Contextualised Stratagem Browsing in Digital Libraries

Von der Fakultät für Ingenieurwissenschaften,
Abteilung Informatik und Angewandte Kognitionswissenschaft
der Universität Duisburg-Essen

zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften (Dr.-Ing.)

genehmigte Dissertation

von

Željko Carević
aus
Wuppertal

1. Gutachter: Prof. Dr.-Ing. Norbert Fuhr
2. Gutachter: Prof. Dr. Stefan Dietze

Datum der mündlichen Prüfung: 09.07.2020

DOI: 10.17185/duepublico/72522
URN: urn:nbn:de:hbz:464-20200811-134729-8

Abstract

Exploratory browsing is an essential part of the information-seeking process, which is often performed with the intent to better define an information need. Especially, Digital Libraries facilitate exploratory browsing by exploiting the inherent structure contained in the descriptive metadata of academic literature. Based on shared characteristics, such as keywords, classifications, or the journal in which an academic publication was published, users are enabled to explore and discover new and potentially relevant content. This type of exploration is referred to as *stratagem browsing*.

Although many modern Digital Libraries allow for stratagem browsing, these types of searches are under-investigated. To this day, it remains unclear in which situations stratagem browsing can support the fulfilment of users' information needs. Furthermore, many existing stratagem browsing implementations are rather out-of-the-box and do not include more advanced methods, such as *contextualisation*.

In order to better support users engaged in these types of searches, this thesis proposes a re-design towards contextualised stratagem browsing in Digital Libraries. By employing the users' interaction context, a re-ranking of stratagem browsing results can be performed which is tailored towards the users' information needs. The applicability of this contextualisation is investigated in a series of three studies: (i) an online survey, (ii) a user study, (iii) a transaction log study.

Following the principles of Interactive Information Retrieval, the online survey and the user study are designed to gain a thorough understanding of stratagem browsing. The online survey aims to assess the usefulness of stratagem browsing during academic searches. The user study is designed to observe the participants information-seeking behaviour with respect to the operationalisation of stratagem browsing. Subsequently, the large-scale transaction log study investigates the effectiveness of contextualisation for stratagem browsing. To this end, two contextualised stratagem browsing variants are developed that employ the users' interaction context for a re-ranking of search results: (1) one variant bases on document similarity, and (2) one variant bases on query expansion. The effectiveness of these two methods is evaluated using the real-life Digital Library Sowiport. An A/B/C-testing is designed to compare the contextual approaches against Sowiport's non-contextual baseline. Amongst others, the rank of the first clicked document in a result list, denoted as the mean first relevant (MFR), and the click-through rate are employed to determine the effectiveness of contextualised stratagem browsing.

The results of this experiment show that both contextualised browsing variants significantly outperform the non-contextualised baseline in terms of MFR and acquire a considerably higher click-through rate.

Zusammenfassung

Exploratives Browsen ist ein wesentlicher Teil des Informationssuchprozesses, häufig mit dem Ziel ein Informationsbedürfnis besser zu definieren. Insbesondere Digitale Bibliotheken fördern exploratives Browsen unter Verwendung der inhärenten Struktur deskriptiver Metadaten. Potenziell relevante Inhalte können dabei auf Basis gemeinsamer Merkmale, wie Keywords, Klassifikationen oder Fachzeitschriften erschlossen werden. Diese Art der Exploration wird als *Stratagem-Browsing* bezeichnet.

Obwohl viele moderne Digitale Bibliotheken Stratagem-Browsing unterstützen, ist diese Art des Informationssuchverhaltens bis heute in der Forschung nicht ausreichend untersucht. Es ist unklar, in welchen Situationen Stratagem-Browsing die Befriedigung eines Informationsbedürfnisses unterstützen kann. Darüber hinaus basiert die Umsetzung des Stratagem-Browsing häufig auf Standardlösungen ohne Einbeziehung fortschrittlicherer Methoden wie z.B. Kontextualisierung.

Um die Benutzer bei dieser Art der Suche besser zu unterstützen, wird in dieser Arbeit eine Neugestaltung hin zu kontextbasiertem Stratagem-Browsing in Digitalen Bibliotheken vorgeschlagen. Unter Anwendung des Interaktionskontextes der Benutzer wird eine Reorganisation von Stratagem-Browsing Ergebnislisten umgesetzt, die auf das Informationsbedürfnis der Benutzer zugeschnitten ist.

Die Anwendbarkeit dieser Kontextualisierung wird in drei Studien untersucht: (i) eine Online-Befragung, (ii) eine Benutzerstudie, (iii) und eine Transaction-Log Studie. Den Prinzipien des Interactive Information Retrieval folgend, sind die Online-Befragung und die Benutzerstudie so konzipiert, dass zunächst ein umfassendes Verständnis des Stratagem-Browsing erarbeitet wird.

Die Online-Befragung zielt darauf ab, die Nützlichkeit des Stratagem-Browsing bei akademischen Recherchen zu beurteilen. Die Benutzerstudie dient der Beobachtung des Informationssuchverhalten der Studienteilnehmer mit Hinblick auf die Operationalisierung des Stratagem-Browsing.

