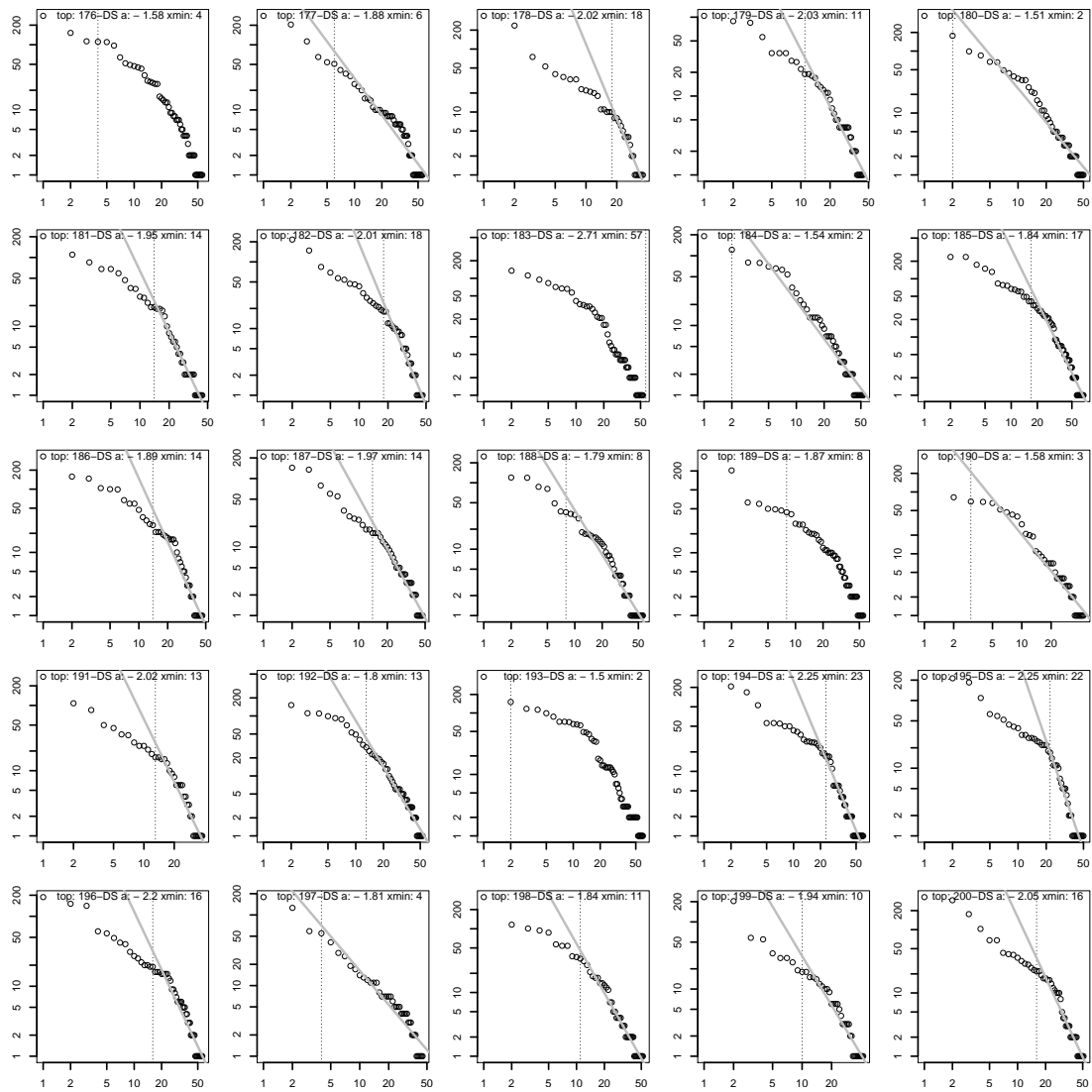


Abbildung B.29: Doppelt logarithmierter Plot der Produktivität der Forschungsmethode für die Topics 176-200 (CLEF Jahrgang 2007).

B.5 Produktivität der Forschungsmethode in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
10.2452/176-DS	method	1,58	0,149	4	0,017	0,149
10.2452/177-DS	method	1,88	0,071	6	<b>0,812</b>	0,071
10.2452/178-DS	method	2,015	0,109	18	<b>0,836</b>	0,109
10.2452/179-DS	method	2,033	0,126	11	<b>0,364</b>	0,126
10.2452/180-DS	method	1,508	0,097	2	<b>0,395</b>	0,097
10.2452/181-DS	method	1,954	0,128	14	<b>0,459</b>	0,128
10.2452/182-DS	method	2,014	0,139	18	<b>0,262</b>	0,139
10.2452/183-DS	method	2,71	0,113	57	<b>0,858</b>	0,113
10.2452/184-DS	method	1,54	0,126	2	<b>0,118</b>	0,126
10.2452/185-DS	method	1,84	0,117	17	<b>0,223</b>	0,117
10.2452/186-DS	method	1,891	0,119	14	<b>0,427</b>	0,119
10.2452/187-DS	method	1,969	0,115	14	<b>0,356</b>	0,115
10.2452/188-DS	method	1,79	0,123	8	<b>0,197</b>	0,123
10.2452/189-DS	method	1,87	0,13	8	0,098	0,13
10.2452/190-DS	method	1,581	0,117	3	<b>0,254</b>	0,117
10.2452/191-DS	method	2,016	0,089	13	<b>0,93</b>	0,089
10.2452/192-DS	method	1,8	0,116	13	<b>0,267</b>	0,116
10.2452/193-DS	method	1,5	0,156	2	0,009	0,156
10.2452/194-DS	method	2,25	0,116	23	<b>0,494</b>	0,116
10.2452/195-DS	method	2,254	0,104	22	<b>0,762</b>	0,104
10.2452/196-DS	method	2,2	0,095	16	<b>0,759</b>	0,095
10.2452/197-DS	method	1,813	0,085	4	<b>0,631</b>	0,085
10.2452/198-DS	method	1,84	0,137	11	<b>0,206</b>	0,137
10.2452/199-DS	method	1,944	0,074	10	<b>0,977</b>	0,074
10.2452/200-DS	method	2,05	0,066	16	<b>0,979</b>	0,066

Tabelle B.29: Einzelwerte der Produktivität der Forschungsmethode für die Topics 176-200 (CLEF Jahrgang 2007). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

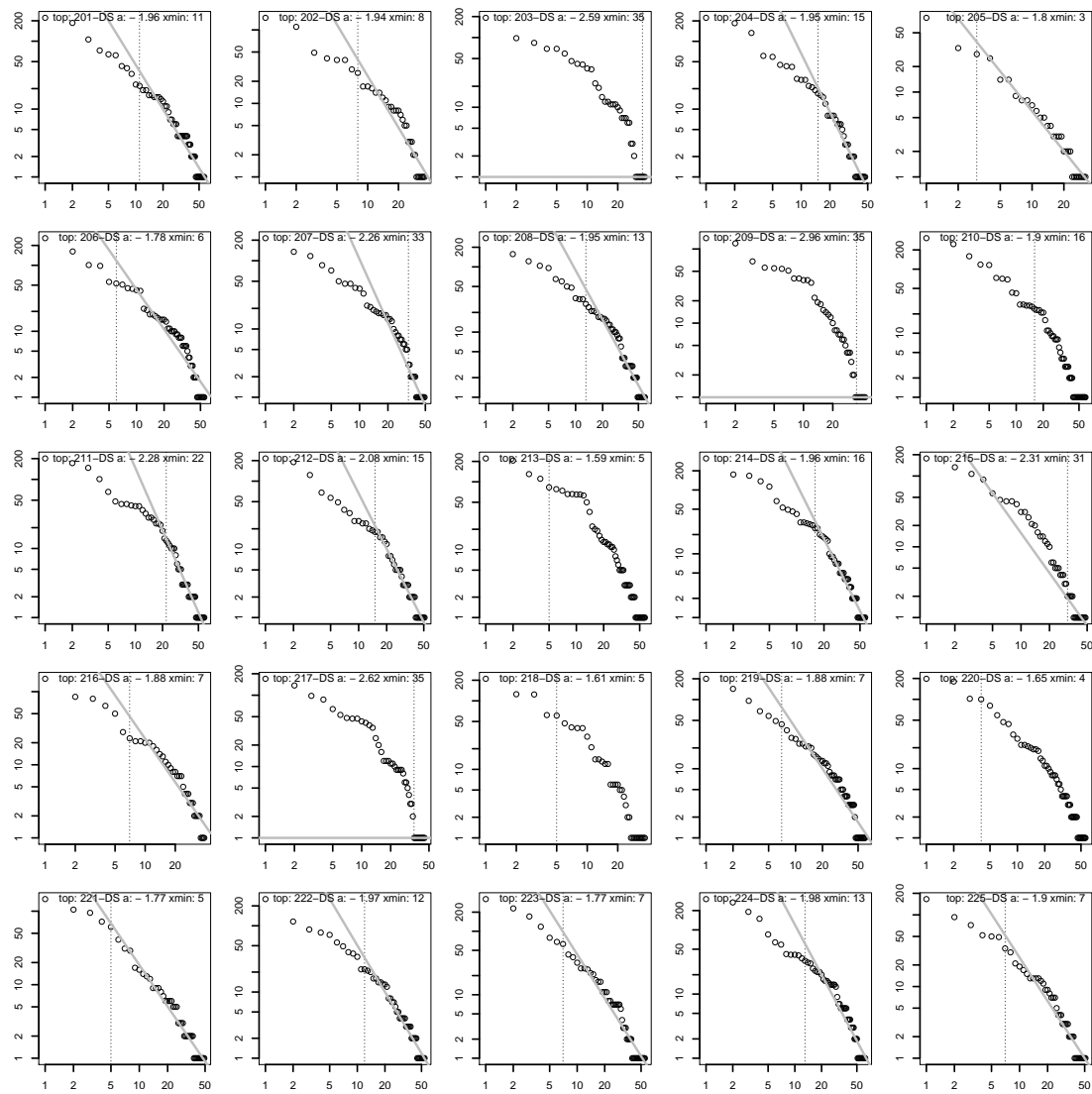


Abbildung B.30: Doppelt logarithmierter Plot der Produktivität der Forschungsmethode für die Topics 201-225 (CLEF Jahrgang 2008).

B.5 Produktivität der Forschungsmethode in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
10.2452/201-DS	method	1,96	0,089	11	<b>0,785</b>	0,089
10.2452/202-DS	method	1,94	0,102	8	<b>0,659</b>	0,102
10.2452/203-DS	method	2,591	0,153	35	<b>0,456</b>	0,153
10.2452/204-DS	method	1,951	0,125	15	<b>0,468</b>	0,125
10.2452/205-DS	method	1,805	0,079	3	<b>0,871</b>	0,079
10.2452/206-DS	method	1,78	0,103	6	<b>0,31</b>	0,103
10.2452/207-DS	method	2,264	0,12	33	<b>0,76</b>	0,12
10.2452/208-DS	method	1,95	0,094	13	<b>0,588</b>	0,094
10.2452/209-DS	method	2,962	0,141	35	<b>0,579</b>	0,141
10.2452/210-DS	method	1,9	0,175	16	0,021	0,175
10.2452/211-DS	method	2,28	0,134	22	<b>0,323</b>	0,134
10.2452/212-DS	method	2,08	0,077	15	<b>0,953</b>	0,077
10.2452/213-DS	method	1,59	0,151	5	0,02	0,151
10.2452/214-DS	method	1,96	0,128	16	<b>0,279</b>	0,128
10.2452/215-DS	method	2,309	0,148	31	<b>0,338</b>	0,148
10.2452/216-DS	method	1,88	0,119	7	<b>0,365</b>	0,119
10.2452/217-DS	method	2,615	0,107	35	<b>0,932</b>	0,107
10.2452/218-DS	method	1,611	0,178	5	0,046	0,178
10.2452/219-DS	method	1,88	0,104	7	<b>0,273</b>	0,104
10.2452/220-DS	method	1,65	0,133	4	0,036	0,133
10.2452/221-DS	method	1,768	0,091	5	<b>0,608</b>	0,091
10.2452/222-DS	method	1,97	0,122	12	<b>0,262</b>	0,122
10.2452/223-DS	method	1,77	0,108	7	<b>0,383</b>	0,108
10.2452/224-DS	method	1,98	0,086	13	<b>0,782</b>	0,086
10.2452/225-DS	method	1,9	0,111	7	<b>0,423</b>	0,111

Tabelle B.30: Einzelwerte der Produktivität der Forschungsmethode für die Topics 201-225 (CLEF Jahrgang 2008). Statistisch signifikante Werte sind **fett** gedruckt.

## B.6 Produktivität der Verlage in GIRT

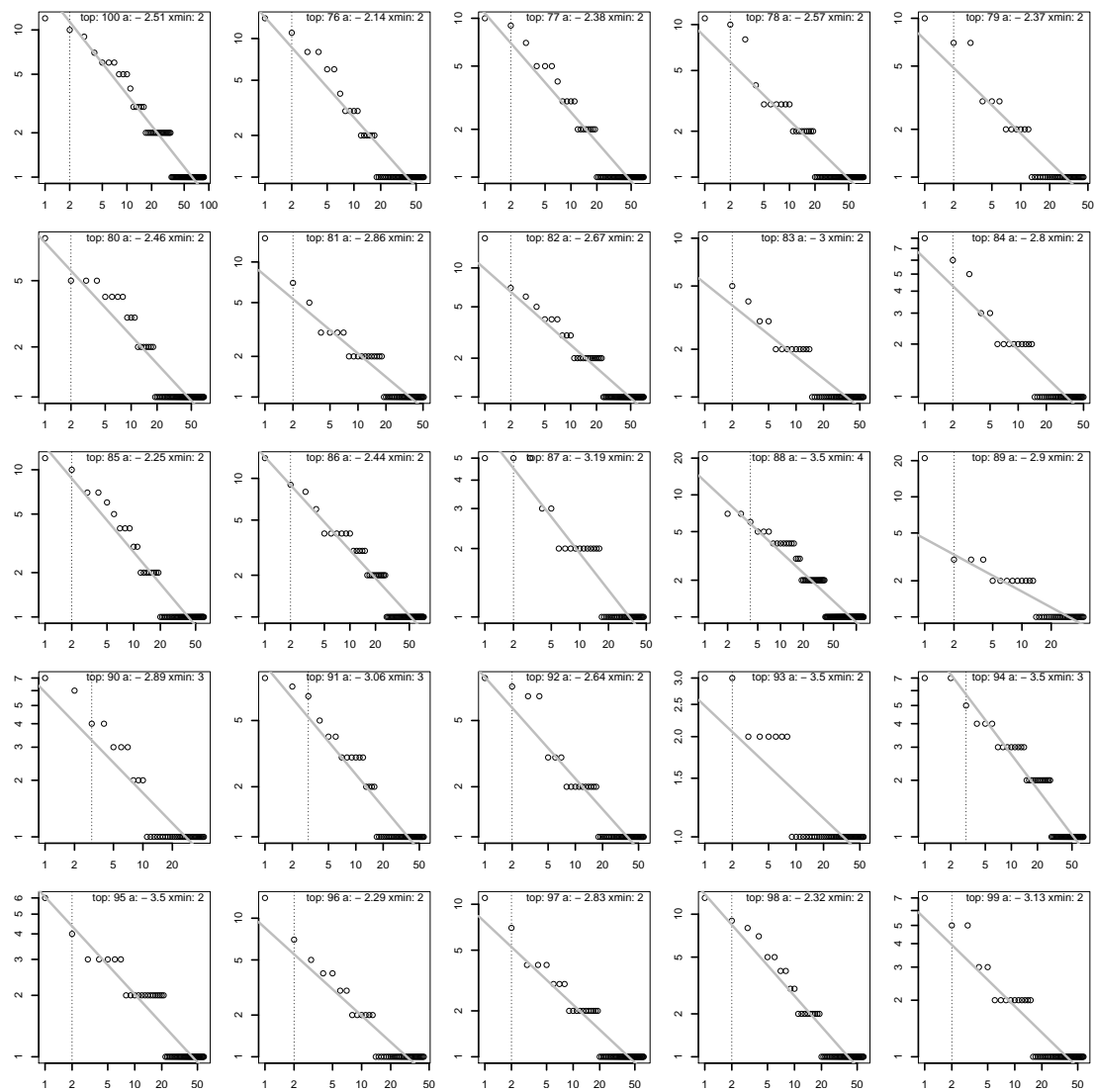


Abbildung B.31: Doppelt logarithmierter Plot der Produktivität der Verlage für die Topics 76-100 (CLEF Jahrgang 2003).

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
76	publisher	2,14	0,117	2	<b>0,427</b>	0,117
77	publisher	2,38	0,086	2	<b>0,677</b>	0,086
78	publisher	2,57	0,073	2	<b>0,733</b>	0,073
79	publisher	2,366	0,131	2	<b>0,436</b>	0,131
80	publisher	2,46	0,14	2	<b>0,165</b>	0,14
81	publisher	2,86	0,055	2	<b>0,888</b>	0,055
82	publisher	2,67	0,047	2	<b>0,937</b>	0,047
83	publisher	3	0,044	2	<b>0,978</b>	0,044
84	publisher	2,801	0,086	2	<b>0,656</b>	0,086
85	publisher	2,25	0,114	2	<b>0,385</b>	0,114
86	publisher	2,44	0,081	2	<b>0,503</b>	0,081
87	publisher	3,194	0,102	2	<b>0,409</b>	0,102
88	publisher	3,5	0,085	4	<b>0,593</b>	0,085
89	publisher	2,902	0,112	2	<b>0,436</b>	0,112
90	publisher	2,894	0,104	3	<b>0,842</b>	0,104
91	publisher	3,06	0,106	3	<b>0,664</b>	0,106
92	publisher	2,64	0,139	2	<b>0,16</b>	0,139
93	publisher	3,5	0,134	2	<b>0,592</b>	0,134
94	publisher	3,5	0,067	3	<b>0,695</b>	0,067
95	publisher	3,5	0,039	2	<b>0,969</b>	0,039
96	publisher	2,292	0,066	2	<b>0,966</b>	0,066
97	publisher	2,83	0,039	2	<b>0,984</b>	0,039
98	publisher	2,32	0,081	2	<b>0,698</b>	0,081
99	publisher	3,13	0,094	2	<b>0,499</b>	0,094
100	publisher	2,51	0,093	2	<b>0,22</b>	0,093

Tabelle B.31: Einzelwerte der Produktivität der Verlage für die Topics 76-100 (CLEF Jahrgang 2003). Statistisch signifikante Werte sind **fett** gedruckt.

B Ermittelte Werte der Power-Law-Analyse

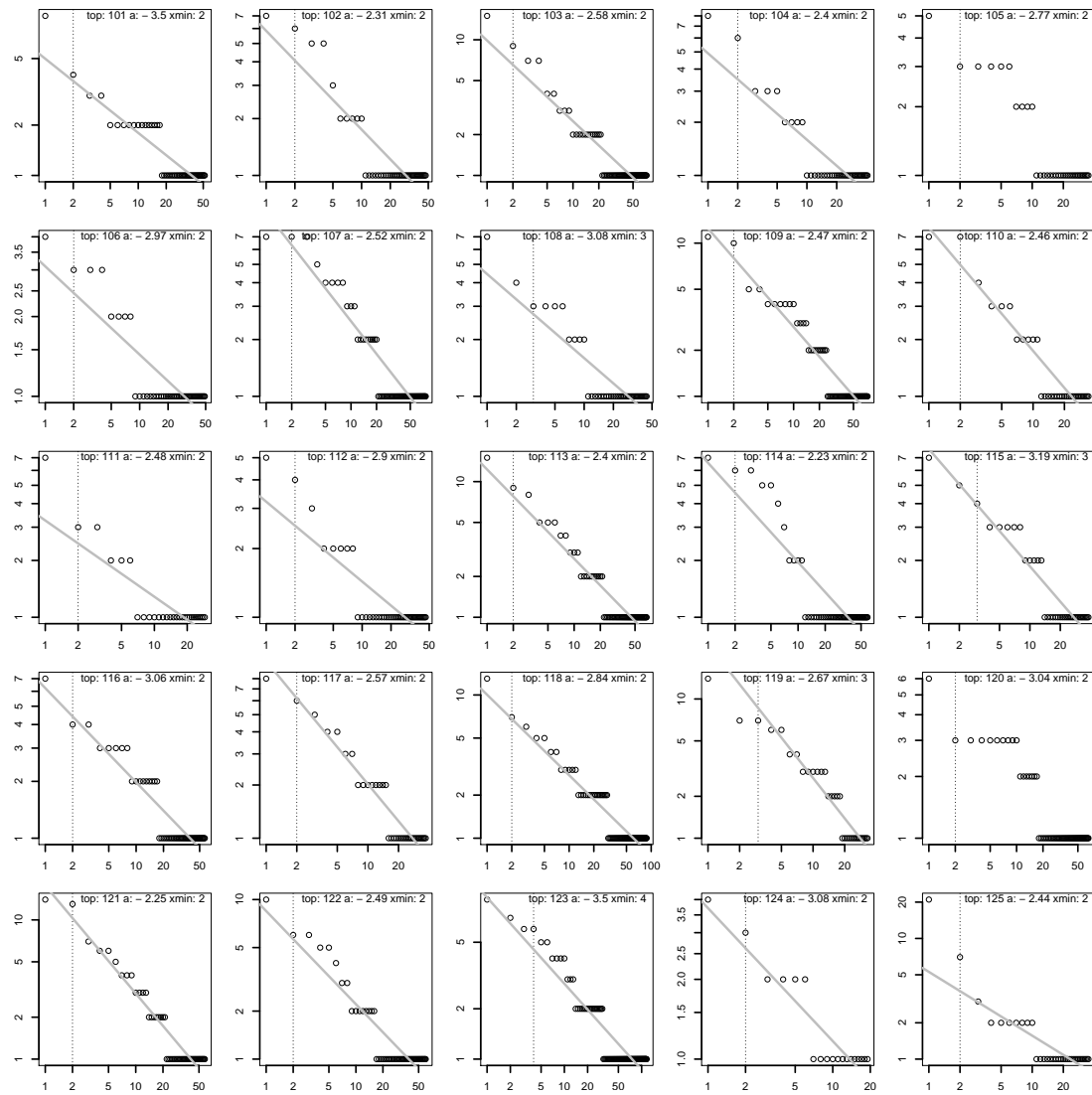


Abbildung B.32: Doppelt logarithmierter Plot der Produktivität der Verlage für die Topics 101-125 (CLEF Jahrgang 2004).



Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
101	publisher	3,5	0,067	2	<b>0,761</b>	0,067
102	publisher	2,314	0,188	2	<b>0,207</b>	0,188
103	publisher	2,58	0,086	2	<b>0,546</b>	0,086
104	publisher	2,404	0,091	2	<b>0,856</b>	0,091
105	publisher	2,773	0,213	2	0,092	0,213
106	publisher	2,969	0,16	2	<b>0,356</b>	0,16
107	publisher	2,52	0,11	2	<b>0,292</b>	0,11
108	publisher	3,083	0,109	3	<b>0,766</b>	0,109
109	publisher	2,47	0,115	2	<b>0,201</b>	0,115
110	publisher	2,455	0,085	2	<b>0,852</b>	0,085
111	publisher	2,483	0,087	2	<b>0,919</b>	0,087
112	publisher	2,899	0,077	2	<b>0,879</b>	0,077
113	publisher	2,4	0,062	2	<b>0,858</b>	0,062
114	publisher	2,23	0,185	2	<b>0,211</b>	0,185
115	publisher	3,188	0,067	3	<b>0,961</b>	0,067
116	publisher	3,06	0,1	2	<b>0,401</b>	0,1
117	publisher	2,568	0,078	2	<b>0,762</b>	0,078
118	publisher	2,84	0,03	2	<b>0,977</b>	0,03
119	publisher	2,671	0,139	3	<b>0,229</b>	0,139
120	publisher	3,04	0,214	2	0,024	0,214
121	publisher	2,25	0,076	2	<b>0,767</b>	0,076
122	publisher	2,49	0,107	2	<b>0,486</b>	0,107
123	publisher	3,5	0,084	4	<b>0,516</b>	0,084
124	publisher	3,083	0,072	2	<b>0,993</b>	0,072
125	publisher	2,44	0,16	2	<b>0,34</b>	0,16

Tabelle B.32: Einzelwerte der Produktivität der Verlage für die Topics 101-125 (CLEF Jahrgang 2004). Statistisch signifikante Werte sind **fett** gedruckt.

B *Ermittelte Werte der Power-Law-Analyse*

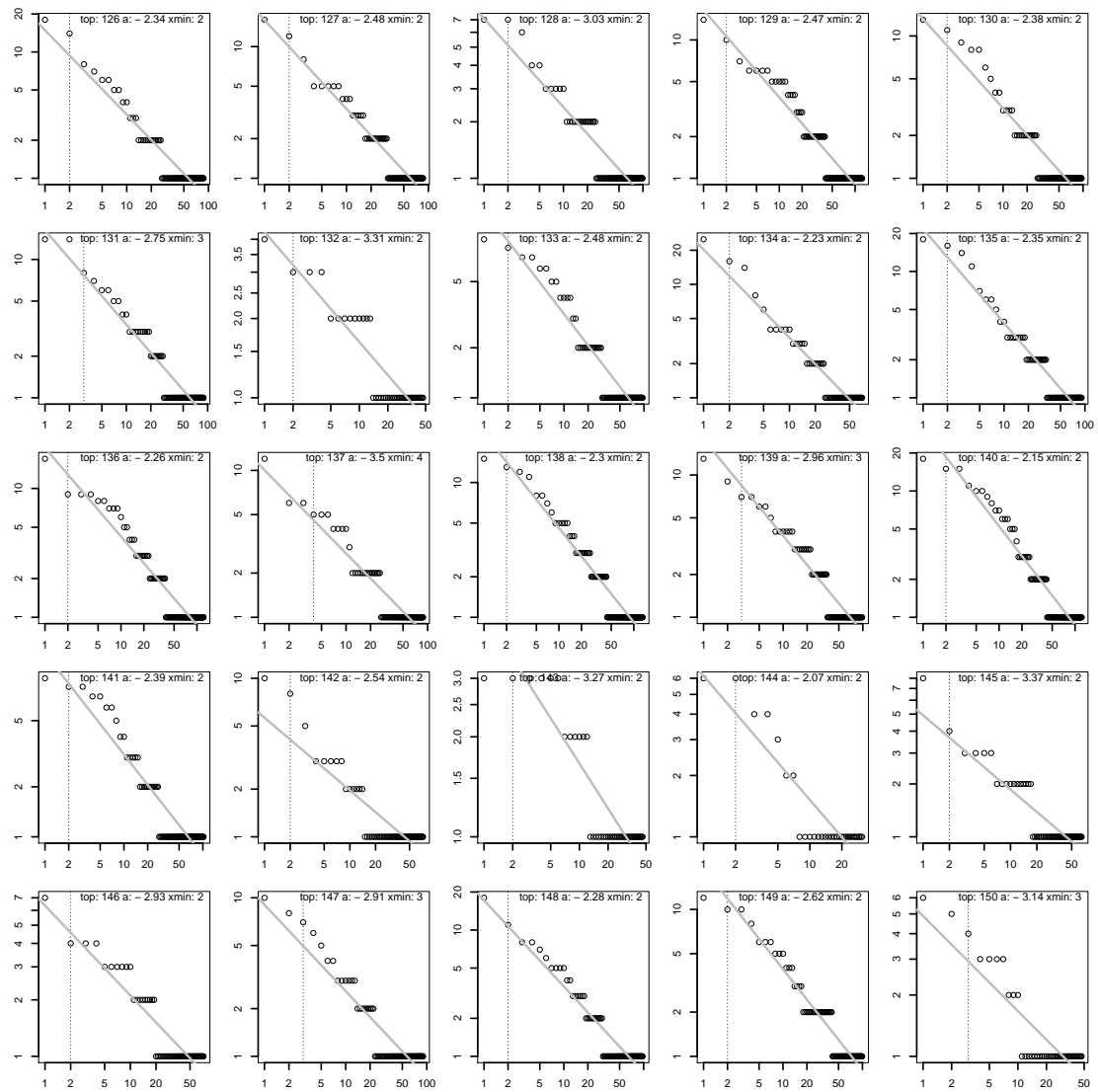


Abbildung B.33: Doppelt logarithmierter Plot der Produktivität der Verlage für die Topics 126-150 (CLEF Jahrgang 2005).

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
126	publisher	2,34	0,068	2	<b>0,721</b>	0,068
127	publisher	2,48	0,058	2	<b>0,729</b>	0,058
128	publisher	3,03	0,042	2	<b>0,919</b>	0,042
129	publisher	2,47	0,098	2	<b>0,149</b>	0,098
130	publisher	2,38	0,078	2	<b>0,54</b>	0,078
131	publisher	2,75	0,058	3	<b>0,911</b>	0,058
132	publisher	3,308	0,072	2	<b>0,854</b>	0,072
133	publisher	2,48	0,115	2	<b>0,132</b>	0,115
134	publisher	2,23	0,069	2	<b>0,741</b>	0,069
135	publisher	2,35	0,045	2	<b>0,898</b>	0,045
136	publisher	2,26	0,076	2	<b>0,39</b>	0,076
137	publisher	3,5	0,117	4	<b>0,393</b>	0,117
138	publisher	2,3	0,074	2	<b>0,342</b>	0,074
139	publisher	2,96	0,065	3	<b>0,724</b>	0,065
140	publisher	2,15	0,088	2	<b>0,246</b>	0,088
141	publisher	2,39	0,087	2	<b>0,413</b>	0,087
142	publisher	2,54	0,098	2	<b>0,608</b>	0,098
143	publisher	3,273	0,195	2	<b>0,105</b>	0,195
144	publisher	2,071	0,212	2	<b>0,38</b>	0,212
145	publisher	3,37	0,034	2	<b>0,992</b>	0,034
146	publisher	2,93	0,132	2	<b>0,152</b>	0,132
147	publisher	2,91	0,08	3	<b>0,64</b>	0,08
148	publisher	2,28	0,078	2	<b>0,532</b>	0,078
149	publisher	2,62	0,065	2	<b>0,438</b>	0,065
150	publisher	3,143	0,096	3	<b>0,934</b>	0,096

Tabelle B.33: Einzelwerte der Produktivität der Verlage für die Topics 126-150 (CLEF Jahrgang 2005). Statistisch signifikante Werte sind **fett** gedruckt.

B Ermittelte Werte der Power-Law-Analyse

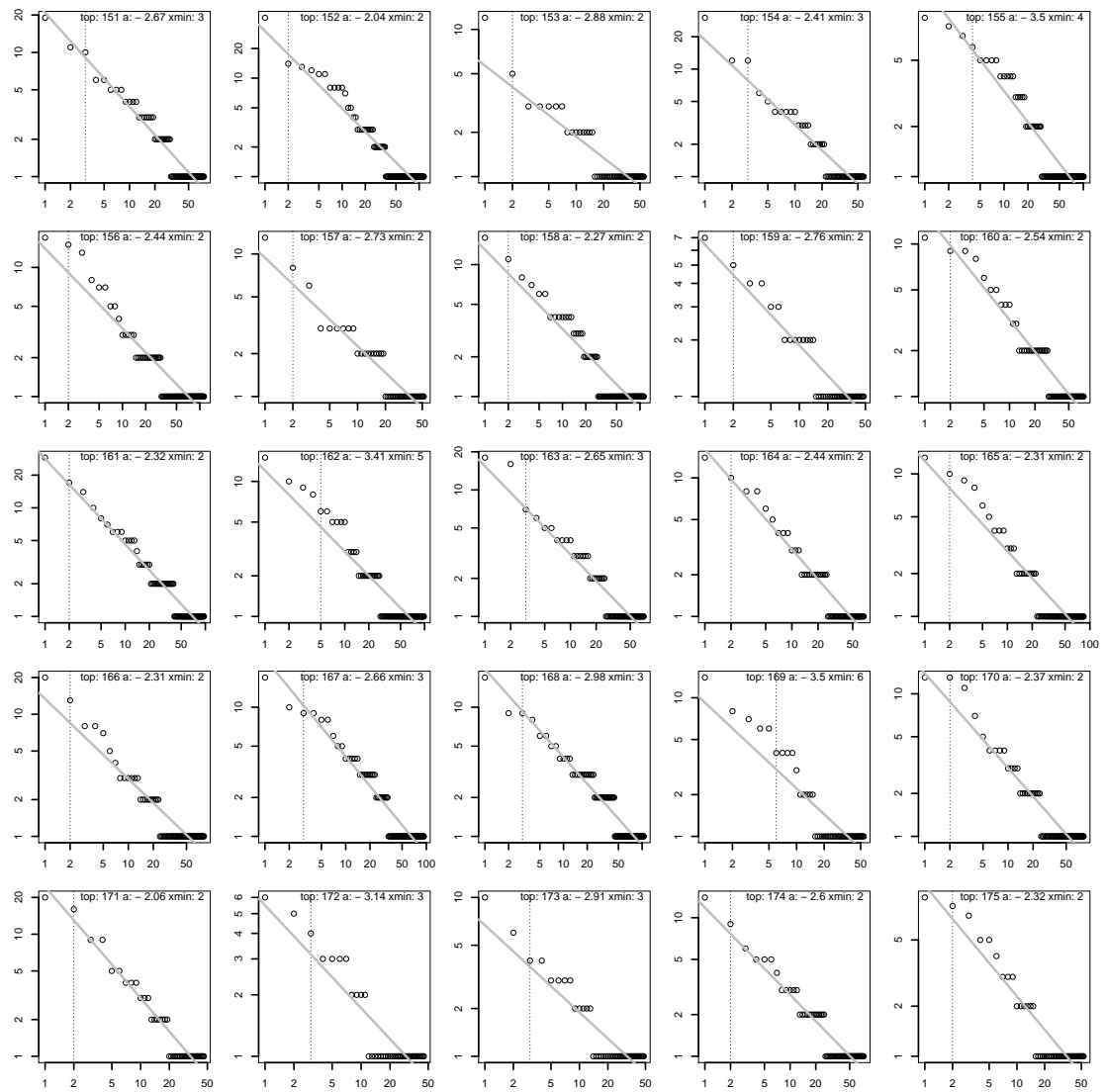


Abbildung B.34: Doppelt logarithmierter Plot der Produktivität der Verlage für die Topics 151-175 (CLEF Jahrgang 2006).

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
151	publisher	2,67	0,054	3	<b>0,911</b>	0,054
152	publisher	2,04	0,094	2	<b>0,309</b>	0,094
153	publisher	2,88	0,084	2	<b>0,657</b>	0,084
154	publisher	2,41	0,096	3	<b>0,544</b>	0,096
155	publisher	3,5	0,084	4	<b>0,69</b>	0,084
156	publisher	2,44	0,066	2	<b>0,622</b>	0,066
157	publisher	2,73	0,087	2	<b>0,502</b>	0,087
158	publisher	2,27	0,11	2	<b>0,209</b>	0,11
159	publisher	2,764	0,071	2	<b>0,835</b>	0,071
160	publisher	2,54	0,072	2	<b>0,556</b>	0,072
161	publisher	2,32	0,08	2	<b>0,311</b>	0,08
162	publisher	3,41	0,105	5	<b>0,606</b>	0,105
163	publisher	2,65	0,074	3	<b>0,741</b>	0,074
164	publisher	2,44	0,056	2	<b>0,867</b>	0,056
165	publisher	2,31	0,066	2	<b>0,819</b>	0,066
166	publisher	2,31	0,064	2	<b>0,831</b>	0,064
167	publisher	2,66	0,095	3	<b>0,362</b>	0,095
168	publisher	2,98	0,049	3	<b>0,773</b>	0,049
169	publisher	3,5	0,139	6	<b>0,672</b>	0,139
170	publisher	2,37	0,052	2	<b>0,935</b>	0,052
171	publisher	2,061	0,074	2	<b>0,868</b>	0,074
172	publisher	3,143	0,096	3	<b>0,866</b>	0,096
173	publisher	2,908	0,068	3	<b>0,951</b>	0,068
174	publisher	2,6	0,057	2	<b>0,833</b>	0,057
175	publisher	2,32	0,092	2	<b>0,727</b>	0,092

Tabelle B.34: Einzelwerte der Produktivität der Verlage für die Topics 151-175 (CLEF Jahrgang 2006). Statistisch signifikante Werte sind **fett** gedruckt.

B Ermittelte Werte der Power-Law-Analyse

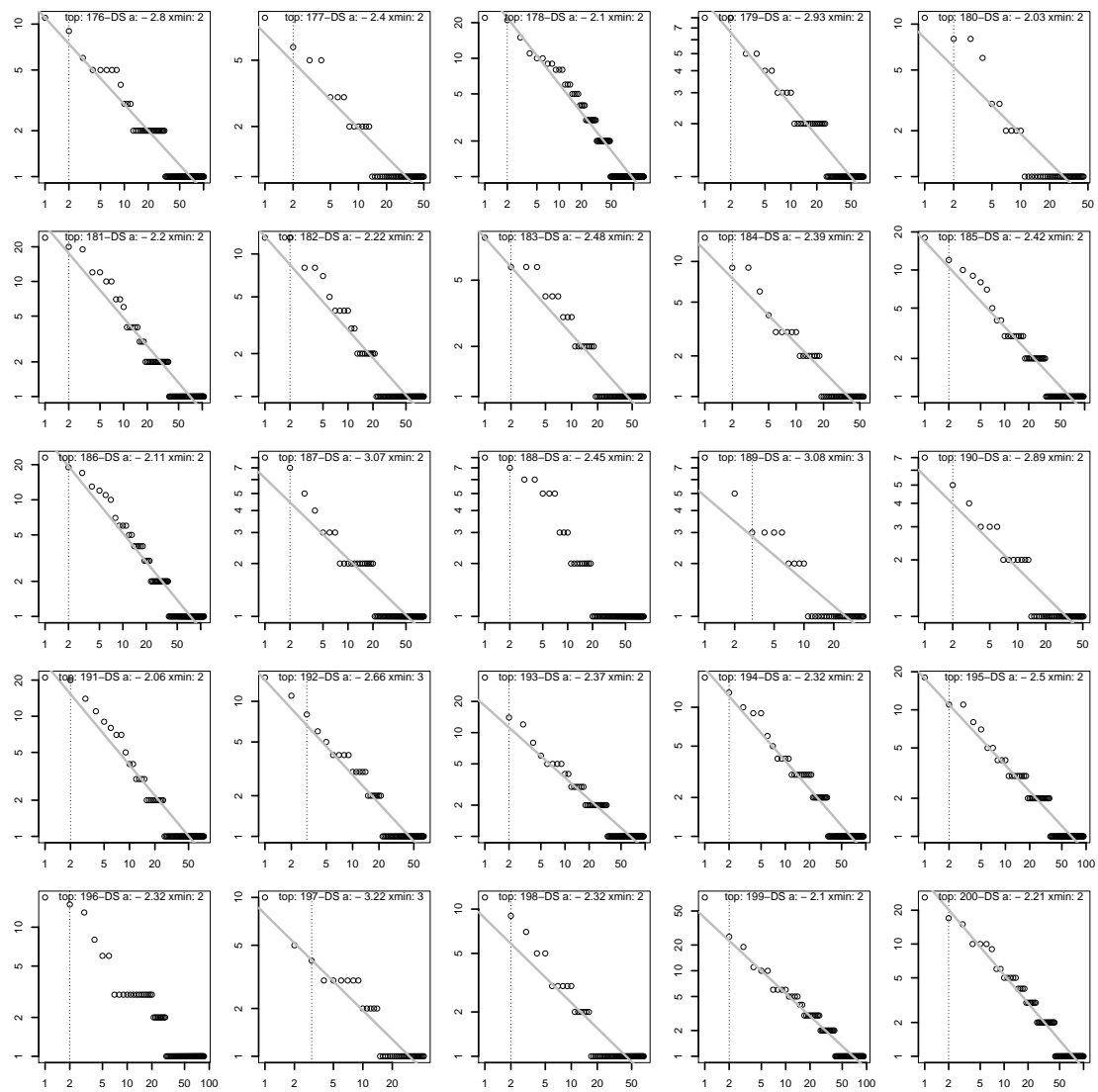


Abbildung B.35: Doppelt logarithmierter Plot der Produktivität der Verlage für die Topics 176-200 (CLEF Jahrgang 2007).