Anschließend wird in einer Transaction-Log Studie die Effektivität der Kontextualisierung für das Stratagem-Browsing evaluiert. Zu diesem Zweck werden zwei kontextbasierte Stratagem-Browsing Varianten entwickelt, die den Interaktionskontext der Nutzer für ein Ranking der Suchergebnisse nutzen: (1) eine Variante basiert auf Dokumentenähnlichkeit und (2) eine Variante auf Anfrageerweiterung. Die Wirksamkeit dieser beiden Methoden wird unter Verwendung der Digitalen Bibliothek Sowiport evaluiert. Ein A/B/C-Test wird durchgeführt, um die kontextabhängigen Ansätze mit der kontextfreien Baseline von Sowiport zu vergleichen. Unter anderem werden der Rang des zuerst angeklickten Dokuments in einer Ergebnisliste, bezeichnet als mean first relevant (MFR), und die Klickrate verwendet, um die Wirksamkeit von kontextbasiertem Stratagem-Browsing zu bestimmen. Die Ergebnisse dieses Experiments zeigen, dass beide kontextbasierten Browsing-Varianten die kontextfreie Baseline in Bezug auf den MFR statistisch signifikant übertreffen und eine wesentlich höhere Klickrate erzielen.

Acknowledgements

I want to take this opportunity to express my gratitude towards my family, friends, and colleagues that have supported me during the last years. First, I want to thank my supervisors Norbert Fuhr and Stefan Dietze. I am especially grateful to Norbert Fuhr for guiding me through this thesis with his knowledge, experience, and patience since we met at the Autumn School on Information Retrieval in 2014. Although he supervised me as an external Ph.D. student, support was always just around the corner and a helpful response only just a few minutes away. Stefan Dietze helped me to improve the conceptual details of contextualisation with a door always open no matter how stressful his own situation. Even though it is common practice to thank one's supervisors, I did not take their support for granted at any moment of my time as a Ph.D student. Furthermore, I want to thank Philipp Mayr, my team leader at GESIS and additional Ph.D. advisor. Philipp helped me develop the research topic of this thesis at a very early stage and has since offered guidance and support through the entire time of this dissertation. Besides my supervisors, I want to thank my family: Mate and Tonka for all their love, unconditional support, and hard work for which I will always be grateful. My sisters, Sandra and Sonja, for all the encouragement and interest in discussing my research. My nieces Lari, Lea, and Marie who made fun of me for not being done with "school" and thus, gave me extra motivation. I am particularly thankful for the support from Heidi Schulze. She helped me through the ups and downs of writing this thesis in all kinds of roles from proof-reader to a therapist. Her contribution, in particular, during the most challenging times, was extraordinary, and she is probably the one person that is as happy as I am that this chapter is done. And of course, my friends at GESIS: Wanja Schaible, who taught me all necessary basics one needs to know as a Ph.D. student and above that many unnecessary but highly entertaining things, repeatedly. Dimitar Dimitrov, for constantly questioning my point of view leading to many hilarious but also useful discussions. Matthäus Zloch, my office neighbour, for his helpfulness and calmness even during the most stressful times. Kata Boland, Felix Bensmann, and Wilko Van Hoek for countless hours of discussing research or any other topic of interest and for every highly appreciated distraction. And, of course, Dagmar Kern, Daniel Hienert, Sascha Schüller, Karam Abdulahhad, Philipp Schaer, and Maria Lusky. I would also like to thank all other colleagues of the WTS department for their support, in particular, Siegfried Schomisch who always took care of urgent matters whenever I was stuck in my research. Last but certainly not least, I want to thank my friends Wit Urban, Tim Adrian Peplies, and Henrik Schulze Neuhoff who would either listen to and motivate me or take my mind off research with a Kölsch. Overall, I am very grateful to be surrounded by a lot of smart and supportive people who have made my time as a Ph.D. student not just hard work, but also a lot of fun and inspiration.

Contents

List of Figures

List of Tables

Listings

1

Introduction

The history of Information Retrieval (IR) goes back far beyond the beginning of the world wide web. The earliest computer-based search systems date back to the late 1940s (Sanderson & Croft, 2012). In a simplified way, IR deals with the structure, analysis, organisation, storage, searching, and retrieval of information (Salton, 1968). While Salton's definition focuses on encompassing the main components of IR, other researchers emphasise the main objective of IR which is to deal with *vagueness*, i.e. the inability of a user to precisely define an information need and *imprecision*, i.e. the imperfection in the representation of the semantics and pragmatics of the stored objects in information systems (Fuhr, 2012). The relationship between vagueness and imprecision is displayed in an adapted representation of the classic IR model in Figure 1.1. On the left-hand side of the figure, one can see the system-oriented perspective that transforms an information object, for example, a text document into an indexed document representation. The right-hand side of the figure represents the user-oriented part of IR. A user is transforming an information need into a query that is then matched against the indexed representation of the document within the information system. Both perspectives are faced with different kinds of challenges. While the information system has to provide an adequate semantic representation of the document itself, the user side is challenged by the transformation of an information need into a query that represents that particular information need in an adequate and precise way.

Despite these challenges, IR research has had and still has a significant impact on the development of today's information systems. One of the reasons for the success of IR research is based on a resource known as *test collection*. A test collection used in conjunction with evaluation measures enables researchers to assess the effectiveness of an IR system. A test collection usually consists of a set of documents, a set of topics (often referred to as queries), and a set of relevance judgements that account for the relevance of documents to queries. IR researchers in possession of an appropriate test collection are enabled to simulate a user of a search system and evaluate the effectiveness of a retrieval method in isolation (Sanderson, 2010). These types of evaluation settings are commonly referred to as the *Cranfield paradigm*, and they have been one of the main driving forces in IR research that became a de facto standard in the IR community.

The benefit of test collections used in shared evaluation campaigns is that researchers ensure the reproducibility of their approach, in order to enable other researchers in that community to compare and compete with their own approach against strong baselines.

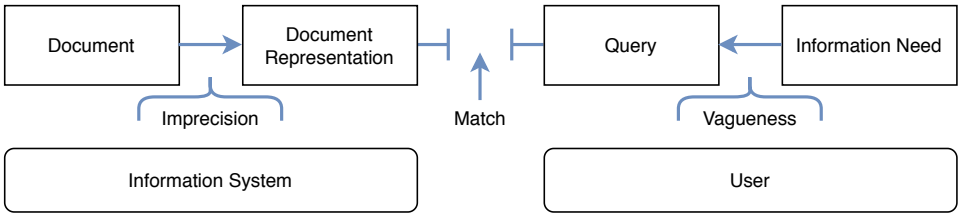


Figure 1.1: Adapted version of the classic IR model as presented in Bates (1989) that incorporates imprecision and vagueness.

1.1 From Information Retrieval to Interactive Information Retrieval

The contribution of the Cranfield paradigm is indisputable, and its application has led to substantial improvements of IR systems. One major drawback of the Cranfield paradigm, however, is that it reduces the process of information seeking to a simplified query-response operation, which is limited to the system-oriented perspective. Hence, IR research following the Cranfield paradigm has been criticised for performing in an artificially and closed environment. This isolation is metaphorically described as the Cranfield cave in which external effects[1] of a user, such as "real" and dynamically evolving information needs or the user's state of knowledge, are ignored. Instead, the Cranfield paradigm is following the underlying assumption of retrieving documents to fixed information needs substituting or even entirely neglecting the user-oriented aspects of IR.

To form a paradigm change to this classical IR research, one can observe a shift away from this mainly system-oriented perspective towards a more user-oriented perspective (Ingwersen & Järvelin, 2005b), today referred to as Interactive Information Retrieval (IIR). IIR understands the process of information seeking as a highly interactive task, that incorporates several dimensions of research, such as the study of information-seeking behaviour, human-computer interaction, cognition, and contextualisation. Today, IIR is considered a sub-discipline of classic IR, both sharing the same goal: to build retrieval systems that support a user in satisfying an information need. The methods, however, are highly different. IIR, in contrast to classic IR, aims to improve retrieval systems by establishing a better understanding of the interactions between a retrieval system and its users. In order to do so, IIR employs different methods and strategies, like the observation of users in naturalistic settings, discourse analysis, and other protocol analyses (Robins, 2000).

[1] Denoted as *Context* by Ingwersen & Järvelin (2007).

1.1.1 Exploratory Search

One can argue that classic IR is sufficient when the information need is well-defined, and the user is aware of what the underlying terminology is. However, in situations in which the search is driven by more complex information needs like decision making, learning and other cognitive tasks, the methods of classic IR are not suitable anymore. Moving further from the system-oriented perspective towards the user-centred interactive information retrieval approach, Marchionini introduced the concept of *exploratory search*.

Exploratory search tasks comprise search activities on the level of *learning* and *investigating*. These activities go beyond simple lookup tasks such as known-item searches (Marchionini, 2006). Although exploratory searches describe highly interactive tasks that involve various search tactics occurring iteratively, the two predominant search activities are *searching* and *browsing*. Browsing refers to the selection of links or categories that produce pre-defined groups of information items. This could, for instance, involve browsing through documents that share a particular metadata. Searching, on the other hand, produces ad hoc collections of information that have not been gathered together before (Hearst, 2009).
Both searching and browsing are complementary and reasonable in different situations. Searching has the advantage that it may lead to useful results when the user is aware of the underlying terminology or looking for a known-item. Browsing, on the contrary, is useful when the user is not aware of appropriate search keywords or not certain of what he or she is looking for. An example for a goal of a user that utilises browsing is to learn more about the terminology in a certain field in order to get familiar with a particular topic.