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
10.2452/176-DS	publisher	2,8	0,101	2	<b>0,129</b>	0,101
10.2452/177-DS	publisher	2,403	0,109	2	<b>0,543</b>	0,109
10.2452/178-DS	publisher	2,1	0,066	2	<b>0,52</b>	0,066
10.2452/179-DS	publisher	2,93	0,038	2	<b>0,941</b>	0,038
10.2452/180-DS	publisher	2,026	0,164	2	<b>0,403</b>	0,164
10.2452/181-DS	publisher	2,2	0,075	2	<b>0,469</b>	0,075
10.2452/182-DS	publisher	2,22	0,108	2	<b>0,387</b>	0,108
10.2452/183-DS	publisher	2,48	0,09	2	<b>0,602</b>	0,09
10.2452/184-DS	publisher	2,39	0,072	2	<b>0,816</b>	0,072
10.2452/185-DS	publisher	2,42	0,056	2	<b>0,756</b>	0,056
10.2452/186-DS	publisher	2,11	0,071	2	<b>0,583</b>	0,071
10.2452/187-DS	publisher	3,07	0,047	2	<b>0,892</b>	0,047
10.2452/188-DS	publisher	2,45	0,155	2	0,098	0,155
10.2452/189-DS	publisher	3,083	0,109	3	<b>0,742</b>	0,109
10.2452/190-DS	publisher	2,89	0,06	2	<b>0,93</b>	0,06
10.2452/191-DS	publisher	2,06	0,101	2	<b>0,411</b>	0,101
10.2452/192-DS	publisher	2,66	0,063	3	<b>0,924</b>	0,063
10.2452/193-DS	publisher	2,37	0,04	2	<b>0,954</b>	0,04
10.2452/194-DS	publisher	2,32	0,111	2	<b>0,131</b>	0,111
10.2452/195-DS	publisher	2,5	0,032	2	<b>0,987</b>	0,032
10.2452/196-DS	publisher	2,32	0,164	2	0,02	0,164
10.2452/197-DS	publisher	3,222	0,109	3	<b>0,644</b>	0,109
10.2452/198-DS	publisher	2,32	0,1	2	<b>0,609</b>	0,1
10.2452/199-DS	publisher	2,1	0,066	2	<b>0,602</b>	0,066
10.2452/200-DS	publisher	2,21	0,051	2	<b>0,756</b>	0,051

Tabelle B.35: Einzelwerte der Produktivität der Verlage für die Topics 176-200 (CLEF Jahrgang 2007). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

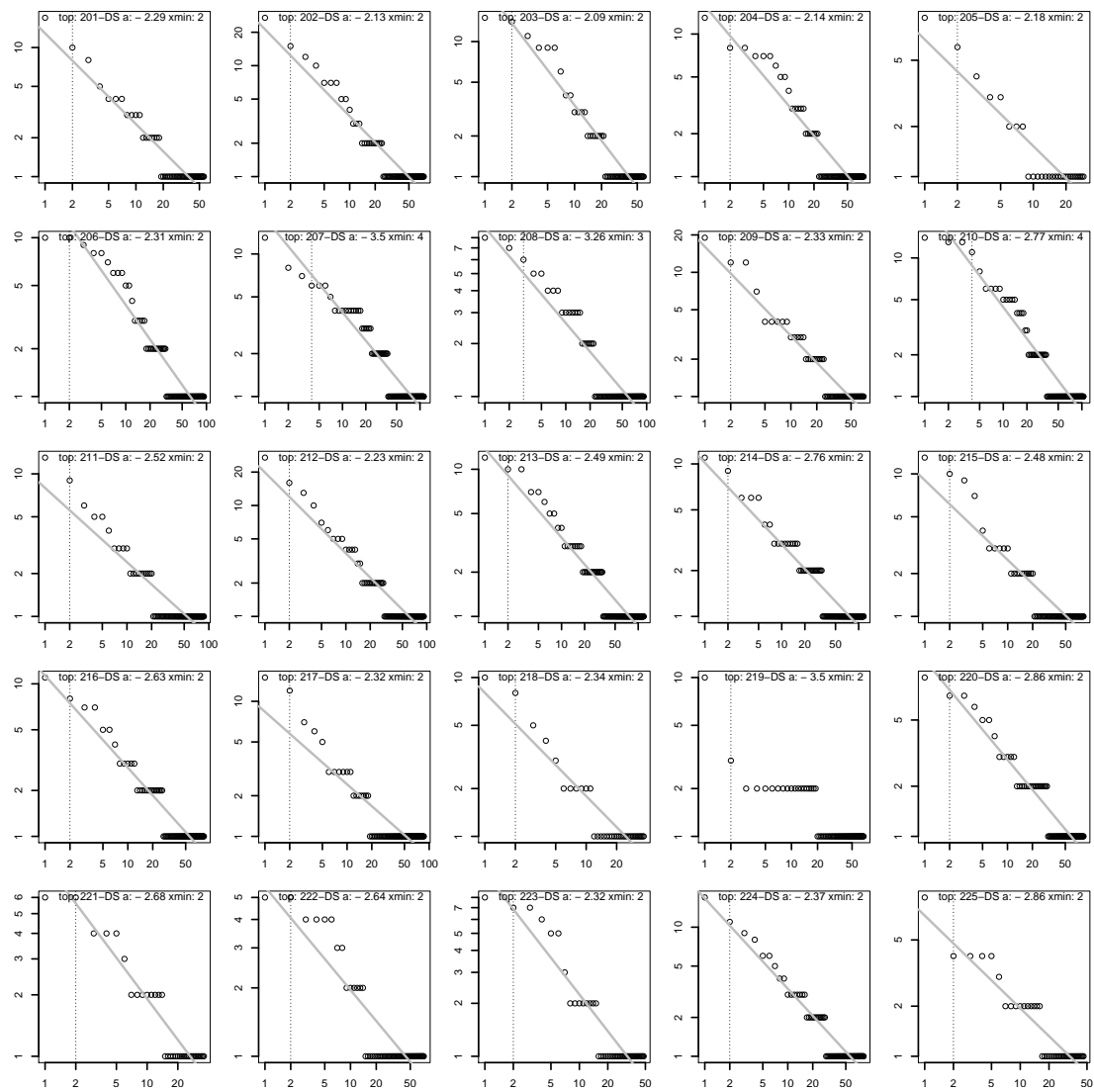


Abbildung B.36: Doppelt logarithmierter Plot der Produktivität der Verlage für die Topics 201-225 (CLEF Jahrgang 2008).



Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
10.2452/201-DS	publisher	2,29	0,078	2	<b>0,801</b>	0,078
10.2452/202-DS	publisher	2,13	0,094	2	<b>0,501</b>	0,094
10.2452/203-DS	publisher	2,09	0,129	2	<b>0,228</b>	0,129
10.2452/204-DS	publisher	2,14	0,113	2	<b>0,325</b>	0,113
10.2452/205-DS	publisher	2,181	0,107	2	<b>0,867</b>	0,107
10.2452/206-DS	publisher	2,31	0,107	2	<b>0,185</b>	0,107
10.2452/207-DS	publisher	3,5	0,128	4	<b>0,178</b>	0,128
10.2452/208-DS	publisher	3,26	0,066	3	<b>0,91</b>	0,066
10.2452/209-DS	publisher	2,33	0,076	2	<b>0,66</b>	0,076
10.2452/210-DS	publisher	2,77	0,133	4	<b>0,103</b>	0,133
10.2452/211-DS	publisher	2,52	0,051	2	<b>0,942</b>	0,051
10.2452/212-DS	publisher	2,23	0,083	2	<b>0,498</b>	0,083
10.2452/213-DS	publisher	2,49	0,045	2	<b>0,882</b>	0,045
10.2452/214-DS	publisher	2,76	0,058	2	<b>0,587</b>	0,058
10.2452/215-DS	publisher	2,48	0,079	2	<b>0,633</b>	0,079
10.2452/216-DS	publisher	2,63	0,056	2	<b>0,766</b>	0,056
10.2452/217-DS	publisher	2,32	0,086	2	<b>0,732</b>	0,086
10.2452/218-DS	publisher	2,336	0,083	2	<b>0,901</b>	0,083
10.2452/219-DS	publisher	3,5	0,197	2	0,03	0,197
10.2452/220-DS	publisher	2,86	0,06	2	<b>0,578</b>	0,06
10.2452/221-DS	publisher	2,681	0,127	2	<b>0,304</b>	0,127
10.2452/222-DS	publisher	2,64	0,165	2	<b>0,123</b>	0,165
10.2452/223-DS	publisher	2,316	0,177	2	<b>0,101</b>	0,177
10.2452/224-DS	publisher	2,37	0,058	2	<b>0,816</b>	0,058
10.2452/225-DS	publisher	2,856	0,111	2	<b>0,338</b>	0,111

Tabelle B.36: Einzelwerte der Produktivität der Verlage für die Topics 201-225 (CLEF Jahrgang 2008). Statistisch signifikante Werte sind **fett** gedruckt.

## B.7 Produktivität des Erscheinungsjahres in GIRT

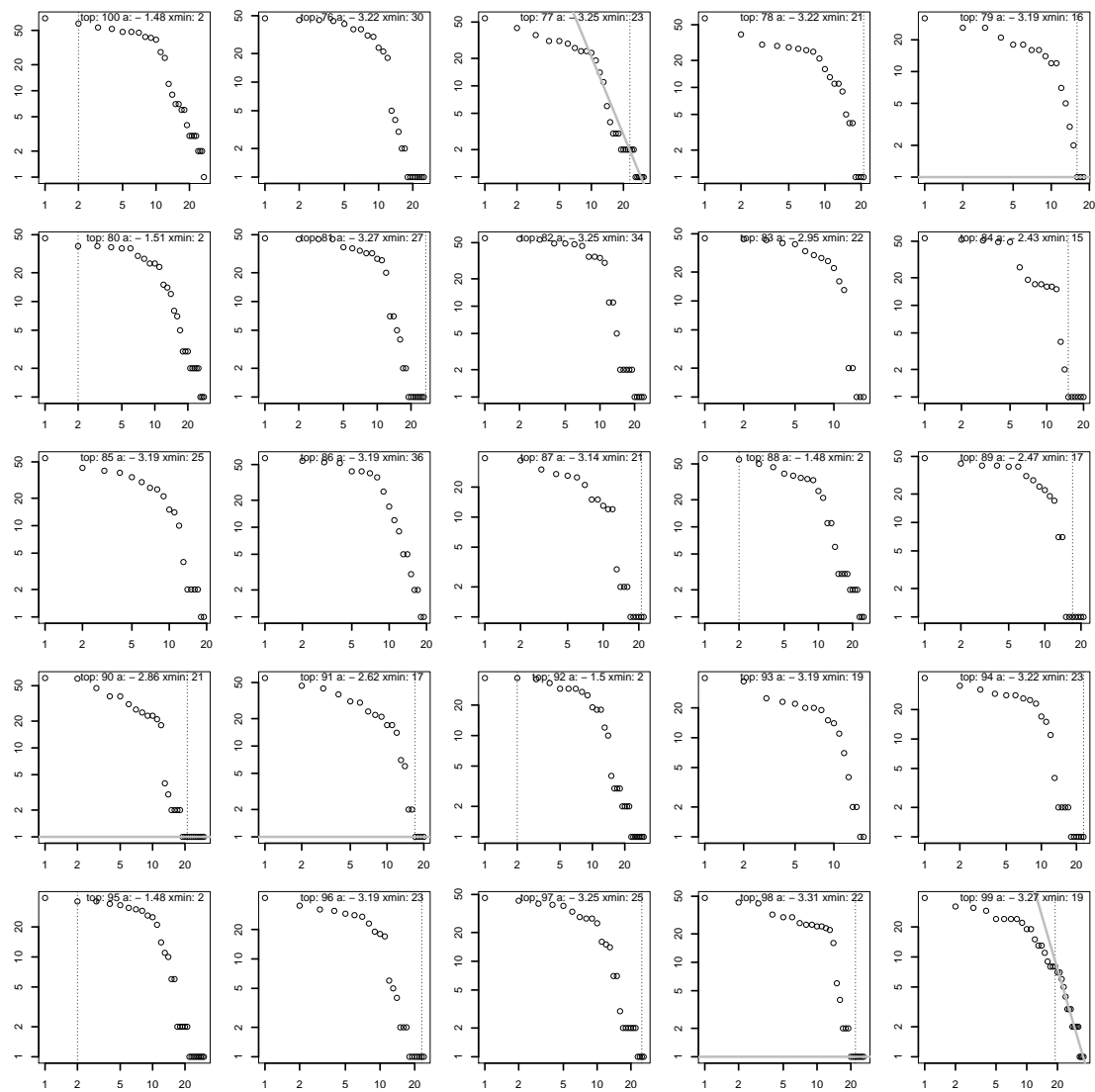


Abbildung B.37: Doppelt logarithmierter Plot der Produktivität des Erscheinungsjahres für die Topics 76-100 (CLEF Jahrgang 2003).

B.7 Produktivität des Erscheinungsjahres in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
76	pubyear	3,222	0,304	30	0,06	0,304
77	pubyear	3,25	0,131	23	<b>0,864</b>	0,131
78	pubyear	3,222	0,248	21	<b>0,111</b>	0,248
79	pubyear	3,188	0,157	16	<b>0,686</b>	0,157
80	pubyear	1,509	0,216	2	0,022	0,216
81	pubyear	3,273	0,245	27	<b>0,148</b>	0,245
82	pubyear	3,25	0,271	34	0,071	0,271
83	pubyear	2,953	0,217	22	<b>0,275</b>	0,217
84	pubyear	2,43	0,265	15	0,046	0,265
85	pubyear	3,188	0,167	25	<b>0,742</b>	0,167
86	pubyear	3,188	0,275	36	0,081	0,275
87	pubyear	3,143	0,217	21	<b>0,371</b>	0,217
88	pubyear	1,477	0,226	2	0,031	0,226
89	pubyear	2,467	0,242	17	0,085	0,242
90	pubyear	2,864	0,164	21	<b>0,535</b>	0,164
91	pubyear	2,618	0,19	17	<b>0,298</b>	0,19
92	pubyear	1,496	0,263	2	0,005	0,263
93	pubyear	3,188	0,199	19	<b>0,466</b>	0,199
94	pubyear	3,222	0,209	23	<b>0,395</b>	0,209
95	pubyear	1,476	0,264	2	0,01	0,264
96	pubyear	3,188	0,211	23	<b>0,317</b>	0,211
97	pubyear	3,25	0,202	25	<b>0,366</b>	0,202
98	pubyear	3,308	0,132	22	<b>0,776</b>	0,132
99	pubyear	3,273	0,177	19	<b>0,211</b>	0,177
100	pubyear	1,481	0,206	2	0,034	0,206

Tabelle B.37: Einzelwerte der Produktivität des Erscheinungsjahres für die Topics 76-100 (CLEF Jahrgang 2003). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

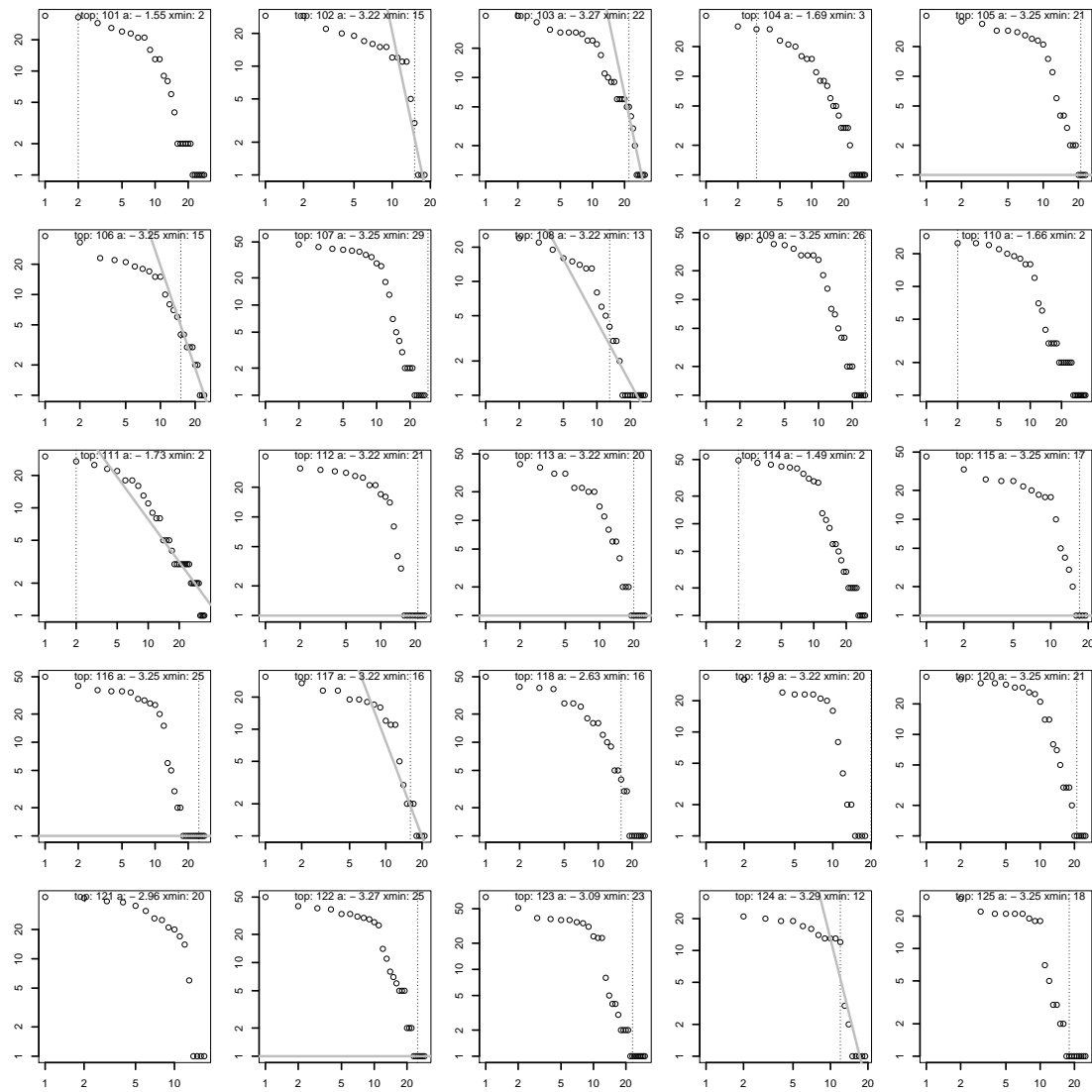


Abbildung B.38: Doppelt logarithmierter Plot der Produktivität des Erscheinungsjahres für die Topics 101-125 (CLEF Jahrgang 2004).

B.7 Produktivität des Erscheinungsjahres in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
101	pubyear	1,552	0,227	2	0,018	0,227
102	pubyear	3,222	0,17	15	<b>0,503</b>	0,17
103	pubyear	3,273	0,187	22	<b>0,283</b>	0,187
104	pubyear	1,687	0,18	3	0,032	0,18
105	pubyear	3,25	0,172	21	<b>0,585</b>	0,172
106	pubyear	3,25	0,099	15	<b>0,974</b>	0,099
107	pubyear	3,25	0,232	29	<b>0,126</b>	0,232
108	pubyear	3,222	0,169	13	<b>0,602</b>	0,169
109	pubyear	3,25	0,223	26	<b>0,21</b>	0,223
110	pubyear	1,662	0,196	2	0,016	0,196
111	pubyear	1,726	0,115	2	<b>0,203</b>	0,115
112	pubyear	3,222	0,231	21	<b>0,223</b>	0,231
113	pubyear	3,222	0,229	20	<b>0,195</b>	0,229
114	pubyear	1,49	0,21	2	0,029	0,21
115	pubyear	3,25	0,128	17	<b>0,919</b>	0,128
116	pubyear	3,25	0,185	25	<b>0,508</b>	0,185
117	pubyear	3,222	0,17	16	<b>0,633</b>	0,17
118	pubyear	2,629	0,229	16	0,08	0,229
119	pubyear	3,222	0,24	20	<b>0,216</b>	0,24
120	pubyear	3,25	0,261	21	0,078	0,261
121	pubyear	2,962	0,223	20	<b>0,179</b>	0,223
122	pubyear	3,273	0,194	25	<b>0,348</b>	0,194
123	pubyear	3,09	0,263	23	0,042	0,263
124	pubyear	3,292	0,126	12	<b>0,807</b>	0,126
125	pubyear	3,25	0,334	18	0,016	0,334

Tabelle B.38: Einzelwerte der Produktivität des Erscheinungsjahres für die Topics 101-125 (CLEF Jahrgang 2004). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

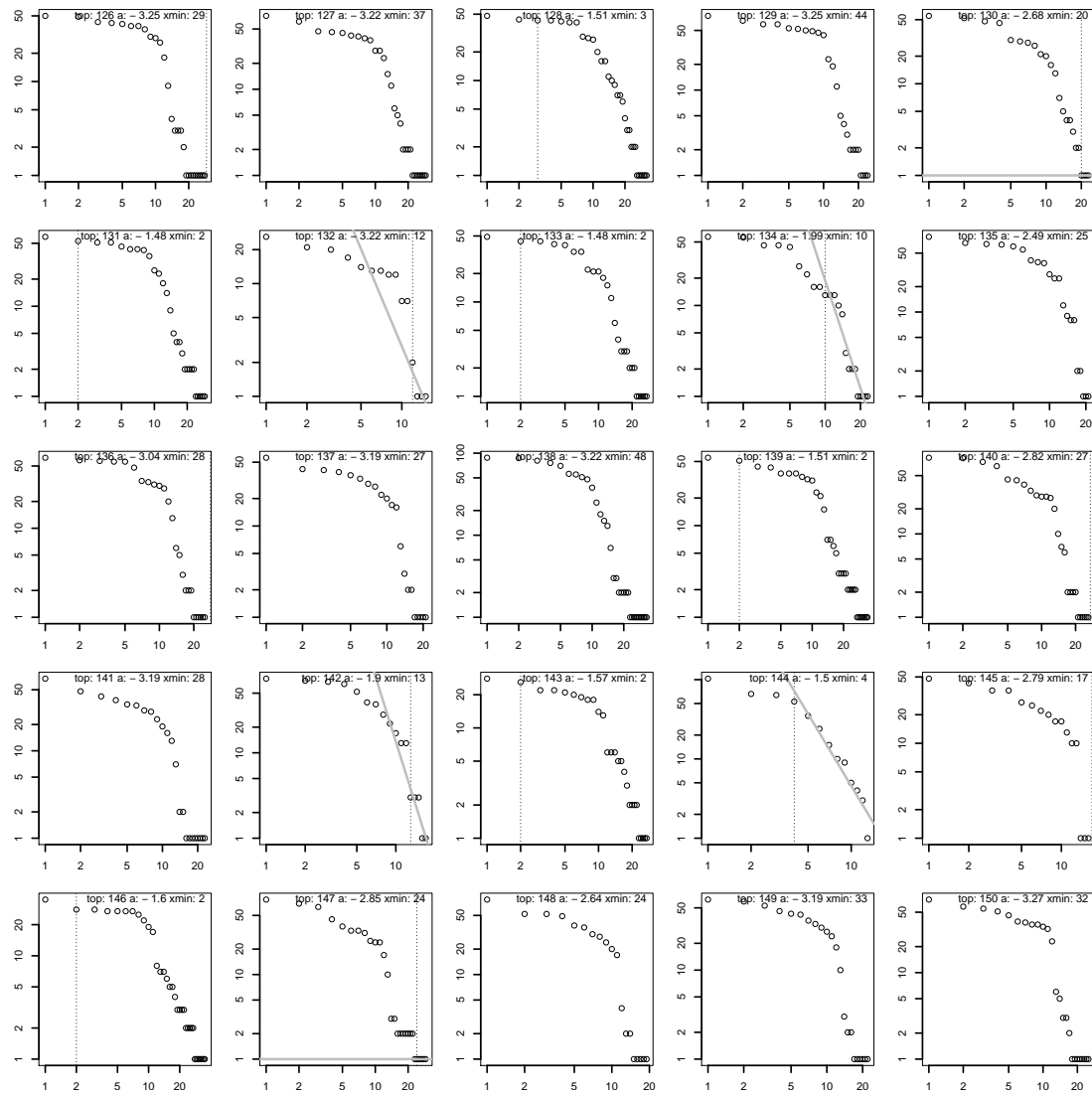


Abbildung B.39: Doppelt logarithmierter Plot der Produktivität des Erscheinungsjahres für die Topics 126-150 (CLEF Jahrgang 2005).

B.7 Produktivität des Erscheinungsjahres in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
126	pubyear	3,25	0,239	29	<b>0,16</b>	0,239
127	pubyear	3,222	0,295	37	0,025	0,295
128	pubyear	1,515	0,219	3	0,016	0,219
129	pubyear	3,25	0,261	44	<b>0,129</b>	0,261
130	pubyear	2,683	0,195	20	<b>0,293</b>	0,195
131	pubyear	1,478	0,228	2	0,023	0,228
132	pubyear	3,222	0,124	12	<b>0,932</b>	0,124
133	pubyear	1,476	0,246	2	0,015	0,246
134	pubyear	1,988	0,188	10	<b>0,158</b>	0,188
135	pubyear	2,494	0,25	25	0,068	0,25
136	pubyear	3,045	0,253	28	0,093	0,253
137	pubyear	3,188	0,182	27	<b>0,57</b>	0,182
138	pubyear	3,222	0,211	48	<b>0,361</b>	0,211
139	pubyear	1,51	0,208	2	0,014	0,208
140	pubyear	2,824	0,163	27	<b>0,56</b>	0,163
141	pubyear	3,188	0,117	28	<b>0,982</b>	0,117
142	pubyear	1,898	0,233	13	<b>0,102</b>	0,233
143	pubyear	1,573	0,215	2	0,017	0,215
144	pubyear	1,5	0,203	4	<b>0,319</b>	0,203
145	pubyear	2,791	0,182	17	<b>0,466</b>	0,182
146	pubyear	1,597	0,177	2	0,023	0,177
147	pubyear	2,845	0,138	24	<b>0,831</b>	0,138
148	pubyear	2,644	0,2	24	<b>0,367</b>	0,2
149	pubyear	3,188	0,207	33	<b>0,337</b>	0,207
150	pubyear	3,273	0,133	32	<b>0,892</b>	0,133

Tabelle B.39: Einzelwerte der Produktivität des Erscheinungsjahres für die Topics 126-150 (CLEF Jahrgang 2005). Statistisch signifikante Werte sind **fett** gedruckt.

B Ermittelte Werte der Power-Law-Analyse

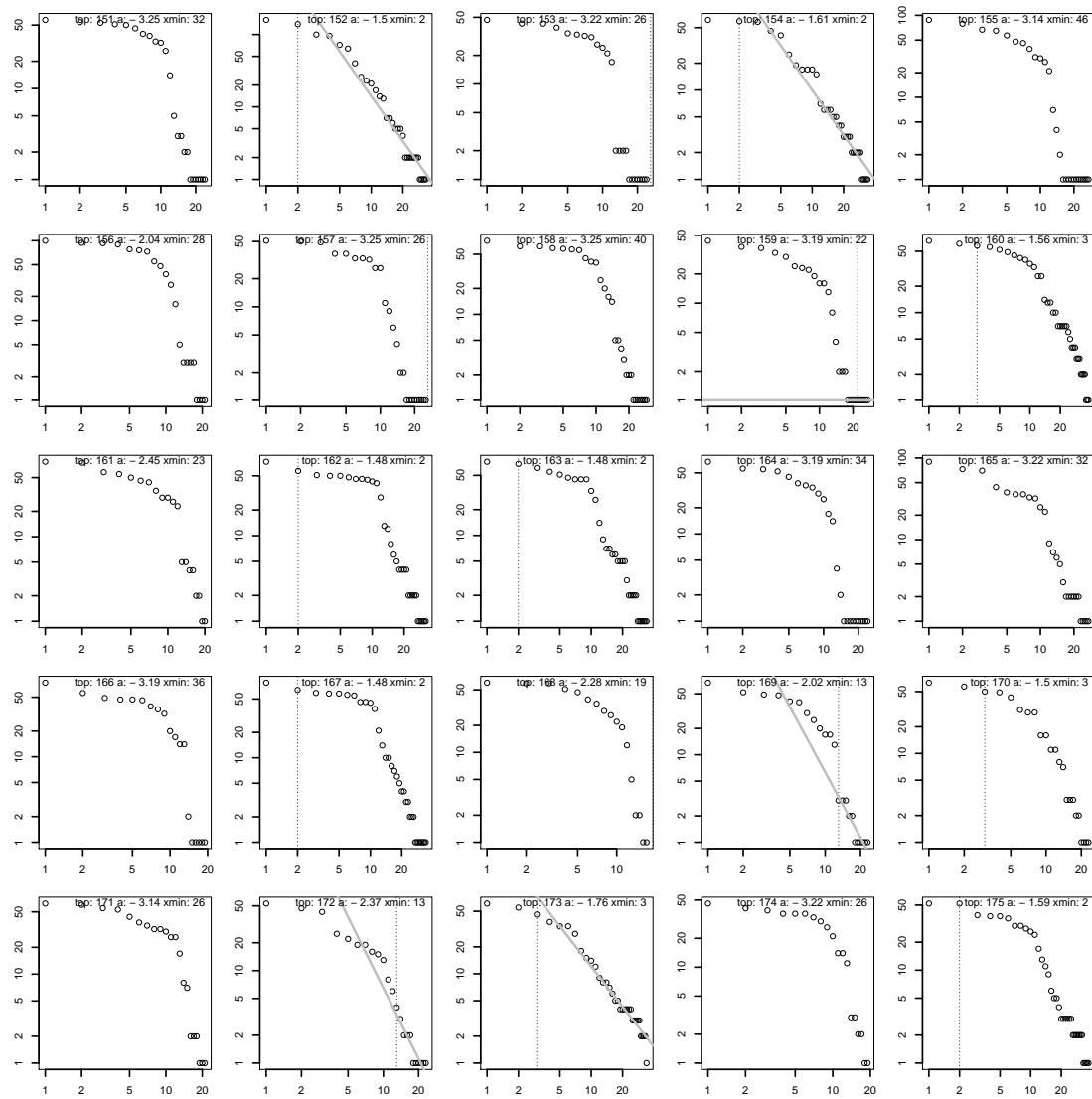


Abbildung B.40: Doppelt logarithmierter Plot der Produktivität des Erscheinungsjahres für die Topics 151-175 (CLEF Jahrgang 2006).



B.7 Produktivität des Erscheinungsjahres in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
151	pubyear	3,25	0,222	32	<b>0,291</b>	0,222
152	pubyear	1,501	0,128	2	<b>0,351</b>	0,128
153	pubyear	3,222	0,25	26	<b>0,144</b>	0,25
154	pubyear	1,608	0,136	2	<b>0,129</b>	0,136
155	pubyear	3,143	0,19	46	<b>0,622</b>	0,19
156	pubyear	2,045	0,308	28	0,007	0,308
157	pubyear	3,25	0,21	26	<b>0,418</b>	0,21
158	pubyear	3,25	0,273	40	0,073	0,273
159	pubyear	3,188	0,171	22	<b>0,583</b>	0,171
160	pubyear	1,561	0,165	3	0,013	0,165
161	pubyear	2,448	0,23	23	<b>0,131</b>	0,23
162	pubyear	1,481	0,228	2	0,01	0,228
163	pubyear	1,481	0,194	2	0,058	0,194
164	pubyear	3,188	0,174	34	<b>0,734</b>	0,174
165	pubyear	3,222	0,195	32	<b>0,402</b>	0,195
166	pubyear	3,188	0,212	36	<b>0,463</b>	0,212
167	pubyear	1,481	0,221	2	0,022	0,221
168	pubyear	2,282	0,221	19	<b>0,168</b>	0,221
169	pubyear	2,018	0,221	13	<b>0,136</b>	0,221
170	pubyear	1,499	0,236	3	0,038	0,236
171	pubyear	3,136	0,147	26	<b>0,741</b>	0,147
172	pubyear	2,368	0,149	13	<b>0,729</b>	0,149
173	pubyear	1,763	0,09	3	<b>0,562</b>	0,09
174	pubyear	3,222	0,232	26	<b>0,233</b>	0,232
175	pubyear	1,591	0,154	2	0,034	0,154

Tabelle B.40: Einzelwerte der Produktivität des Erscheinungsjahres für die Topics 151-175 (CLEF Jahrgang 2006). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

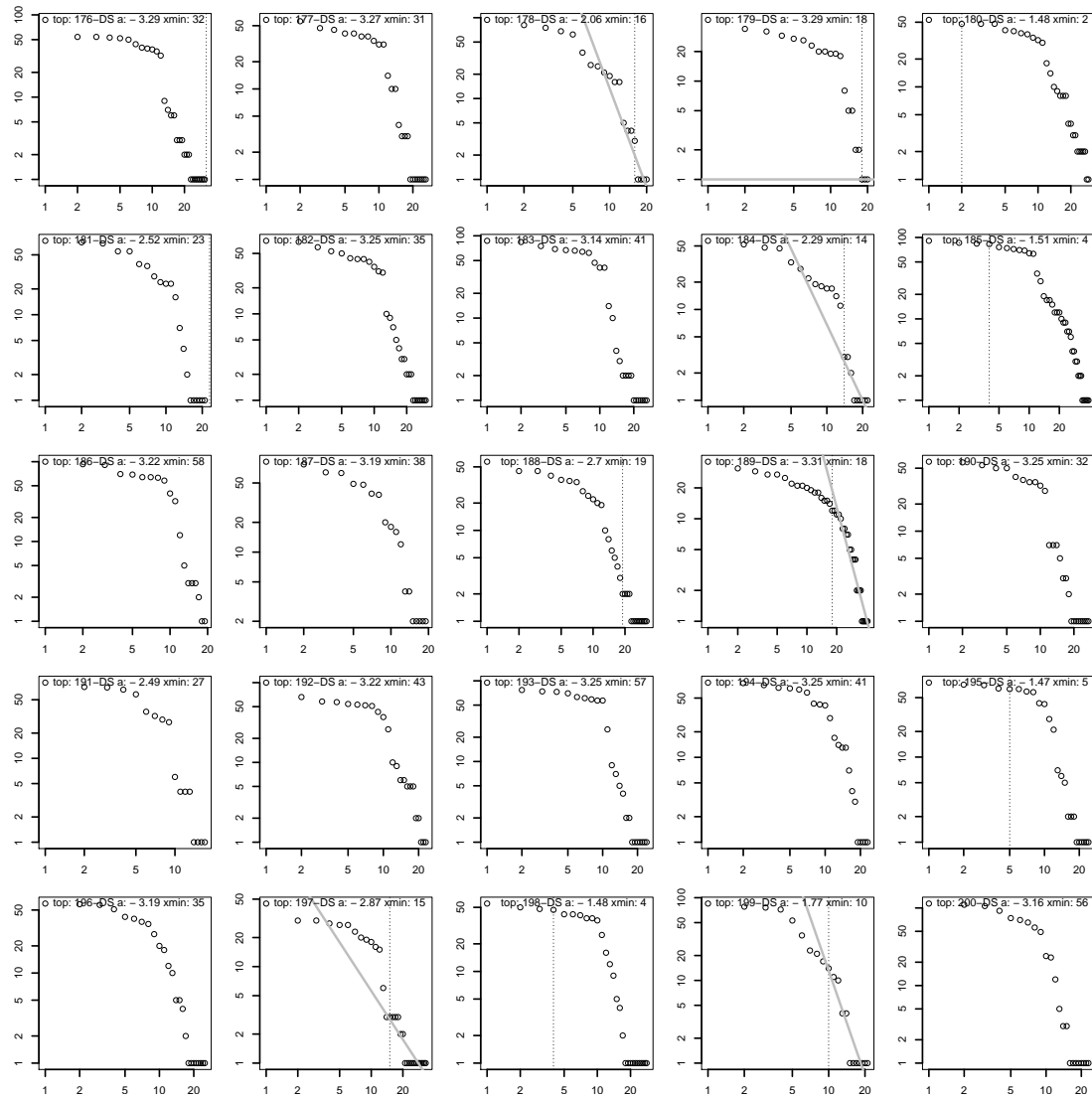


Abbildung B.41: Doppelt logarithmierter Plot der Produktivität des Erscheinungsjahres für die Topics 176-200 (CLEF Jahrgang 2007).

B.7 Produktivität des Erscheinungsjahres in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
10.2452/176-DS	pubyear	3,292	0,187	32	<b>0,404</b>	0,187
10.2452/177-DS	pubyear	3,273	0,196	31	<b>0,437</b>	0,196
10.2452/178-DS	pubyear	2,063	0,214	16	<b>0,188</b>	0,214
10.2452/179-DS	pubyear	3,292	0,11	18	<b>0,931</b>	0,11
10.2452/180-DS	pubyear	1,481	0,214	2	0,021	0,214
10.2452/181-DS	pubyear	2,518	0,226	23	<b>0,15</b>	0,226
10.2452/182-DS	pubyear	3,25	0,206	35	<b>0,262</b>	0,206
10.2452/183-DS	pubyear	3,136	0,353	41	0,006	0,353
10.2452/184-DS	pubyear	2,292	0,163	14	<b>0,467</b>	0,163
10.2452/185-DS	pubyear	1,511	0,195	4	0,012	0,195
10.2452/186-DS	pubyear	3,222	0,267	58	<b>0,174</b>	0,267
10.2452/187-DS	pubyear	3,188	0,221	38	<b>0,369</b>	0,221
10.2452/188-DS	pubyear	2,696	0,25	19	0,026	0,25
10.2452/189-DS	pubyear	3,308	0,159	18	<b>0,177</b>	0,159
10.2452/190-DS	pubyear	3,25	0,204	32	<b>0,338</b>	0,204
10.2452/191-DS	pubyear	2,493	0,283	27	<b>0,114</b>	0,283
10.2452/192-DS	pubyear	3,222	0,239	43	<b>0,146</b>	0,239
10.2452/193-DS	pubyear	3,25	0,354	57	0,025	0,354
10.2452/194-DS	pubyear	3,25	0,284	41	0,061	0,284
10.2452/195-DS	pubyear	1,467	0,337	5	0,004	0,337
10.2452/196-DS	pubyear	3,188	0,256	35	<b>0,158</b>	0,256
10.2452/197-DS	pubyear	2,87	0,207	15	<b>0,18</b>	0,207
10.2452/198-DS	pubyear	1,478	0,317	4	0,005	0,317
10.2452/199-DS	pubyear	1,77	0,179	10	<b>0,415</b>	0,179
10.2452/200-DS	pubyear	3,161	0,209	56	<b>0,542</b>	0,209

Tabelle B.41: Einzelwerte der Produktivität des Erscheinungsjahres für die Topics 176-200 (CLEF Jahrgang 2007). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

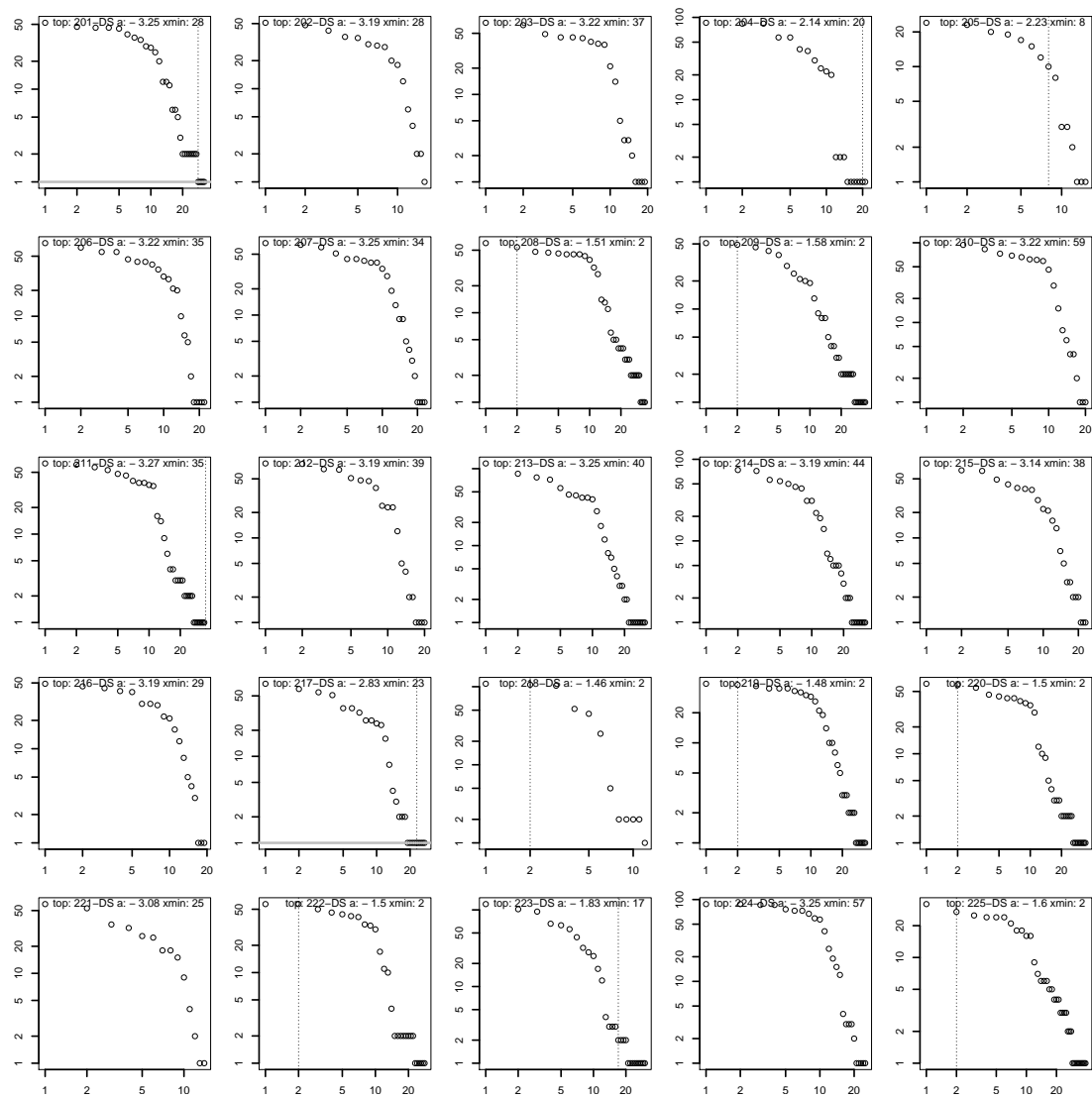


Abbildung B.42: Doppelt logarithmierter Plot der Produktivität des Erscheinungsjahres für die Topics 201-225 (CLEF Jahrgang 2008).