1.1.2 Stratagems in Digital Libraries

The characteristics of learning and investigating, in the context of exploratory search, are frequently associated with scholarly information systems such as Digital Libraries (DLs). Particularly, search activities on the level of browsing play a fundamental role in DLs. The importance of browsing is reflected in many of the most renowned models of information-seeking behaviour (e.g. Ellis, 1989; Kuhlthau, 1991; Bates, 1989). Today, DLs facilitate browsing by exploiting structured metadata that annotate the content of scholarly DLs. This enables users to explore the content based on shared characteristics like keywords, classifications, or author information. These types of searchers are referred to as *stratagems*, and they are today's implementation of a conceptual search activity which has been introduced as part of an information-seeking model by Bates (1990).
Following the definition of Bates, a stratagem: ".. is a complex of a number of moves and/or tactics, and generally involves both a particular identified information search domain anticipated to be productive by the searcher, and a mode of tackling the particular file organization of that domain." (Bates, 1990, p. 6). Hence,

a stratagem could be, for instance, a so-called ”journal run” where a user identifies a journal to be useful for his or her research and browses the latest publications of that journal. A further example of a stratagem is to follow references in a particular document that might lead to potentially relevant material.

Stratagems represent a specialised form of browsing rooted in the area of DLs. Due to the strong relationship between stratagems and browsing, these types of search activities will be referred to as *stratagem browsing* throughout this thesis.

1.2 Motivation and Problem Statement

Although stratagem browsing enables the exploration of a DL, the system support on this level is rather low. Modern DLs usually support stratagem browsing as simple Boolean filters, that disregard information about the present user, his or her information need, or state of knowledge. Instead, stratagem browsing in DLs is provided from a solely system-oriented perspective in the form of a *filter query* $\rightarrow$ *response* mechanism, contradicting the principles of IIR.

Many DLs facilitate stratagem browsing by implementing the descriptive metadata of a document as hyperlinks which, when utilised, generate a result list of records that share that particular information with the seed document. A click on a particular keyword, for instance, will thus, generate a result list containing all documents that share this keyword. The content of these filtered lists is usually organised on the basis of generic features, such as "publication date" or in alphabetic order.
Organising content from stratagem browsing by such generic features is intuitive, but not suitable for complex search tasks such as learning. Situations in which stratagem browsing is employed on the basis of rather generic concepts may easily lead to result lists with thousands of documents, which are nearly impossible to be assessed by a user.

The gap between the importance of browsing during exploratory search and the limited system support described above, is even more evident when looking at the following consideration of browsing during exploratory search:

> "Browsing may be a hypothesis-generation activity, whereby hypotheses are generated about the causes of observed phenomena or the best ways to resolve an information problem. During hypothesis generation, users will visit multiple documents to better understand what information is available and familiarize themselves with the topic." (White & Roth, 2009, p.17).

According to this, browsing is utilised, in order to form a hypothesis to get familiar with a topic. Current DLs however, only supply the technical means of stratagem browsing. The intentions, the current state of knowledge and the preceding trails

of search activities that contribute to generate a hypothesis are disregarded by the retrieval system. Instead, an unassociated list of documents is generated that share selected characteristics. Thus, this type of system behaviour forms a mismatch between the intentions of the user and the system-oriented response, which is inappropriate for tasks on the level of stratagem browsing.

To provide a user-oriented support for stratagem browsing that is in line with the principles of IIR, this thesis presents a contextualised browsing approach that tailors documents from stratagem browsing towards the user's information need. Instead of applying simple Boolean filters, a stratagem browsing approach is developed and evaluated in which the search is aligned along with the *interaction context* of the user. This forms an implicit representation of the user's information need.

The underlying assumption is that the incorporation of the interaction context can lead to a new generation of search systems that can be created, designed, and developed to increase the performance of context-insensitive search systems (White, 2016). Today contextualisation is supported in a wide range of applications, such as query expansion (e.g. Chirita et al., 2007; Shen et al., 2005) and recommender systems (e.g. Adomavicius & Tuzhilin, 2011). Yet, employing the interaction context on the level of stratagem browsing has so far been underexposed.

In summary, the goal of this thesis is to enhance stratagems browsing by incorporating the users' interaction context and thus, tailor search results based on previous search activities. Instead of filtering the documents based on shared characteristics, as it is the current state-of- the-art in many DLs, a contextualised stratagem browsing is proposed that extends the basic filtering by re-ranking the results with respect to the users' interaction context.

Following these considerations, the underlying hypothesis under investigation in this thesis is:

Hypothesis: *The employment of the user's interaction context on the level of stratagem browsing will lead to the development of more effective retrieval systems.*

On the basis of this hypothesis, three research questions are derived which are discussed in the following section.

1.2.1 Research Questions

The present thesis aims to answer the following research questions.

RQ 1 What kind of stratagem browsing variants do users employ during information seeking, and how do the respondents assess their usefulness?
Problem: Although DLs are designed to facilitate stratagem browsing, the usefulness of these types of search activities has not been thoroughly evaluated empirically. Therefore, it is difficult to assess their value during information seeking. For instance, scanning the content of a renowned journal may undoubtedly lead to interesting and useful content. How frequent a user may employ such a strategy and whether it is useful for continuing a search after finding a relevant document, however, has not been assessed up to now.
Approach: The first research question is addressed by conducting an online survey in which the respondents are asked to rate the usefulness of different stratagem browsing variants. The research question aims to better understand the usage of different types of stratagems that were derived from the examples proposed by Bates (1990). The usefulness of certain stratagems is investigated with regards to the users' state of search which is distinguished as follows: a) the user has found a relevant document and wants to find similar documents and b) the user performs stratagem browsing without a preceding document.