B.7 Produktivität des Erscheinungsjahres in GIRT

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
10.2452/201-DS	pubyear	3,25	0,2	28	<b>0,215</b>	0,2
10.2452/202-DS	pubyear	3,188	0,208	28	<b>0,516</b>	0,208
10.2452/203-DS	pubyear	3,222	0,245	37	<b>0,305</b>	0,245
10.2452/204-DS	pubyear	2,136	0,209	20	<b>0,303</b>	0,209
10.2452/205-DS	pubyear	2,227	0,263	8	0,069	0,263
10.2452/206-DS	pubyear	3,222	0,184	35	<b>0,561</b>	0,184
10.2452/207-DS	pubyear	3,25	0,238	34	<b>0,173</b>	0,238
10.2452/208-DS	pubyear	1,512	0,179	2	0,032	0,179
10.2452/209-DS	pubyear	1,577	0,16	2	0,083	0,16
10.2452/210-DS	pubyear	3,222	0,265	59	<b>0,175</b>	0,265
10.2452/211-DS	pubyear	3,273	0,218	35	<b>0,129</b>	0,218
10.2452/212-DS	pubyear	3,188	0,251	39	<b>0,336</b>	0,251
10.2452/213-DS	pubyear	3,25	0,173	40	<b>0,644</b>	0,173
10.2452/214-DS	pubyear	3,188	0,165	44	<b>0,722</b>	0,165
10.2452/215-DS	pubyear	3,143	0,168	38	<b>0,759</b>	0,168
10.2452/216-DS	pubyear	3,188	0,252	29	<b>0,149</b>	0,252
10.2452/217-DS	pubyear	2,827	0,167	23	<b>0,615</b>	0,167
10.2452/218-DS	pubyear	1,455	0,295	2	0,07	0,295
10.2452/219-DS	pubyear	1,481	0,213	2	0,032	0,213
10.2452/220-DS	pubyear	1,499	0,221	2	0,019	0,221
10.2452/221-DS	pubyear	3,083	0,184	25	<b>0,818</b>	0,184
10.2452/222-DS	pubyear	1,496	0,239	2	0,016	0,239
10.2452/223-DS	pubyear	1,827	0,222	17	0,097	0,222
10.2452/224-DS	pubyear	3,25	0,326	57	0,014	0,326
10.2452/225-DS	pubyear	1,598	0,157	2	0,046	0,157

Tabelle B.42: Einzelwerte der Produktivität des Erscheinungsjahres für die Topics 201-225 (CLEF Jahrgang 2008). Statistisch signifikante Werte sind **fett** gedruckt.

## B.8 Produktivität der Thesaurusterme in GIRT

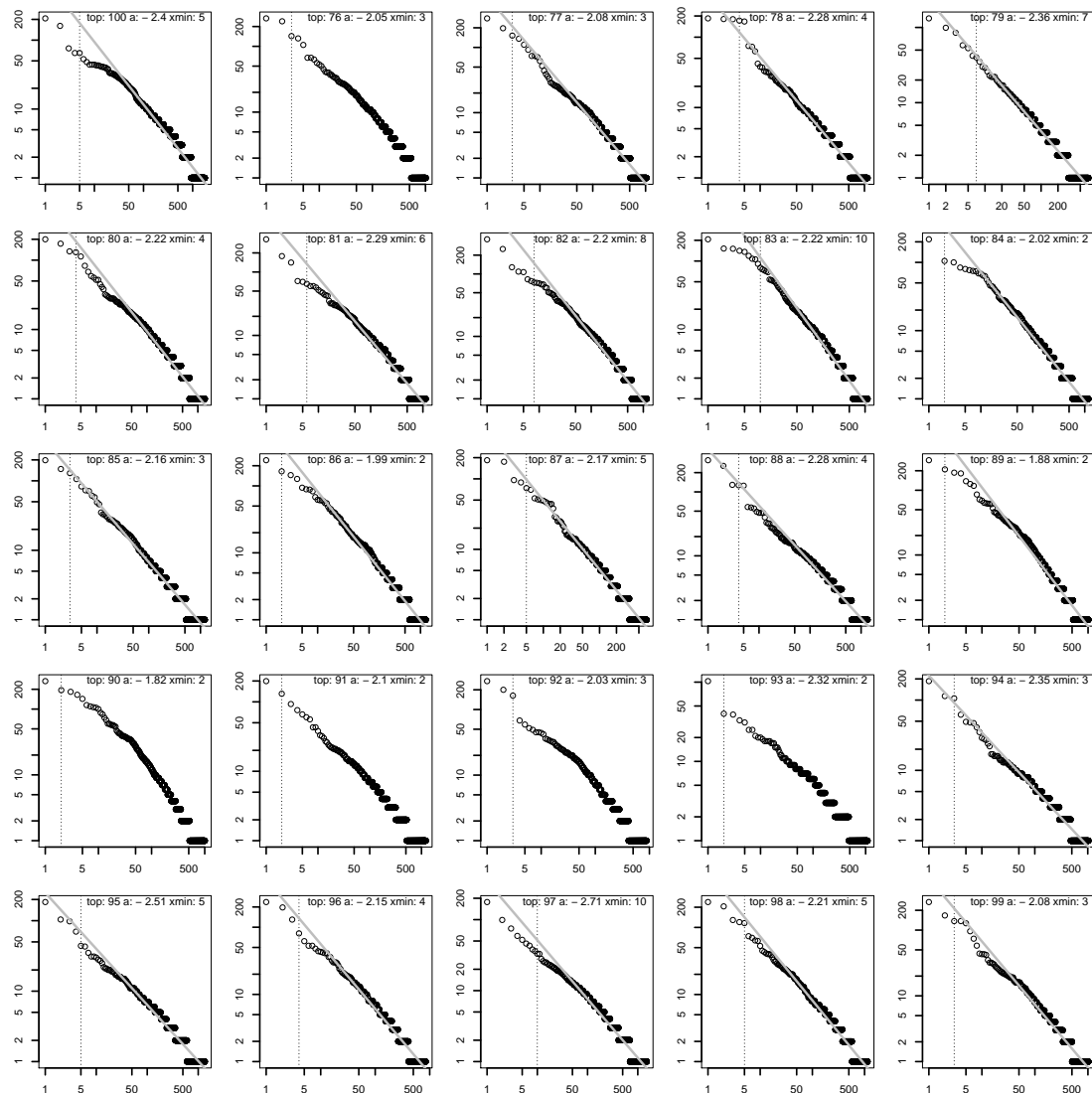


Abbildung B.43: Doppelt logarithmierter Plot der Produktivität der Thesaurusterme für die Topics 76-100 (CLEF Jahrgang 2003).

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
76	subject	2,05	0,037	3	0,082	0,037
77	subject	2,08	0,03	3	<b>0,25</b>	0,03
78	subject	2,28	0,031	4	<b>0,422</b>	0,031
79	subject	2,36	0,026	7	<b>0,725</b>	0,026
80	subject	2,22	0,03	4	<b>0,232</b>	0,03
81	subject	2,29	0,035	6	<b>0,558</b>	0,035
82	subject	2,2	0,036	8	<b>0,382</b>	0,036
83	subject	2,22	0,036	10	<b>0,514</b>	0,036
84	subject	2,02	0,023	2	<b>0,24</b>	0,023
85	subject	2,16	0,016	3	<b>0,873</b>	0,016
86	subject	1,99	0,023	2	<b>0,25</b>	0,023
87	subject	2,17	0,037	5	<b>0,401</b>	0,037
88	subject	2,28	0,028	4	<b>0,362</b>	0,028
89	subject	1,88	0,027	2	<b>0,205</b>	0,027
90	subject	1,82	0,039	2	0,014	0,039
91	subject	2,1	0,03	2	0,073	0,03
92	subject	2,03	0,04	3	0,096	0,04
93	subject	2,32	0,035	2	0,019	0,035
94	subject	2,35	0,031	3	<b>0,321</b>	0,031
95	subject	2,51	0,023	5	<b>0,848</b>	0,023
96	subject	2,15	0,038	4	<b>0,29</b>	0,038
97	subject	2,71	0,027	10	<b>0,755</b>	0,027
98	subject	2,21	0,035	5	<b>0,215</b>	0,035
99	subject	2,08	0,036	3	<b>0,126</b>	0,036
100	subject	2,4	0,034	5	<b>0,16</b>	0,034

Tabelle B.43: Einzelwerte der Produktivität der Thesaurusterme für die Topics 76-100 (CLEF Jahrgang 2003). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

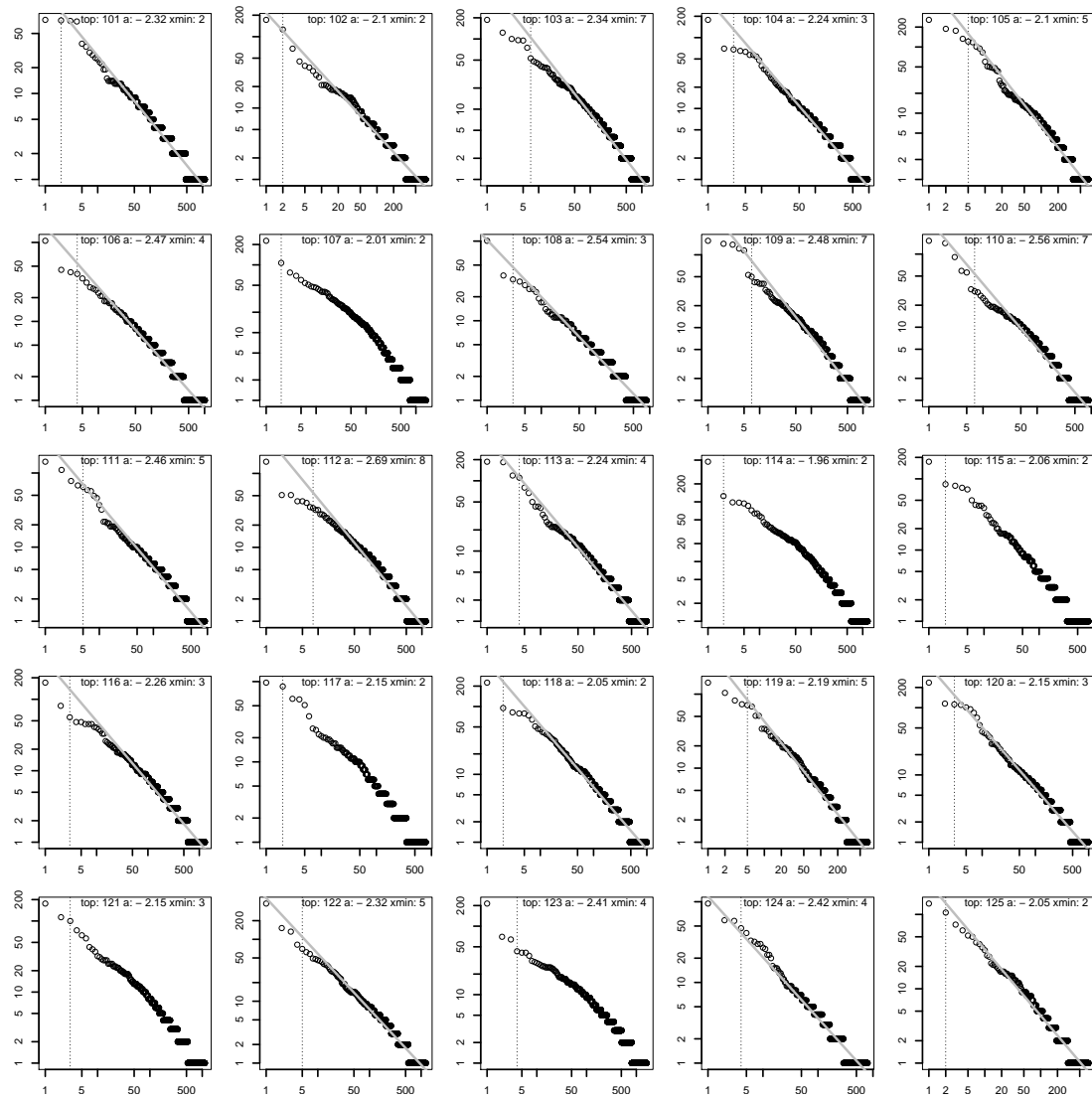


Abbildung B.44: Doppelt logarithmierter Plot der Produktivität der Thesaurusterme für die Topics 101-125 (CLEF Jahrgang 2004).



Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
101	subject	2,32	0,022	2	<b>0,343</b>	0,022
102	subject	2,1	0,033	2	<b>0,162</b>	0,033
103	subject	2,34	0,039	7	<b>0,232</b>	0,039
104	subject	2,24	0,026	3	<b>0,532</b>	0,026
105	subject	2,1	0,048	5	<b>0,122</b>	0,048
106	subject	2,47	0,027	4	<b>0,726</b>	0,027
107	subject	2,01	0,039	2	0	0,039
108	subject	2,54	0,024	3	<b>0,736</b>	0,024
109	subject	2,48	0,034	7	<b>0,625</b>	0,034
110	subject	2,56	0,047	7	<b>0,371</b>	0,047
111	subject	2,46	0,027	5	<b>0,845</b>	0,027
112	subject	2,69	0,043	8	<b>0,43</b>	0,043
113	subject	2,24	0,036	4	<b>0,27</b>	0,036
114	subject	1,96	0,03	2	0,071	0,03
115	subject	2,06	0,038	2	0,097	0,038
116	subject	2,26	0,033	3	<b>0,128</b>	0,033
117	subject	2,15	0,04	2	0,038	0,04
118	subject	2,05	0,026	2	<b>0,231</b>	0,026
119	subject	2,19	0,053	5	<b>0,118</b>	0,053
120	subject	2,15	0,032	3	<b>0,299</b>	0,032
121	subject	2,15	0,038	3	0,056	0,038
122	subject	2,32	0,022	5	<b>0,907</b>	0,022
123	subject	2,41	0,044	4	0,012	0,044
124	subject	2,42	0,032	4	<b>0,647</b>	0,032
125	subject	2,05	0,031	2	<b>0,256</b>	0,031

Tabelle B.44: Einzelwerte der Produktivität der Thesaurusterme für die Topics 101-125 (CLEF Jahrgang 2004). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

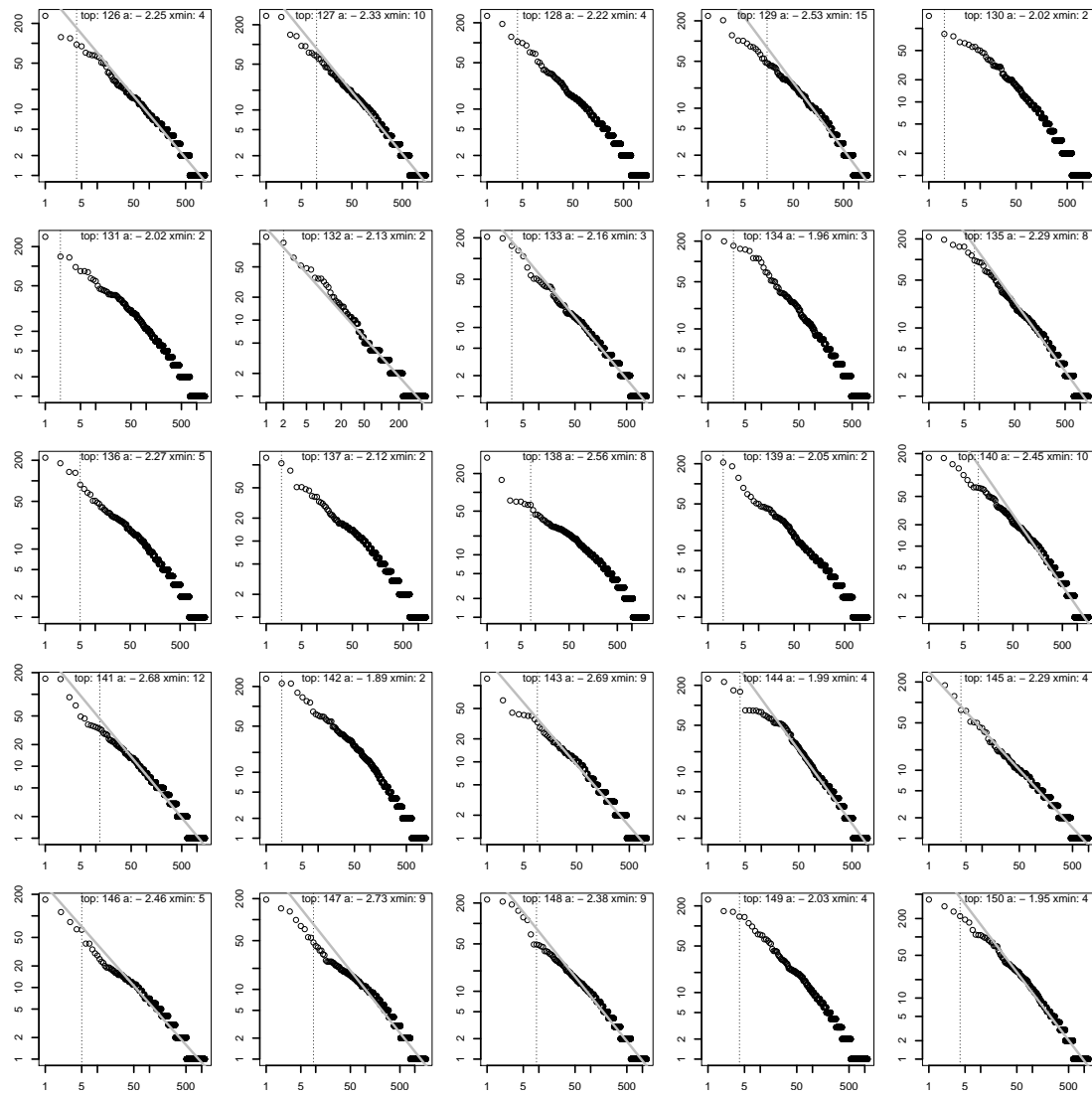


Abbildung B.45: Doppelt logarithmierter Plot der Produktivität der Thesaurusterme für die Topics 126-150 (CLEF Jahrgang 2005).

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
126	subject	2,25	0,022	4	<b>0,784</b>	0,022
127	subject	2,33	0,026	10	<b>0,788</b>	0,026
128	subject	2,22	0,034	4	0,093	0,034
129	subject	2,53	0,041	15	<b>0,544</b>	0,041
130	subject	2,02	0,035	2	0,014	0,035
131	subject	2,02	0,03	2	0,013	0,03
132	subject	2,13	0,033	2	<b>0,343</b>	0,033
133	subject	2,16	0,026	3	<b>0,36</b>	0,026
134	subject	1,96	0,04	3	0,054	0,04
135	subject	2,29	0,035	8	<b>0,493</b>	0,035
136	subject	2,27	0,048	5	0,017	0,048
137	subject	2,12	0,041	2	0,001	0,041
138	subject	2,56	0,05	8	0,023	0,05
139	subject	2,05	0,037	2	0,004	0,037
140	subject	2,45	0,034	10	<b>0,614</b>	0,034
141	subject	2,68	0,036	12	<b>0,66</b>	0,036
142	subject	1,89	0,036	2	0,01	0,036
143	subject	2,69	0,032	9	<b>0,516</b>	0,032
144	subject	1,99	0,032	4	<b>0,361</b>	0,032
145	subject	2,29	0,028	4	<b>0,472</b>	0,028
146	subject	2,46	0,044	5	<b>0,101</b>	0,044
147	subject	2,73	0,039	9	<b>0,405</b>	0,039
148	subject	2,38	0,027	9	<b>0,805</b>	0,027
149	subject	2,03	0,04	4	0,068	0,04
150	subject	1,95	0,034	4	<b>0,174</b>	0,034

Tabelle B.45: Einzelwerte der Produktivität der Thesaurusterme für die Topics 126-150 (CLEF Jahrgang 2005). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

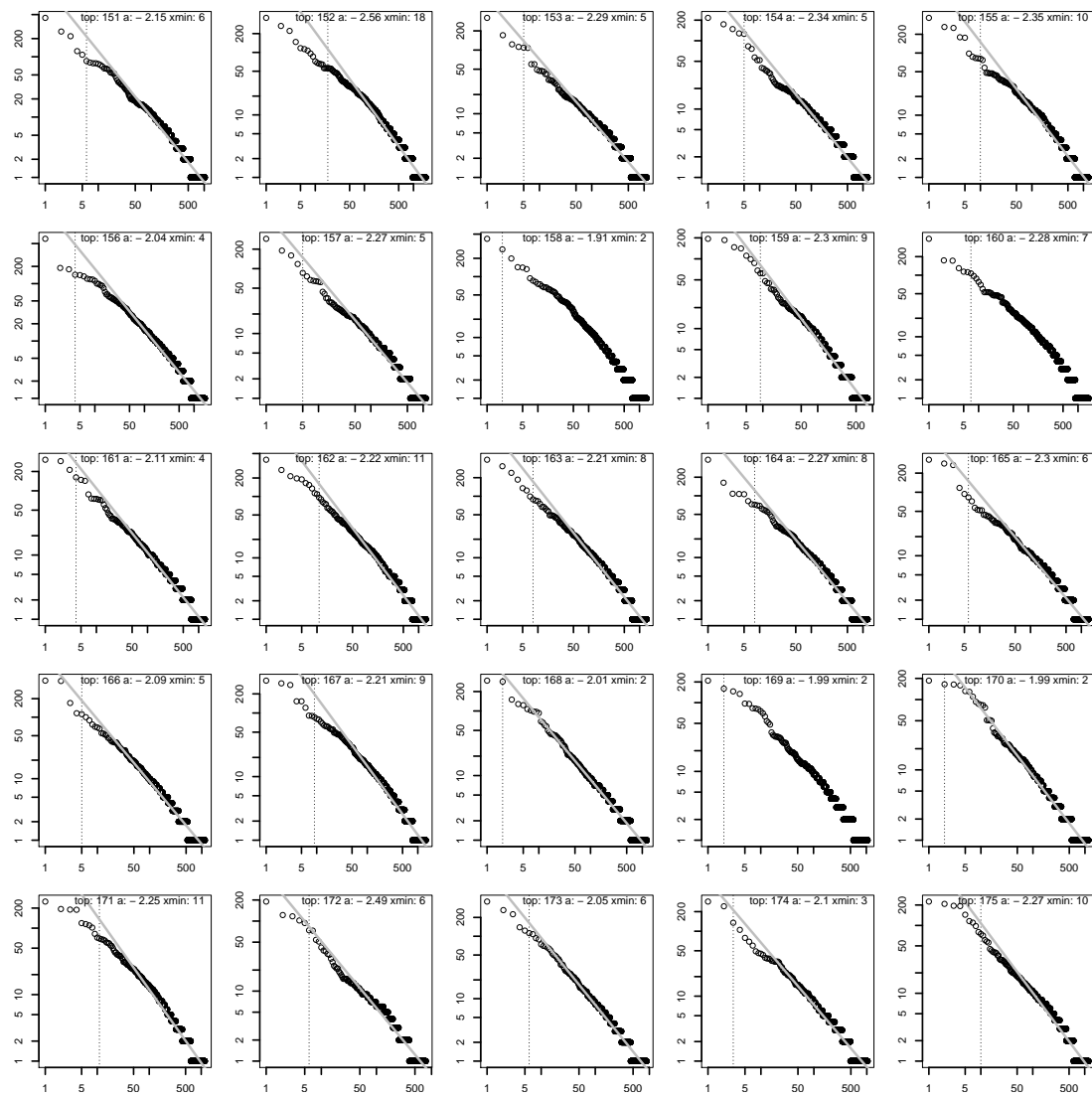


Abbildung B.46: Doppelt logarithmierter Plot der Produktivität der Thesaurusterme für die Topics 151-175 (CLEF Jahrgang 2006).

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
151	subject	2,15	0,044	6	<b>0,179</b>	0,044
152	subject	2,56	0,038	18	<b>0,547</b>	0,038
153	subject	2,29	0,022	5	<b>0,922</b>	0,022
154	subject	2,34	0,03	5	<b>0,525</b>	0,03
155	subject	2,35	0,051	10	<b>0,163</b>	0,051
156	subject	2,04	0,031	4	<b>0,132</b>	0,031
157	subject	2,27	0,035	5	<b>0,425</b>	0,035
158	subject	1,91	0,032	2	0,015	0,032
159	subject	2,3	0,037	9	<b>0,54</b>	0,037
160	subject	2,28	0,048	7	0,028	0,048
161	subject	2,11	0,027	4	<b>0,329</b>	0,027
162	subject	2,22	0,03	11	<b>0,783</b>	0,03
163	subject	2,21	0,033	8	<b>0,422</b>	0,033
164	subject	2,27	0,042	8	<b>0,225</b>	0,042
165	subject	2,3	0,029	6	<b>0,628</b>	0,029
166	subject	2,09	0,032	5	<b>0,487</b>	0,032
167	subject	2,21	0,036	9	<b>0,366</b>	0,036
168	subject	2,01	0,028	2	<b>0,114</b>	0,028
169	subject	1,99	0,029	2	0,063	0,029
170	subject	1,99	0,025	2	<b>0,226</b>	0,025
171	subject	2,25	0,037	11	<b>0,435</b>	0,037
172	subject	2,49	0,049	6	<b>0,251</b>	0,049
173	subject	2,05	0,025	6	<b>0,76</b>	0,025
174	subject	2,1	0,028	3	<b>0,447</b>	0,028
175	subject	2,27	0,021	10	<b>0,966</b>	0,021

Tabelle B.46: Einzelwerte der Produktivität der Thesaurusterme für die Topics 151-175 (CLEF Jahrgang 2006). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

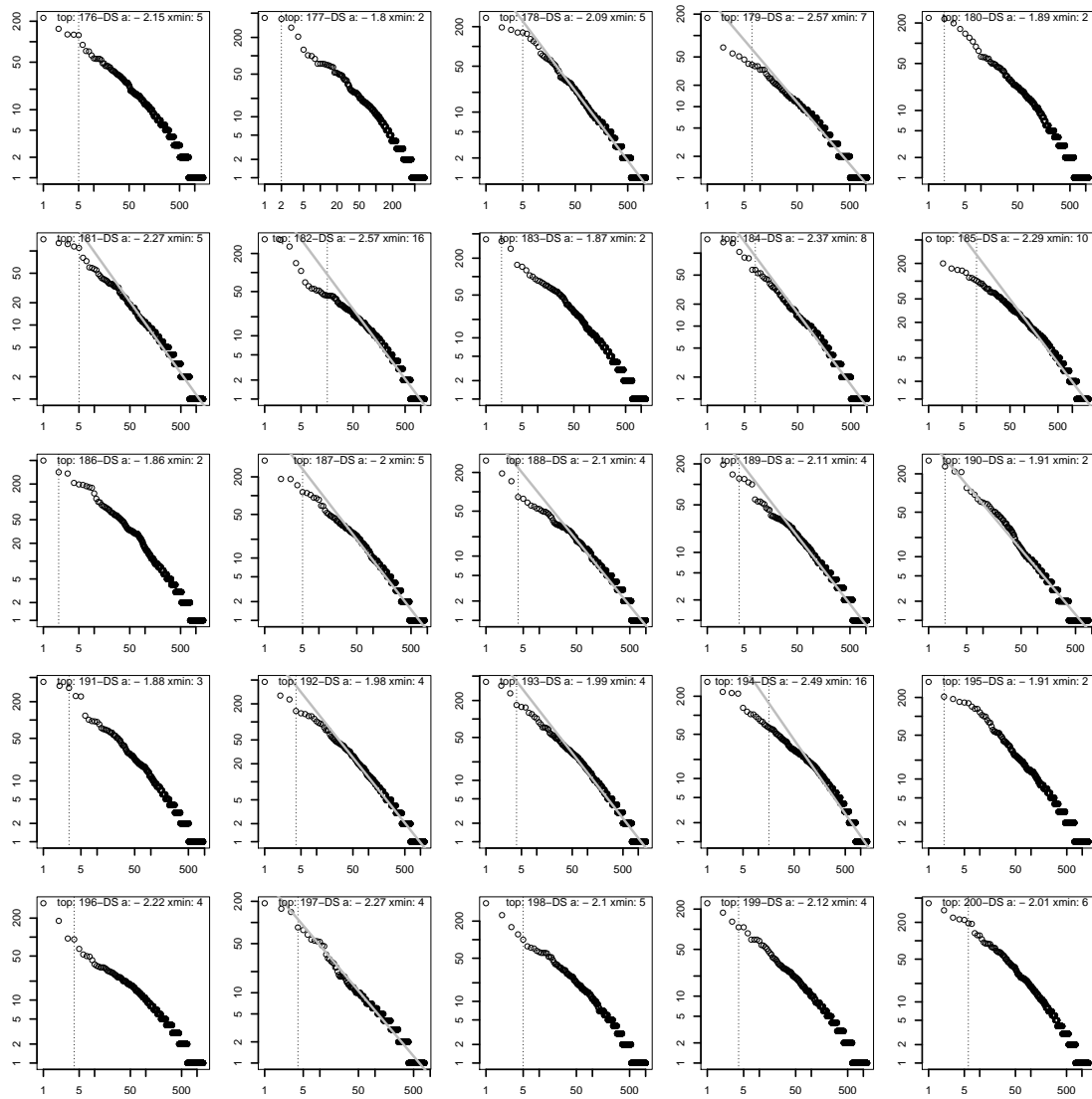


Abbildung B.47: Doppelt logarithmierter Plot der Produktivität der Thesaurusterme für die Topics 176-200 (CLEF Jahrgang 2007).

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
10.2452/176-DS	subject	2,15	0,042	5	0,04	0,042
10.2452/177-DS	subject	1,8	0,037	2	0,051	0,037
10.2452/178-DS	subject	2,09	0,028	5	<b>0,637</b>	0,028
10.2452/179-DS	subject	2,57	0,023	7	<b>0,959</b>	0,023
10.2452/180-DS	subject	1,89	0,034	2	0,014	0,034
10.2452/181-DS	subject	2,27	0,029	5	<b>0,46</b>	0,029
10.2452/182-DS	subject	2,57	0,051	16	<b>0,316</b>	0,051
10.2452/183-DS	subject	1,87	0,036	2	0,009	0,036
10.2452/184-DS	subject	2,37	0,034	8	<b>0,692</b>	0,034
10.2452/185-DS	subject	2,29	0,04	10	<b>0,12</b>	0,04
10.2452/186-DS	subject	1,86	0,027	2	0,063	0,027
10.2452/187-DS	subject	2	0,042	5	<b>0,171</b>	0,042
10.2452/188-DS	subject	2,1	0,04	4	<b>0,111</b>	0,04
10.2452/189-DS	subject	2,11	0,026	4	<b>0,495</b>	0,026
10.2452/190-DS	subject	1,91	0,019	2	<b>0,821</b>	0,019
10.2452/191-DS	subject	1,88	0,04	3	0,026	0,04
10.2452/192-DS	subject	1,98	0,031	4	<b>0,207</b>	0,031
10.2452/193-DS	subject	1,99	0,033	4	<b>0,103</b>	0,033
10.2452/194-DS	subject	2,49	0,039	16	<b>0,605</b>	0,039
10.2452/195-DS	subject	1,91	0,029	2	0,043	0,029
10.2452/196-DS	subject	2,22	0,038	4	0,096	0,038
10.2452/197-DS	subject	2,27	0,024	4	<b>0,865</b>	0,024
10.2452/198-DS	subject	2,1	0,052	5	0,019	0,052
10.2452/199-DS	subject	2,12	0,036	4	0,098	0,036
10.2452/200-DS	subject	2,01	0,044	6	0,051	0,044

Tabelle B.47: Einzelwerte der Produktivität der Thesaurusterme für die Topics 176-200 (CLEF Jahrgang 2007). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

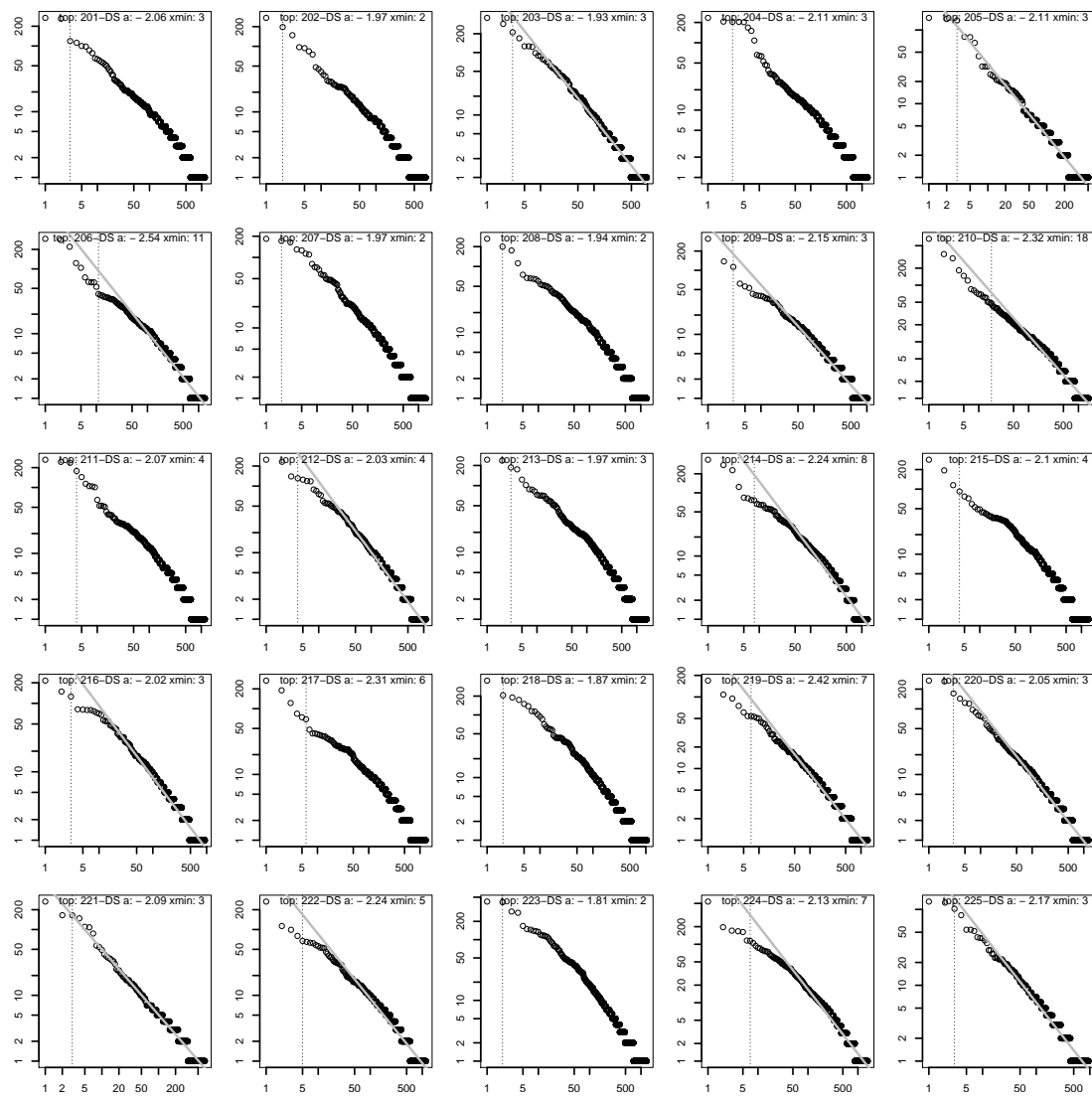


Abbildung B.48: Doppelt logarithmierter Plot der Produktivität der Thesaurusterme für die Topics 201-225 (CLEF Jahrgang 2008).



Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
10.2452/201-DS	subject	2,06	0,044	3	0,009	0,044
10.2452/202-DS	subject	1,97	0,044	2	0,009	0,044
10.2452/203-DS	subject	1,93	0,026	3	<b>0,402</b>	0,026
10.2452/204-DS	subject	2,11	0,036	3	0,047	0,036
10.2452/205-DS	subject	2,11	0,029	3	<b>0,818</b>	0,029
10.2452/206-DS	subject	2,54	0,042	11	<b>0,498</b>	0,042
10.2452/207-DS	subject	1,97	0,037	2	0,002	0,037
10.2452/208-DS	subject	1,94	0,037	2	0,006	0,037
10.2452/209-DS	subject	2,15	0,026	3	<b>0,449</b>	0,026
10.2452/210-DS	subject	2,32	0,03	18	<b>0,864</b>	0,03
10.2452/211-DS	subject	2,07	0,043	4	0,011	0,043
10.2452/212-DS	subject	2,03	0,03	4	<b>0,372</b>	0,03
10.2452/213-DS	subject	1,97	0,044	3	0	0,044
10.2452/214-DS	subject	2,24	0,034	8	<b>0,469</b>	0,034
10.2452/215-DS	subject	2,1	0,056	4	0	0,056
10.2452/216-DS	subject	2,02	0,027	3	<b>0,476</b>	0,027
10.2452/217-DS	subject	2,31	0,049	6	0,05	0,049
10.2452/218-DS	subject	1,87	0,032	2	0,062	0,032
10.2452/219-DS	subject	2,42	0,034	7	<b>0,387</b>	0,034
10.2452/220-DS	subject	2,05	0,032	3	<b>0,15</b>	0,032
10.2452/221-DS	subject	2,09	0,031	3	<b>0,514</b>	0,031
10.2452/222-DS	subject	2,24	0,026	5	<b>0,715</b>	0,026
10.2452/223-DS	subject	1,81	0,031	2	0,029	0,031
10.2452/224-DS	subject	2,13	0,033	7	0,351	0,033
10.2452/225-DS	subject	2,17	0,026	3	0,607	0,026

Tabelle B.48: Einzelwerte der Produktivität der Thesaurusterme für die Topics 201-225 (CLEF Jahrgang 2008). Statistisch signifikante Werte sind **fett** gedruckt.

## B.9 Produktivität der Autorennamen in iSearch

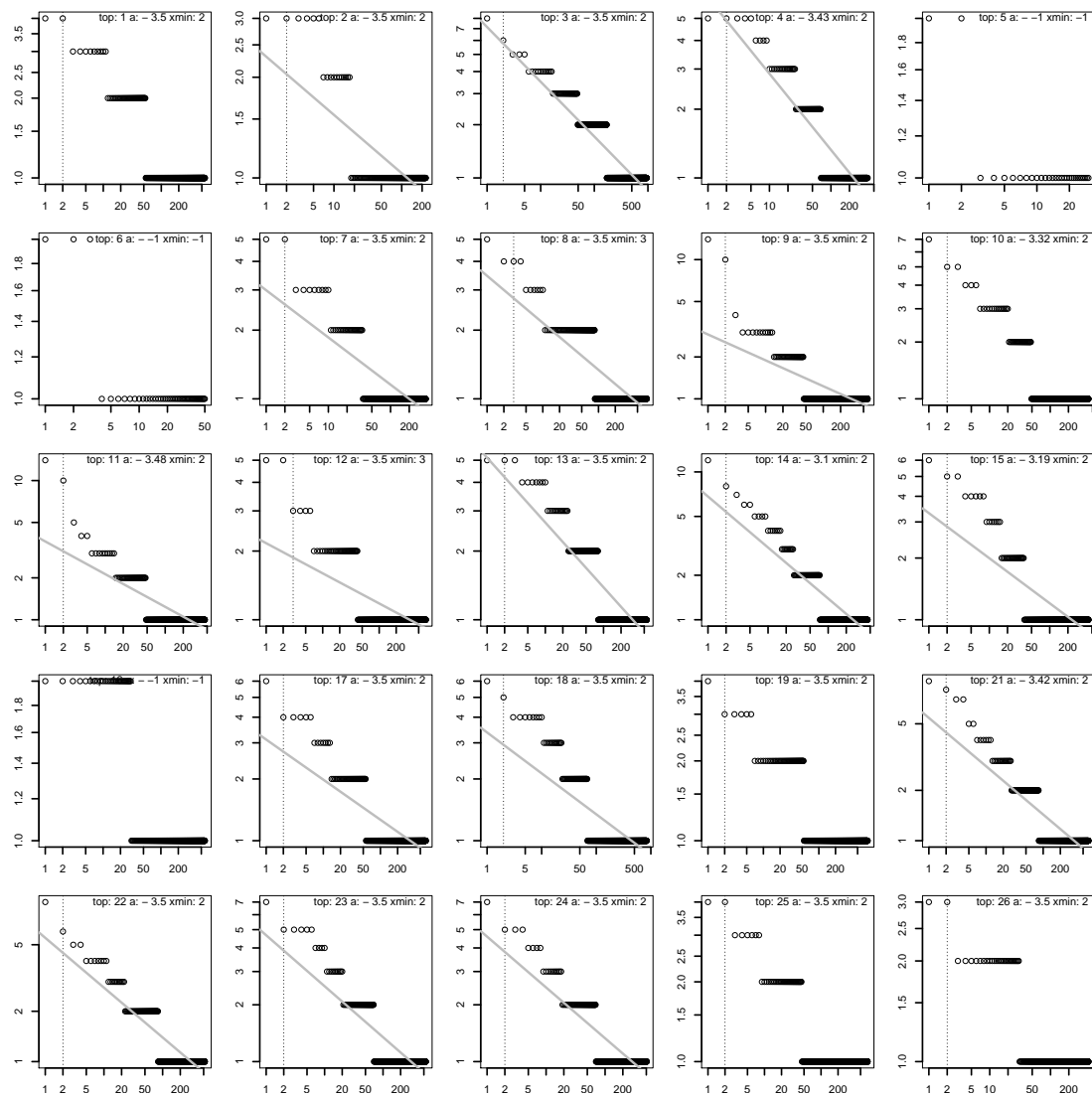


Abbildung B.49: Doppelt logarithmierter Plot der Produktivität der Autorennamen für die iSearch Topics 001-026 (das Topic 20 fehlt systembedingt).