RQ 2 What are the most frequently applied stratagem browsing variants in a state-of-the-art DL, and how is their usage in comparison to other search strategies like, for example, query searches? How is the perceived relevance of stratagem browsing opportunities?
Problem: Due to the nature of an online survey, the first research question is only suitable to provide a first impression on the empirically assessed value of stratagem browsing. A statement about the specific usefulness of stratagem browsing, however, is only possible by actually observing participants during an exploratory search task. Consequently, the aim of this research question is to quantify how frequent stratagem browsing is applied during an exploratory search. Are users rather (re-)formulating queries until they narrowed down their search towards relevant content, or are users relying on stratagem browsing, in order to get familiar with the topic?
Approach: The second research question is addressed by conducting a user study in which 32 participants are asked to solve a given search task. The study follows a scenario in which the participants are looking for related content to an already retrieved relevant document. Starting from a relevant seed document, the participants are observed with regard to the search tactics applied. In addition to measuring the frequency of stratagem browsing, an eye-tracking device is utilised, in order to measure the perceived relevance of stratagem browsing. The goal of utilising an eye-tracking device is to gain knowledge about a potential imbalance between the actual usage of a

stratagem (for example, a click on a keyword) and the perceived value of a stratagem (for example, fixating a keyword but not utilising it).

RQ 3 Can the effectiveness of exploratory search on the level of browsing be improved by employing contextual ranking features in comparison to a non-contextual ranking feature?
Problem: This is the central research question of the present thesis. While the first two research questions aim to better understand the users' intentions during stratagem browsing, this research question is dedicated to the analysis of potential benefits of contextualised stratagem browsing. By today, contextualisation has been mainly applied to query searches. In comparison, the second main activity during exploratory search, browsing is as of today underexposed.
Approach: The first step towards an answer to this research question is to implement contextualised stratagem browsing in a retrieval system. This is done in *Sowiport*, a DL for the social sciences. In order to measure the effectiveness of contextualised browsing, a transaction log study is conducted. Two contextualised stratagem browsing variants are developed which are compared against a non-contextualised baseline. The effectiveness of each approach is evaluated by measuring: a) the position of the first clicked item which is referred to as "mean first relevant" (MFR) (Fuhr, 2018), b) the click-through rate, and c) the usefulness in terms of implicit relevance feedback.

1.3 Contextualised Browsing in Digital Libraries

A schematic visualisation of contextualised stratagem browsing is presented in Figure 1.2. The figure shows a common sequence of interactions, which can be found in many DLs. In this example, a fictitious user is seeking information on the topic *violence and sports*.
After entering a query "violence" (1), inspecting two documents (2), and refining the query to "violence and sports", (3) the user has found a document of interest in a result set (4: highlighted green in the figure) that he or she inspects in detail (5 and 6). Each of the interactions (1 to 6) contributes to the development of the interaction context. To seek further related content, the user could now, for instance, click at the journal in which the document was published in or click on a particular keyword that is contained in the current document. Each of these interactions (7) leads to a new result list containing documents that share the same attribute with the seed document, which is also part of the result list (8). Contextualised stratagem browsing aims to re-rank these result lists on the basis of the interaction context in order to tailor search results towards an approximation of the user's information need.

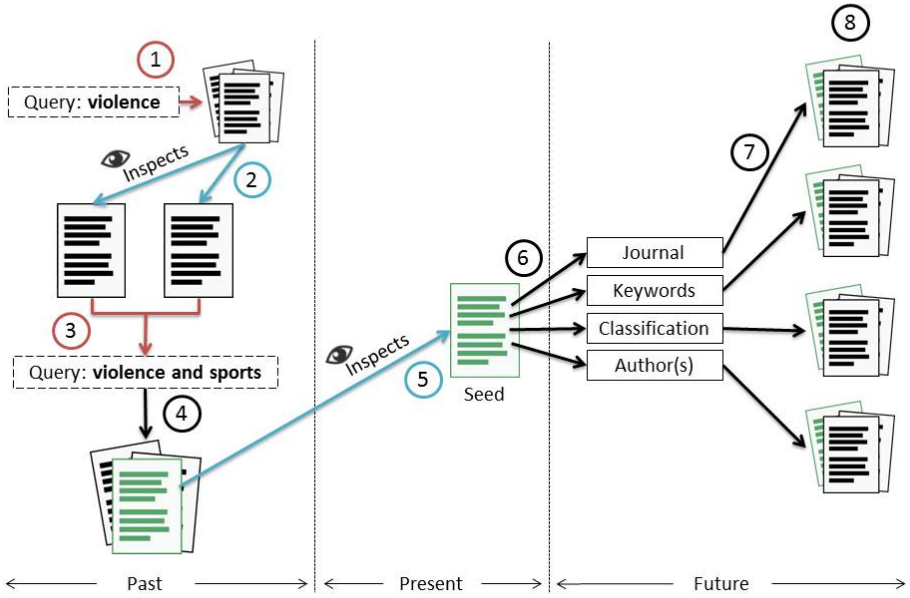


Figure 1.2: Schematic visualisation of contextualised stratagem browsing.