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
1	author	3,5	0,096	2	0,078	0,096
2	author	3,5	0,134	2	<b>0,203</b>	0,134
3	author	3,5	0,043	2	<b>0,143</b>	0,043
4	author	3,43	0,07	2	<b>0,123</b>	0,07
5	author	-1	-1	-1	-1	-1
6	author	-1	-1	-1	-1	-1
7	author	3,5	0,077	2	<b>0,32</b>	0,077
8	author	3,5	0,145	3	<b>0,502</b>	0,145
9	author	3,5	0,067	2	<b>0,329</b>	0,067
10	author	3,32	0,088	2	0,099	0,088
11	author	3,48	0,031	2	<b>0,836</b>	0,031
12	author	3,5	0,145	3	<b>0,552</b>	0,145
13	author	3,5	0,044	2	<b>0,392</b>	0,044
14	author	3,1	0,039	2	<b>0,496</b>	0,039
15	author	3,19	0,056	2	<b>0,476</b>	0,056
16	author	-1	-1	-1	-1	-1
17	author	3,5	0,066	2	<b>0,277</b>	0,066
18	author	3,5	0,043	2	<b>0,525</b>	0,043
19	author	3,5	0,189	2	0	0,189
21	author	3,42	0,016	2	<b>0,956</b>	0,016
22	author	3,5	0,038	2	<b>0,488</b>	0,038
23	author	3,5	0,029	2	<b>0,74</b>	0,029
24	author	3,5	0,042	2	<b>0,505</b>	0,042
25	author	3,5	0,125	2	<b>0,038</b>	0,125
26	author	3,5	0,236	2	0	0,236

Tabelle B.49: Einzelwerte der Produktivität der Autorennamen für die Topics 001-026 (das Topic 20 fehlt systembedingt). Statistisch signifikante Werte sind **fett** gedruckt.

B Ermittelte Werte der Power-Law-Analyse

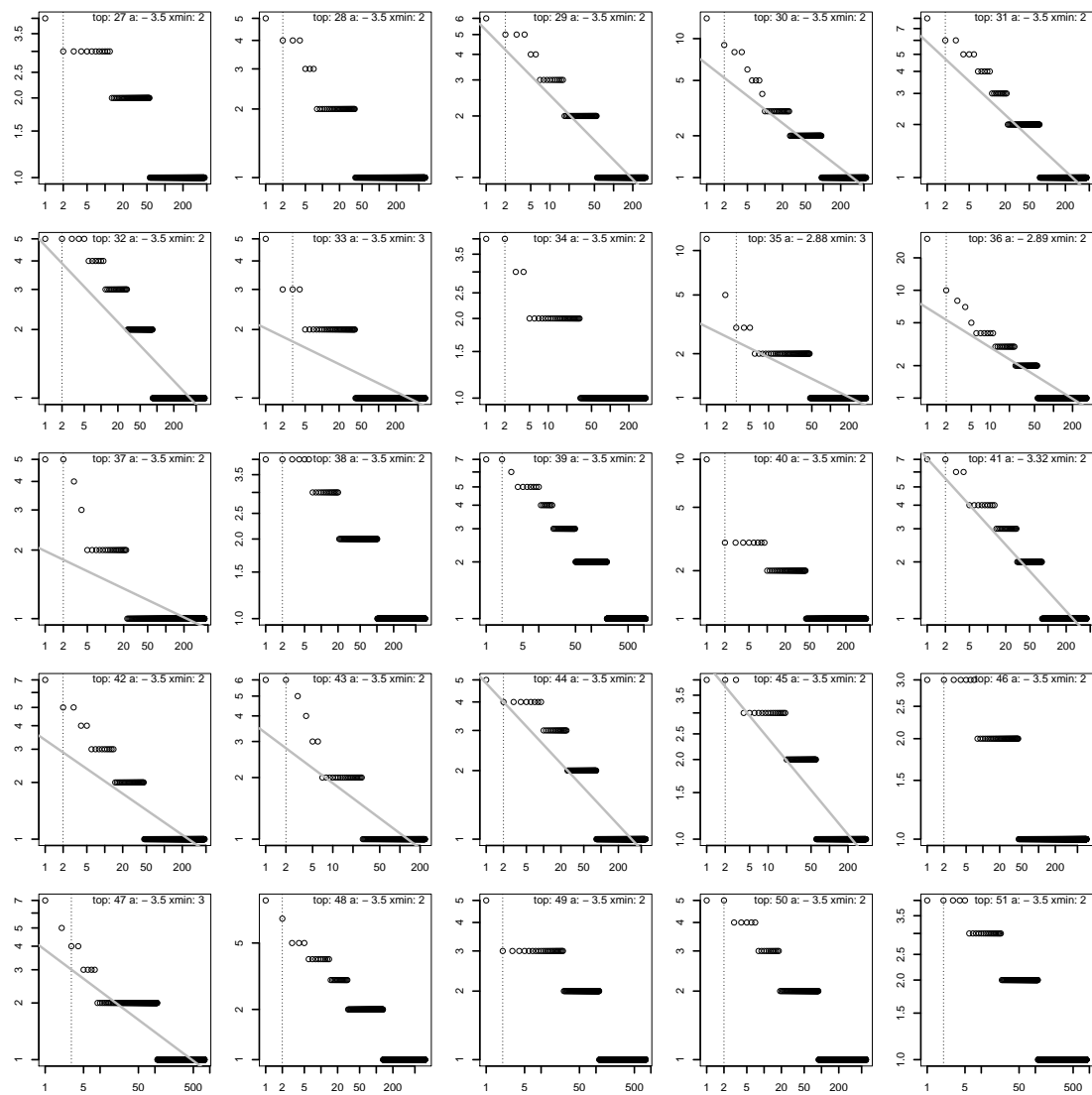


Abbildung B.50: Doppelt logarithmierter Plot der Produktivität der Autorennamen für die iSearch Topics 027-026.

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
27	author	3,5	0,116	2	0,015	0,116
28	author	3,5	0,113	2	0,091	0,113
29	author	3,5	0,029	2	<b>0,87</b>	0,029
30	author	3,5	0,037	2	<b>0,479</b>	0,037
31	author	3,5	0,019	2	<b>0,94</b>	0,019
32	author	3,5	0,047	2	<b>0,329</b>	0,047
33	author	3,5	0,192	3	<b>0,532</b>	0,192
34	author	3,5	0,178	2	0,016	0,178
35	author	2,88	0,141	3	<b>0,851</b>	0,141
36	author	2,89	0,044	2	<b>0,525</b>	0,044
37	author	3,5	0,121	2	<b>0,189</b>	0,121
38	author	3,5	0,105	2	0,006	0,105
39	author	3,5	0,054	2	0,037	0,054
40	author	3,5	0,111	2	0,077	0,111
41	author	3,32	0,051	2	<b>0,217</b>	0,051
42	author	3,5	0,023	2	<b>0,967</b>	0,023
43	author	3,5	0,08	2	<b>0,39</b>	0,08
44	author	3,5	0,06	2	<b>0,168</b>	0,06
45	author	3,5	0,084	2	<b>0,109</b>	0,084
46	author	3,5	0,14	2	0,022	0,14
47	author	3,5	0,067	3	<b>0,997</b>	0,067
48	author	3,5	0,075	2	0,016	0,075
49	author	3,5	0,125	2	0	0,125
50	author	3,5	0,098	2	0,014	0,098
51	author	3,5	0,095	2	0,009	0,095

Tabelle B.50: Einzelwerte der Produktivität der Autorennamen für die Topics 027-051.  
Statistisch signifikante Werte sind **fett** gedruckt.

B Ermittelte Werte der Power-Law-Analyse

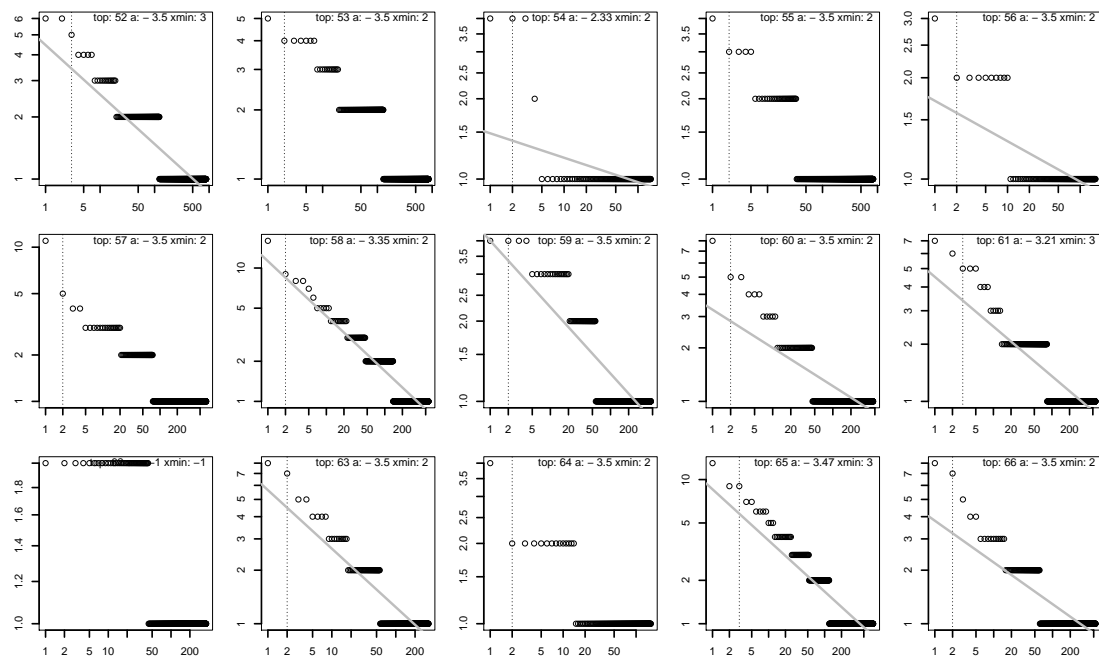


Abbildung B.51: Doppelt logarithmierter Plot der Produktivität der Autorennamen für die iSearch Topics 052-066.

B.9 Produktivität der Autorennamen in iSearch

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
52	author	3,5	0,096	3	<b>0,52</b>	0,096
53	author	3,5	0,154	2	0	0,154
54	author	2,33	0,413	2	<b>0,238</b>	0,413
55	author	3,5	0,151	2	0,03	0,151
56	author	3,5	0,203	2	<b>0,116</b>	0,203
57	author	3,5	0,08	2	0,084	0,08
58	author	3,35	0,009	2	<b>0,995</b>	0,009
59	author	3,5	0,072	2	<b>0,204</b>	0,072
60	author	3,5	0,058	2	<b>0,408</b>	0,058
61	author	3,21	0,13	3	<b>0,354</b>	0,13
62	author	-1	-1	-1	-1	-1
63	author	3,5	0,015	2	<b>0,998</b>	0,015
64	author	3,5	0,231	2	0,036	0,231
65	author	3,47	0,034	3	<b>0,677</b>	0,034
66	author	3,5	0,053	2	<b>0,392</b>	0,053

Tabelle B.51: Einzelwerte der Produktivität der Autorennamen für die Topics 052-066. Statistisch signifikante Werte sind **fett** gedruckt.

## B.10 Produktivität der Autorennamen in iSearch

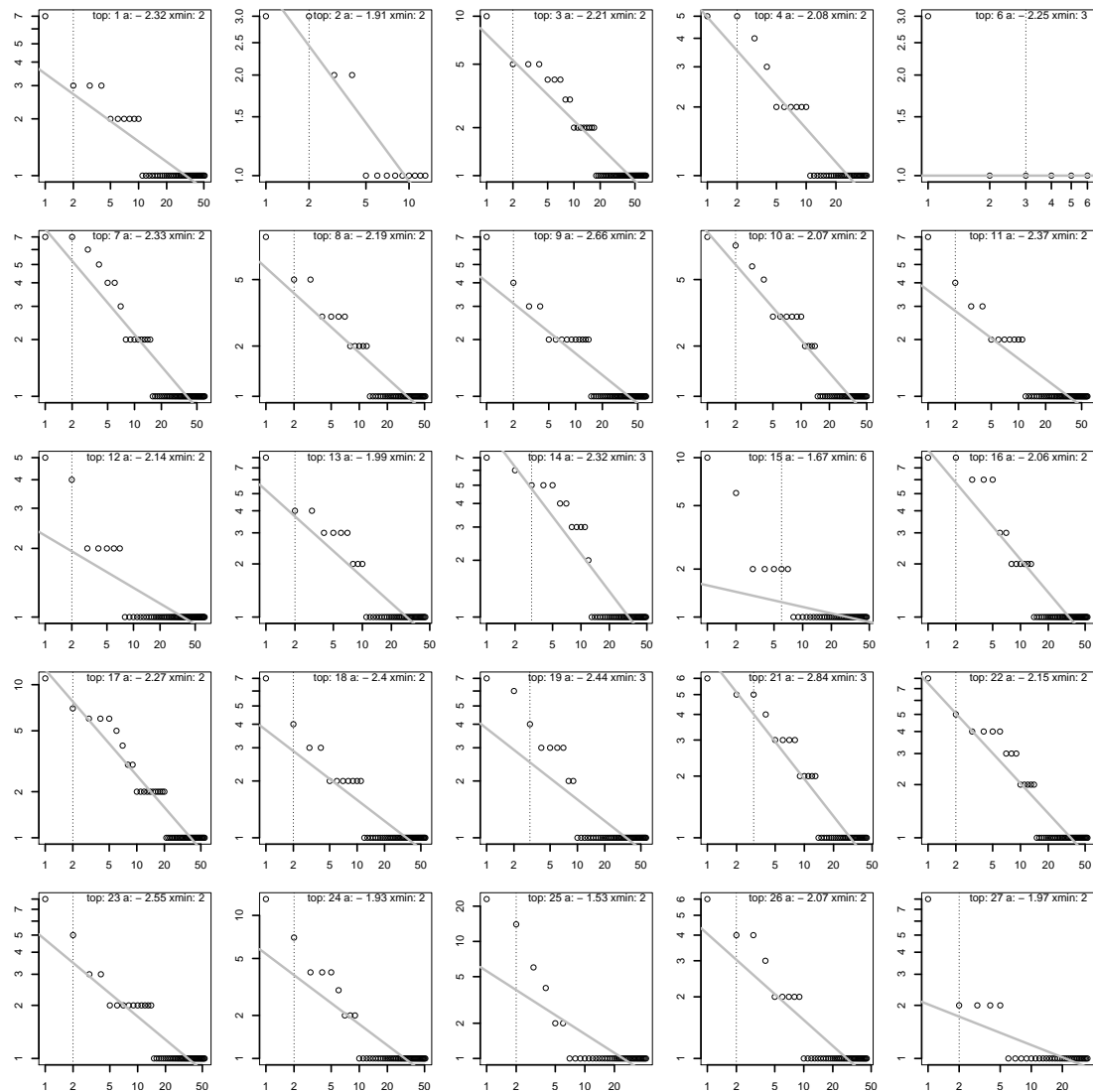


Abbildung B.52: Doppelt logarithmierter Plot der Produktivität der Publikationsquelle für die iSearch Topics 001-027 (die Topics 5 und 20 fehlen systembedingt).



Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
1	source	2,32	0,158	2	<b>0,319</b>	0,158
2	source	1,912	0,193	2	<b>0,712</b>	0,193
3	source	2,21	0,105	2	<b>0,504</b>	0,105
4	source	2,082	0,131	2	<b>0,621</b>	0,131
6	source	2,25	0,239	3	<b>0,92</b>	0,239
7	source	2,33	0,1	2	<b>0,634</b>	0,1
8	source	2,19	0,07	2	<b>0,977</b>	0,07
9	source	2,66	0,114	2	<b>0,43</b>	0,114
10	source	2,07	0,141	2	<b>0,353</b>	0,141
11	source	2,37	0,096	2	<b>0,769</b>	0,096
12	source	2,14	0,198	2	<b>0,371</b>	0,198
13	source	1,991	0,141	2	<b>0,549</b>	0,141
14	source	2,32	0,126	3	<b>0,619</b>	0,126
15	source	1,673	0,233	6	<b>0,602</b>	0,233
16	source	2,06	0,168	2	<b>0,218</b>	0,168
17	source	2,27	0,085	2	<b>0,621</b>	0,085
18	source	2,4	0,089	2	<b>0,82</b>	0,089
19	source	2,44	0,123	3	<b>0,782</b>	0,123
21	source	2,84	0,14	3	<b>0,324</b>	0,14
22	source	2,15	0,1	2	<b>0,716</b>	0,1
23	source	2,55	0,138	2	<b>0,245</b>	0,138
24	source	1,93	0,134	2	<b>0,757</b>	0,134
25	source	1,531	0,168	2	<b>0,836</b>	0,168
26	source	2,071	0,083	2	<b>0,953</b>	0,083
27	source	1,967	0,234	2	<b>0,387</b>	0,234

Tabelle B.52: Einzelwerte der Produktivität der Publikationsquelle für die Topics 001-027 (die Topics 5 und 20 fehlen systembedingt). Statistisch signifikante Werte sind **fett** gedruckt.

## B Ermittelte Werte der Power-Law-Analyse

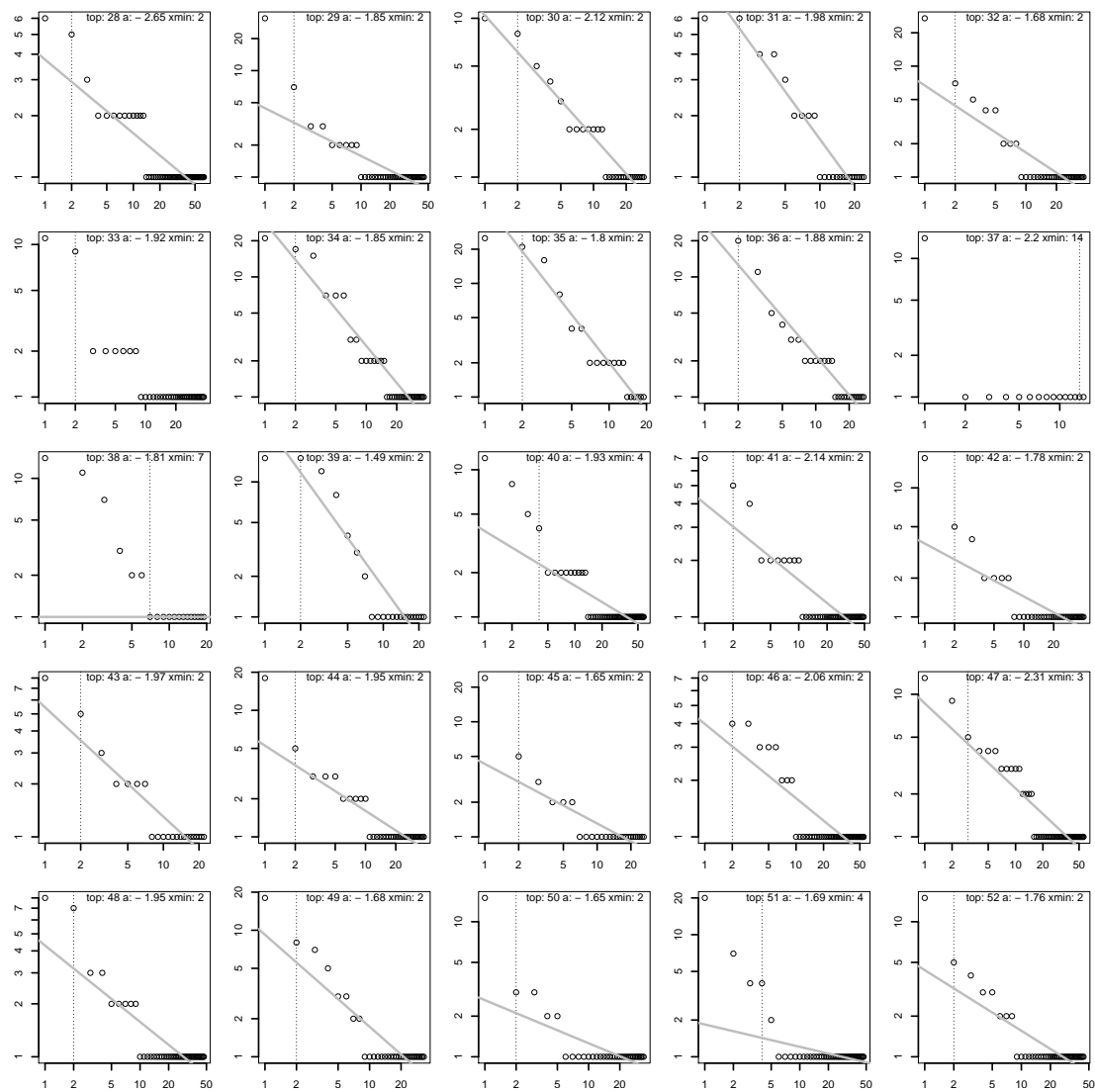


Abbildung B.53: Doppelt logarithmierter Plot der Produktivität der Publikationsquelle für die iSearch Topics 028-052.