1.4 Contribution

This thesis has two main contributions.

The first contribution follows the principles of IIR by establishing a better understanding of stratagem browsing. The methods to construct this type of knowledge range from an online survey to a user study. The latter involves observations of the users' search strategies employed during the search task, gaze data provided by an eye-tracking device, as well as qualitative feedback on the usefulness of stratagem browsing. Both, the online survey and the user study contribute to research by providing empirical results for the usefulness of stratagem browsing in DLs.

The second main contribution is the implementation and evaluation of two contextualised browsing variants in a DL. The first contextualisation variant bases on document similarity and the second variant bases on the interaction context that, amongst others, contains information about submitted queries and different document metadata encountered during the session of a user. The results of this study show that a contextualisation of stratagem browsing significantly outperforms the baseline in terms of the position of the first clicked document in the result set.

First and foremost, the results of this experiment contribute to research by pro-

viding experimental confirmation for a significant increase in the effectiveness of stratagem browsing by employing contextualisation methods. Furthermore, both contextualisation approaches are developed on the basis of existing methods that are well-known in IR, such as implicit relevance feedback, and thus can be easily reproduced by designers of DLs.

To the best of my knowledge, a contextualisation of stratagem browsing has not been studied yet. Thus, this thesis serves as a foundation for the field of contextualised stratagem browsing that allows for algorithmic reproduction by other researches in this field.

1.5 Publications

The present thesis is based on the following publications, which subsequently form the core chapters of the thesis.

(P1) TPDL 2016: Zeljko Carevic and Philipp Mayr. "Survey on High-Level Search Activities based on the Stratagem Level in Digital Libraries." In proceedings of the 20th International Conference on Theory and Practice of Digital Libraries (pp. 54-66). Springer, Cham, 2016. (Carevic & Mayr, 2016).

(P2) IJDL 2018: Zeljko Carevic, Maria Lusky, Wilko van Hoek and Philipp Mayr. "Investigating Exploratory Search Activities based on the Stratagem Level in Digital Libraries." In International Journal on Digital Libraries (pp. 231-251). Springer, 2018. (Carevic et al., 2018a).

(P3) JCDL 2018: Zeljko Carevic, Sascha Schüller, Philipp Mayr and Norbert Fuhr. "Contextualised Browsing in a Digital Library's Living Lab." In proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries (pp. 89-98). (Carevic et al., 2018b).

In addition to the core publications, the following list of publications has had an impact on the results of this thesis.

(P4) BIR 2014: Zeljko Carevic and Philipp Schaer. "On the Connection Between Citation-based and Topical Relevance Ranking: Results of a Pretest using iSearch." In proceedings of the 1st Workshop on Bibliometric-enhanced Information Retrieval (pp. 37-44), 2014. (Carevic & Schaer, 2014).

(P5) BIR 2015: Zeljko Carevic and Philipp Mayr. "Extending search facilities via bibliometric-enhanced Stratagems." In proceedings of the 2nd Workshop on Bibliometric-enhanced Information Retrieval (pp. 40-46), 2015. (Carevic & Mayr, 2015)

(P6) TPDL 2017: Wilko van Hoek and Zeljko Carevic. "Building user groups based on a structural representation of user search sessions." In proceedings of the 21th International Conference on Theory and Practice of Digital Libraries (pp. 459–470). Springer, 2017. (van Hoek & Carevic, 2017)

1.6 Structure

The thesis is structured into four parts. These are described briefly in the following:

Part I: Fundamentals and Related Work
The first part of the thesis is primarily dedicated to the introduction of fundamental theoretical concepts necessary for a better understanding of the thesis. This part is divided into three sections. Section 2 is dedicated to introducing concepts and models of information seeking. While there is a plethora of models in this field, the focus is on those models most relevant for the present thesis. In Section 3, the notion of context is defined on the basis of the nested model of context stratification. Fundamental technical means necessary for a practical realisation of contextualisation are introduced in the form of implicit and explicit relevance feedback. Finally, Digital Libraries (DLs), which represent the use case of the present thesis, are discussed in Section 4. A definition of DLs is presented, followed by a discussion on two central models that provide the theoretical foundations of DLs: a) the DELOS reference model and b) the 5S model. Thereafter, the notion of metadata in the context of DLs is briefly explained. Finally, Sowiport is introduced, which is a real-life DL employed for the investigation of contextualised stratagem browsing.

Part II: Empirical Studies on Browsing in Digital Libraries
This covers the results of the two empirical studies and is divided into two sections: An online survey was conducted in order to investigate the usefulness of stratagem browsing. A detailed presentation of the survey design and the corresponding results can be found in Section 5. In the second empirical study, an observational study on stratagem browsing behaviour was conducted. A total of 32 participants were asked to solve a task on searching related documents to a given topic. The detailed setup and the results of this user study are presented in Section 6.

Part III: Towards Contextual Browsing in Digital Libraries
Based on the observations made in the survey and the user study, two contextualisation approaches are presented in the third part of this thesis. This part is structured into two sections. In Section 7, a detailed explanation and implementation of the two contextualisation approaches is presented. Both approaches are evaluated and compared against the non-contextual baseline of Sowiport by employing A/B/C testing with online users. The results of this experiment are presented in Section 8.

Part IV: Discussion and Conclusion
The fourth part of this thesis presents a thorough discussion of the results provided in Part II and Part III. In addition, the potential strengths and weaknesses of each study are discussed, along with possible implications for the design of DLs and future work.

Part I

Fundamentals and Related Work

2

Concepts and Models of Information Seeking

In Information Retrieval (IR) a change can be observed that moves away from a mainly system-oriented perspective that is aligning with the *Cranfield* paradigm towards a more user-oriented perspective which is referred to as Interactive Information Retrieval (IIR). The principles of IIR extend the traditional query and response or *best-match* principles to a highly dynamic and interactive process that puts the user into the centre of attention. This user-centred perspective of information seeking involves research and development with respect to various complex processes such as the concrete actions and tactics that a user employs (e.g. queries, navigation, browsing), and cognitive (e.g. information need) and emotional aspects (e.g. uncertainty, confusion) which are experienced throughout the different stages of the search process. During the past decades, a considerable amount of research has been invested in the development of theoretical models that capture and formalise the various dimensions involved in information seeking. Although many of the models available today were developed several years ago, often by observing people during physical library searches, they still have a substantial impact on fundamental aspects of today's modern search systems. Hence, gaining a better understanding of the different aspects involved is crucial for the development of effective information systems and evenly important for the objective of the present thesis, namely the development of contextualised browsing capabilities.

2.1 Information-Seeking Behaviour

In order to differentiate the models present in this section and to determine their particular scope, it is necessary to first introduce and distinguish three of the most relevant conceptual levels from a top-down perspective. Wilson (1999) suggested a nested model of information behaviour, displayed in Figure 2.1. In this model, information behaviour is considered as the general field of investigation which resides on the outer rim of a nested model incorporating information-seeking behaviour and information search behaviour.

Wilson defines **information behaviour** as "the totality of human behaviour in relation to sources and channels of information, including both active and passive information seeking, and information use." (Wilson, 2000, p. 49). Following this

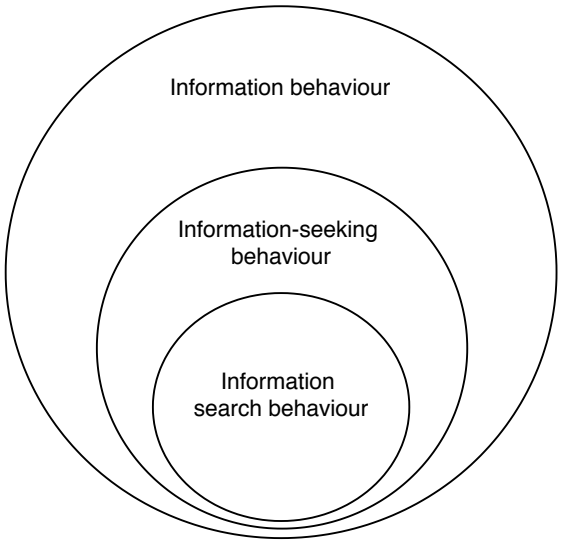


Figure 2.1: Nested model of information behaviour. Adapted from Wilson (1999).

rather generic definition, information behaviour represents the highest degree of abstraction, which is also visible in the illustration in Figure 2.1.
Information-seeking behaviour is considered a sub-set of information behaviour that involves the purposive seeking for information in order to satisfy an information need. To satisfy an information need, one could interact with one or more channels such as talking to colleagues, go to a library or use any form of computer-based system such as a Digital Library (see Section 4) or a web search.
Finally, **information search behaviour** comprises the behaviour that a user employs while interacting with an information system. This involves the concrete interactions such as querying, browsing or judging the relevance of a document. Information search behaviour thus represents the micro-level of information behaviour.

A modified, and in the context of this thesis more accurate, variant of Wilson's model is proposed by Ingwersen & Järvelin (2005b). The central modification of the original model is the replacement of information search behaviour by the notion of (I)IR. This is displayed in Figure 2.2.

Wilson associated (I)IR as a part of the information search behaviour and hence, on the micro-level, which is isolated from information-seeking processes. The modified model of information behaviour, however, argues that in fact many (I)IR studies incorporate both, the concrete interactions with an IR system, and more informal information-seeking activities that are usually incorporated within information-seeking behaviour.[2] Thus, the intuition behind the modified model is that although

[2] A similar argumentation can be found in Belkin (1993).