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
28	source	2,65	0,164	2	<b>0,133</b>	0,164
29	source	1,855	0,158	2	<b>0,532</b>	0,158
30	source	2,117	0,092	2	<b>0,828</b>	0,092
31	source	1,981	0,11	2	<b>0,854</b>	0,11
32	source	1,676	0,132	2	<b>0,855</b>	0,132
33	source	1,916	0,27	2	0,07	0,27
34	source	1,853	0,167	2	<b>0,166</b>	0,167
35	source	1,799	0,155	2	<b>0,32</b>	0,155
36	source	1,877	0,108	2	<b>0,704</b>	0,108
37	source	2,2	0,347	14	<b>0,878</b>	0,347
38	source	1,81	0,171	7	<b>0,904</b>	0,171
39	source	1,49	0,213	2	<b>0,589</b>	0,213
40	source	1,93	0,131	4	<b>0,899</b>	0,131
41	source	2,136	0,18	2	<b>0,225</b>	0,18
42	source	1,779	0,146	2	<b>0,802</b>	0,146
43	source	1,971	0,082	2	<b>0,985</b>	0,082
44	source	1,945	0,153	2	<b>0,445</b>	0,153
45	source	1,651	0,159	2	<b>0,815</b>	0,159
46	source	2,06	0,122	2	<b>0,79</b>	0,122
47	source	2,31	0,136	3	<b>0,337</b>	0,136
48	source	1,954	0,119	2	<b>0,815</b>	0,119
49	source	1,676	0,113	2	<b>0,928</b>	0,113
50	source	1,65	0,193	2	<b>0,744</b>	0,193
51	source	1,688	0,207	4	<b>0,781</b>	0,207
52	source	1,756	0,117	2	<b>0,912</b>	0,117

Tabelle B.53: Einzelwerte der Produktivität der Publikationsquelle für die Topics 028-052. Statistisch signifikante Werte sind **fett** gedruckt.

B *Ermittelte Werte der Power-Law-Analyse*

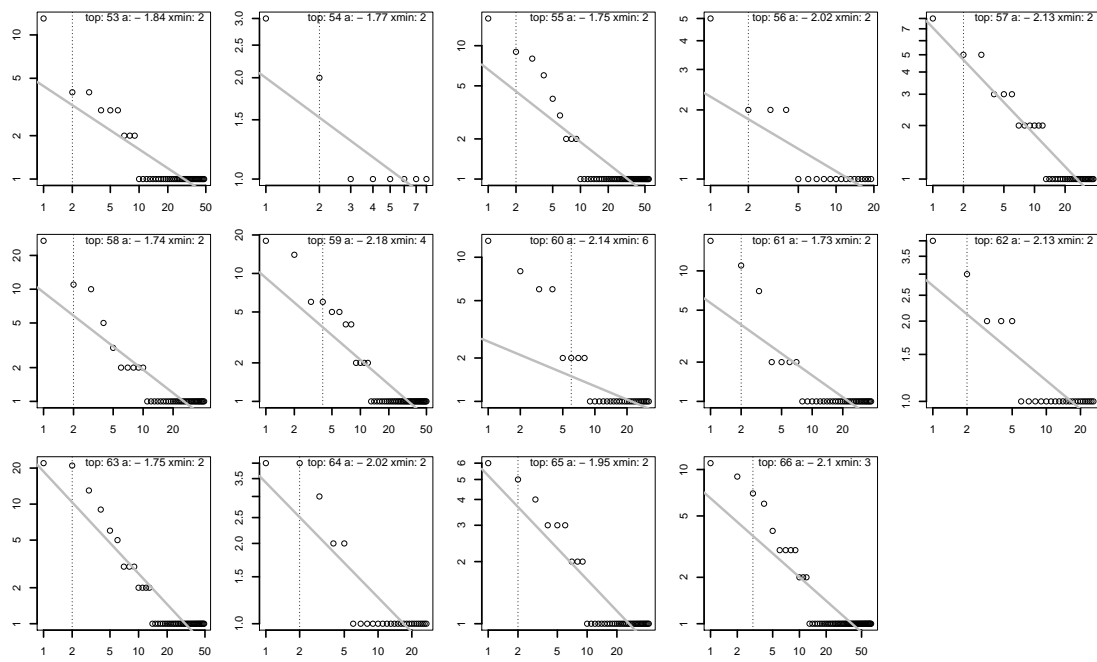


Abbildung B.54: Doppelt logarithmierter Plot der Produktivität der Publikationsquelle für die iSearch Topics 053-066.

Topic	Entität	$\alpha$	$D$	$x_{min}$	$p$	gof
53	source	1,837	0,17	2	<b>0,473</b>	0,17
54	source	1,767	0,192	2	<b>0,895</b>	0,192
55	source	1,75	0,115	2	<b>0,91</b>	0,115
56	source	2,016	0,142	2	<b>0,834</b>	0,142
57	source	2,126	0,061	2	<b>0,988</b>	0,061
58	source	1,736	0,134	2	<b>0,722</b>	0,134
59	source	2,18	0,151	4	<b>0,535</b>	0,151
60	source	2,144	0,189	6	<b>0,769</b>	0,189
61	source	1,726	0,197	2	<b>0,498</b>	0,197
62	source	2,133	0,114	2	<b>0,914</b>	0,114
63	source	1,752	0,107	2	<b>0,783</b>	0,107
64	source	2,025	0,134	2	<b>0,856</b>	0,134
65	source	1,945	0,122	2	<b>0,788</b>	0,122
66	source	2,1	0,096	3	<b>0,89</b>	0,096

Tabelle B.54: Einzelwerte der Produktivität der Publikationsquelle für die Topics 053-066. Statistisch signifikante Werte sind **fett** gedruckt.



## C Publikationen im Rahmen der Dissertation

Im Rahmen dieser Dissertation sind mehrere Publikationen entstanden, die jeweils unterschiedliche Aspekte dieser Arbeit auf internationalen Konferenzen, Workshops und in Fachjournalen präsentierten.

### **Lotkaische Informetrie und nicht-textuelles Reranking**

Die folgenden Papiere beschäftigen sich mit der grundsätzlichen Idee, Produktivitätszahlen und die lotkaischen Informetrie zur Anreicherung des IR zu verwenden. Das Papier für den *ASIS&T European Workshop* (AES) konnte den von EMC<sup>2</sup> gestifteten *Best-Paper-Award* gewinnen (Schaer, 2011). Ein weiteres Papier erschien in dem Fachjournal *Scientometrics* und behandelte die Verbindung von wissenschaftlicher Modellbildung und IR (Mutschke u. a., 2011). Auf dem *Internationalen Symposium für Informationswissenschaft* (ISI) 2011 wurden erstmalig Überlegungen zu einer Kombination unterschiedlicher Rankingverfahren, die auf wissenschaftlichen Modellen basieren, präsentiert (Mayr u. a., 2011a).

### **Kookkurrenzanalyse und nicht-textuelle Anfrageerweiterungen**

Untersuchungen zum Informationsverhalten von Nutzern in digitalen Bibliotheken am Beispiel des Informationsportals Sowiport und die Implementierung und Evaluation der Vorgängerversion der in dieser Arbeit verwendeten Kookkurrenzanalyse wurden auf der internationalen Konferenz *Theories and Practices in Digital Libraries* (TPDL, ehemals ECDL) vorgestellt (Hienert u. a., 2011a). Die im Abschnitt 7.1.3 beschriebenen Ergebnisse zur Anfrageerweiterung mit Autorennamen wurden auf der TPDL 2012 präsentiert (Schaer u. a., 2012c), ebenso wie weitere Experimente mit den entwickelten STR-Komponenten. In diesem Papier stand der Einsatz von speziell trainierten STR-Modulen, die mit Anfrageklassen arbeiten, im Fokus (Lüke u. a., 2012). Die STR-Module wurden für die internationale IR-Evaluationskampagne CLEF im *Cultural-Heritage-Track* angewendet (Schaer u. a., 2012a).

### **IR-Mehrwertdienste in digitalen Bibliotheken und technische Grundlagen**

Das in dieser Dissertation entwickelte System und die im IRM-Projekt implementierten Dienste wurden auf dem *73rd ASIS&T Annual Meeting* präsentiert (Schaer u. a., 2010a). Ein technisches System zur selbstständigen Generierung von Modulen zur Anfrageerweiterung, die als STR in eigene IR-Systeme eingebunden werden können, wurde auf der *64. Jahrestagung der DGI* 2012 vorgestellt (Schaer u. a., 2012b).

### **IR-Evaluation**

Mit Fragen der IR-Evaluation und dem Problem der statistischen Auswertung von Relevanzurteilen beschäftigen sich zwei weitere Papiere. Ersteres wurde auf dem *Jahrestreffen*

## C Publikationen im Rahmen der Dissertation

der Fachgruppe Information Retrieval der Gesellschaft für Informatik (FGIR der GI) vorgestellt (Schaer u. a., 2010b). Das zweite Papier beschäftigt sich mit widersprüchlichen Relevanzbewertungen von Assessoren und wurde auf der internationalen *Conference and Labs of the Evaluation Forum* (CLEF) vorgestellt (Schaer, 2012).

### **Digitale Bibliotheken und Informationswissenschaft**

Weitere Arbeiten, die in den größeren Rahmen von Informationsversorgung, Open-Access und digitalen Bibliotheken gehören, wurden auf der *11. Tagung der Deutschen ISKO* (International Society for Knowledge Organization, Schaer, 2010) und dem *63. Jahrestagung der DGI* (Wilde u. a., 2010) präsentiert. Im internationalen *Grey Journal* erschien ein Artikel zur Intergration von grauer Literatur und Open-Access-Artikeln in das System Sowiport (Stempfhuber u. a., 2008).

### **Weitere Themen**

Seit 2009 schreibt der Autor für das Lexikon der Bibliotheks- und Informationswissenschaft zu den Themen Kerninformatik, IR, Web-Standards, Netzwerke und Software-Ergonomie. Insgesamt sind bis Ende 2012 über 100 Artikel für das Lexikon entstanden (Umlauf u. Gradmann, 2009). Die DFG-geförderten Projekte *IRM* und *IRM2*, die die Grundlage für diese Dissertation bildeten, wurden in den Papieren von Mayr u. a. (2011b) und Mayr u. a. (2011c) vorgestellt. Weiterhin sind die folgenden Papiere dem Bereich Software-Ergonomie und Interaktionsdesign zuzuordnen und damit Bestandteil des Abschnitts 4.5: Schaer u. Heuser (2006) und Hienert u. a. (2011b,c).



## D Software im Rahmen der Dissertation

Das im Rahmen des IRM-Projekts entwickelte Software-Framework IRSA<sup>1</sup> (Information Retrieval Service Assessment), das zur Evaluation in dieser Arbeit verwendet wurde, ist als Open-Source-Software (Apache License V2.0) auf der Plattform SourceForge verfügbar. Eine Sammlung von Tools und Skripten zur Datenaufbereitung und -analyse ist bei Google Code<sup>2</sup> ebenfalls unter Open-Source-Bedingungen (Apache License V2.0) zu finden. Das entwickelte Framework besteht hauptsächlich aus den folgenden Softwarekomponenten (s. Abb. D.1a), wie auch in Schaer u. a. (2011, S. 6 ff.) ausführlich beschrieben:

- der quelloffenen Suchmaschine Solr auf Lucene-Basis (als grundlegende Retrieval-technologie) und einer PostgreSQL-Datenbank (zur Datenhaltung für die Assessments und Dokumente),
- dem ebenfalls quelloffenen Web-Framework Grails auf Basis von Spring und der Skriptsprache Groovy (für den Service-Layer auf REST-Basis und die grafischen Benutzerschnittstellen),
- Plugins für das Grails-Framework und für Solr, die für das Reranking und die Anfrageerweiterungen zuständig sind.

Grails ist das zentrale Framework, das die einzelnen Softwarekomponenten zusammenfügt und für den Nutzer in Form eines interaktiven Web-Prototyps<sup>3</sup> zur Verfügung stellt (s. Abb. D.1b). Über die Oberfläche des Prototyps kann ein beliebiges Suchergebnis nach alternativen Rankingmethoden (Kernzeitschriften und Autorenzentralität) umsortiert gelistet werden. Darüber hinaus kann der Benutzer auf der rechten Seite der Oberfläche die Liste der Kernzeitschriften bzw. zentralen Autoren einsehen und diese für eine Anfrageerweiterung nutzen. Für das Reranking nach Autorenzentralität wurden Softwaremodule verwendet, die bereits in den DFG-Projekten Daffodil und infoconnex zum Einsatz kamen. Für den Einsatz im Prototyp war nur eine geringfügige Anpassung der Software nötig, um das Ranking als Solr-Search-Plugin in den Prototyp integrieren zu können. Neben dem Grails-Framework wird zur Suche auf den freien Suchserver Solr zurückgegriffen, welcher u.a. die Möglichkeit zur direkten Facettierung und einer Einbindung in beliebige eigene Projekte mittels eines RESTful-Webservices bietet.

Zusätzlich zu den Service-Schnittstellen und den interaktiven Prototypen können mit Hilfe von IRSA Relevance-Assessments durchgeführt werden. Hierzu werden die gleichen Komponenten und Services verwendet, die auch für die Prototypen genutzt werden.

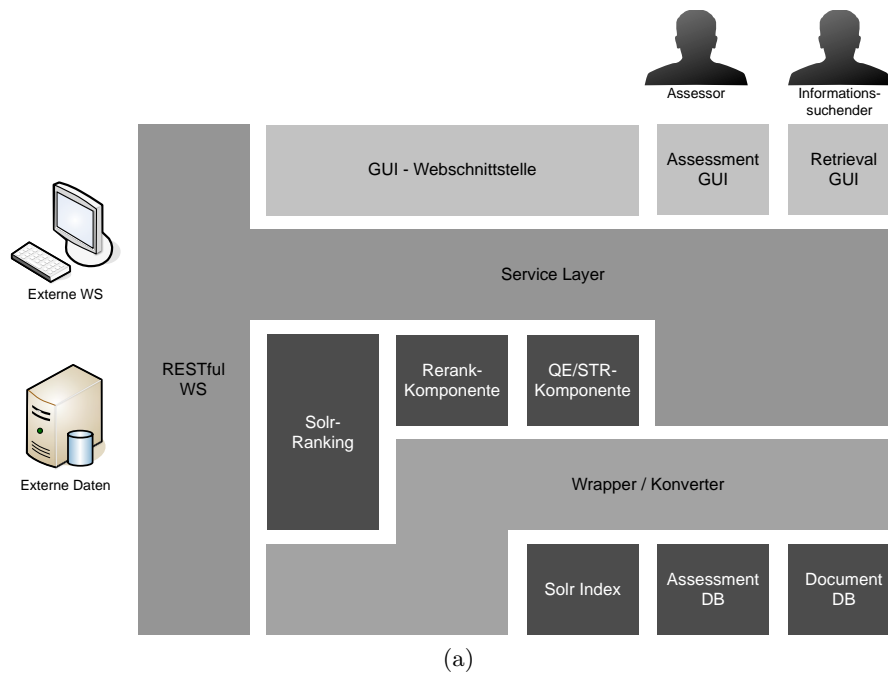
---

<sup>1</sup><http://sourceforge.net/projects/irsa>

<sup>2</sup><http://code.google.com/p/evaltools/>

<sup>3</sup><http://multiweb.gesis.org/irsa/IRMPrototype/index>

## D Software im Rahmen der Dissertation



### Information Retrieval Value-added Services

Query:

Only show metadata sets which include an abstract

Applied filters: collection:(fis\*)

Automatic Query Expansion  Rerank the result list

Expanded query with the following terms: [Erziehungswissenschaft, vergleichende, Untersuchung, empirische, Forschung, empirische, Sportsoziologie]

Total hits: 11295

1. Zum Themenschwerpunkt 'Bildungsforschung: disziplinäre, methodische und empirische Beispiele'. (2005)  
Zeitschriftenaufsatz in *Bildungsforschung* ( 1860-8213 ) by Schaffert, Sandra; Schmidt, Bernhard
2. Möglichkeiten und Grenzen empirischer Bildungsforschung.  
Sammelwerksbeitrag in *Jugend und Bildung*. ( 3-86649-151-4; 978-3-86649-151-9 ) by Tippelt, Rudolf; Schmidt, Bernhard
3. Rahmenprogramm zur Förderung der empirischen Bildungsforschung. (2007)  
Monographie
4. Allgemeine Erziehungswissenschaft versus Empirische Schulforschung?  
Sammelwerksbeitrag in *Bildungsforschung und Bildungsreform*. ( 3-7799-0938-3; 9783779909385 ) by Ruhloff, Jörg
5. Elemente des Rahmenprogramms des BMBF zur strukturellen Förderung der empirischen Bildungsforschung in Deutschland. (2007)  
Monographie

### Interactive query enhancement

term cloud ▼

Unternehmen Sportökonomie  
**Erziehungswissenschaft,**  
vergleichende Karriere,  
berufliche Individualisierung  
Untersuchungsergebnis  
**Untersuchung, empirische**  
Wirkungsgrad Zuschauersport  
Wissenschaftliches Arbeiten  
Vereinsmitglied **Forschung,**  
**empirische** Bewegungshandlung  
Beobachtung Geschlechtsabhängigkeit  
Forschungsstand Betriebsport  
Motivationsforschung  
Sportsoziologie

search term suggestions ▼

core journals ▼

central authors ▼

- Schmid, Günther (0.022784915)
- Klieme, Eckhard (0.0153759485)
- Bellmann, Lutz (0.015198311)
- Baethge, Martin (0.013915186)
- Faulstich, Peter (0.011628525)
- Baumert, Jürgen (0.010552036)

(b)

Abbildung D.1: Das IRSA-Evaluationsframework: (a) schematische Architekturskizze und (b) Screenshot des interaktiven Prototyps.

# E Lebenslauf

Philipp Schaer, geboren am 7. Mai 1980 in Troisdorf.

## Ausbildung

- 07/2011–05/2013 **Doktorand**, Fachbereich 4 der Universität Koblenz-Landau  
10/2000–01/2006 **Diplom-Informatiker** (Dipl.-Inform.), Universität Koblenz-Landau, Koblenz, Abschlussnote: sehr gut (1,5)  
07/1990–06/1999 **Abitur**, Heinrich-Böll Gymnasium, Troisdorf

## Wissenschaftlicher und beruflicher Werdegang

- 01/2007–heute **Wissenschaftlicher Mitarbeiter** in der Abteilung Wissenstechnologien für die Sozialwissenschaften, GESIS – Leibniz-Institut für die Sozialwissenschaften, Köln (vormals Informationszentrum Sozialwissenschaften, Bonn)  
04/2013–heute **Lehrbeauftragter** für das Fach Online-Recherche, Fachhochschule Köln, Köln  
01/2006–06/2009 **Lehrbeauftragter** für das Fach Software-Ergonomie, Universität Koblenz-Landau, Koblenz  
01/2006–12/2006 **Wissenschaftlicher Mitarbeiter** in der Forschungsgruppe Information Retrieval und Software-Ergonomie, Universität Koblenz-Landau, Koblenz  
08/2003–12/2005 **Studentische Hilfskraft** in der Forschungsgruppe Information Retrieval und Software-Ergonomie, Universität Koblenz-Landau, Koblenz  
07/2002–09/2002 **Studentische Hilfskraft**, Fraunhofer-Institut für Angewandte Informationstechnik FIT, St. Augustin  
07/2000–09/2000 **Praktikant**, GMD – Forschungszentrum Informationstechnik, St. Augustin  
07/1999–06/2000 **Zivildienst als Krankenpfleger**, St. Johannes Krankenhaus, Troisdorf

## Auszeichnungen

**Best Student Paper** ASIS&T European Workshop, 1.–2. Juni 2011, Cork, Irland