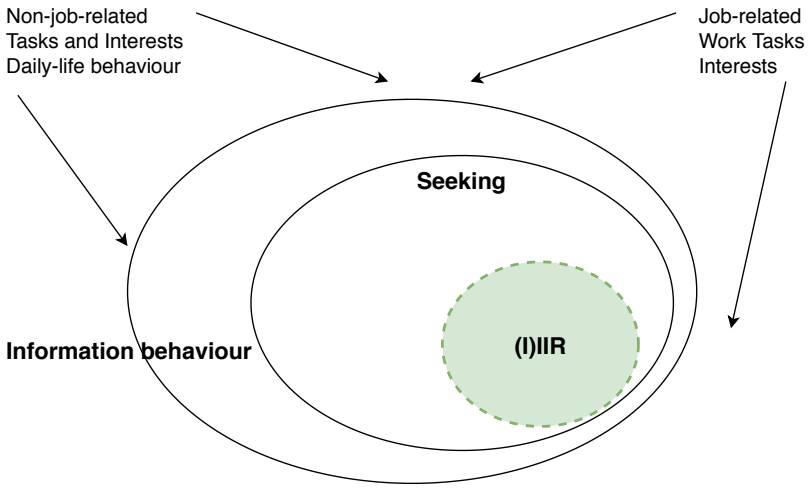


Figure 2.2: Modified version of Wilson's nested model of information behaviour incorporating I(IR). Adapted from Ingwersen & Järvelin (2005b).

it is reasonable to place IR and IIR in the innermost nested part of the model, they should also be considered as a component of information-seeking behaviour.
In addition to incorporating and emphasising the role of IIR, Ingwersen & Järvelin (2005b) incorporated the underlying situational reasons for any information behaviour which are described as non-job related and job-related in Figure 2.2. These situational reasons are closely related to the notion of *work task*, which is discussed in Section 2.6.1. With respect to the nested model of Ingwersen and Järvelin, this thesis is allocated within the dimension of information seeking explicitly incorporating (I)IR related aspects and any other interactions related to the purposive seeking for information.

So far, information seeking was introduced from a highly generic point of view. A more in-depth discussion of information seeking requires the introduction of models that focus on different perspectives which are always dependent on the characteristics of interest and the intended degree of abstraction. Introducing and covering all the different developments in this field is a challenging task, and thus, it is reasonable to introduce this topic with an overview of those models that are essential for the present thesis. These are displayed in Table 2.1 with many of the given attributions inspired by White (2016). Table 2.1 shows the different models discussed in the following sections and provides a general attribution regarding the nature of the discussed models.[3]

The upcoming sections are organised according to the list of attributions provided in column one of Table 2.1.

[3]This list is by no means comprehensive, and there exist a plethora of other models that would be worth mentioning. However, here only the models relevant for this thesis are presented.

Table 2.1: Models of information seeking.

Process-oriented	Behavioural Model of Information Seeking Strategies (Ellis, 1989) Model of the Information Search Process (Kuhlthau, 1991)
Stratified	Stratified Model of Interaction (Saracevic, 1996)
Strategic Information Gathering	Strategic Levels of Information Seeking (Bates, 1990) Berrypicking Model (Bates, 1989)
Exploratory	Exploratory Search (Marchionini, 2006)

2.2 Process-oriented Models of Information Seeking

Process-oriented models of information seeking disintegrate the search process into multi-stage phases that users undergo while seeking for information. In the following, two prominent examples of process-oriented models are introduced: a behavioural model of information seeking strategies by Ellis, and Kuhlthau's model of the information search process.

2.2.1 Ellis' Behavioural Model of Information Seeking Strategies

Ellis (1989) presented a behavioural model of information seeking strategies. Based on semi-structured interviews that were conducted with social scientists, he observed six generic search features. During the interviews, the participants were encouraged to describe their work and the activities employed during information seeking. On the basis of these interviews, Ellis derived the following six generic features which were partly or entirely present in the interviews.

- *Starting*
  The starting feature describes the initial state of seeking information in a new area or on a new topic. A common activity, belonging to the starting feature, is looking for a key paper to start the search. The information seeker is often either already familiar with this key paper or is told of by, for instance, colleagues who are more familiar with the topic.

- *Chaining*
  Chaining describes search activities that involve following citation connections between resources and thus, can be of two kinds: *forward chaining* (to inspect a list of documents citing a particular document) and *backward chaining* (to follow a reference in a document).

- *Browsing*
  Browsing is defined in various ways. However, in this context, it is referred to as a semi-directed or semi-structured search in an area of potential interest.

Examples for browsing are looking through the table of content in a well-known journal or looking through conference proceedings. It involves the scanning of resources that represent topically connected content.

- *Differentiating*
  To estimate the quality or nature of a source in an area of interest. An experienced researcher may be aware of specific journals that contain mature work or a high degree of specialisation and thus, may prefer such a source over different not well-known sources.

- *Monitoring*
  To continuously monitor or track the development in a particular area of interest. Today, various types of monitoring services exist, such as *RSS-Feeds* or *Current Awareness Services*. By subscribing to a topic of interest, one can get updated as soon as new potentially relevant material is available.

- *Extracting*
  Extracting refers to activities in which a user selectively examines content from a source to identify relevant material on a particular topic of interest. The source may be, for instance, conference proceedings, the content of an archive or journals.

For each of these features, Ellis discussed their impact on the design of retrieval systems. Given the feature *browsing*, for instance, he argued that it is desirable for searchers that a system should enable browsing on any information directly accessible, such as browsing through a list of authors or journals and conference proceedings. Today, many of these considerations are integrated into modern retrieval systems. This will be demonstrated at different points throughout this thesis (e.g. Section 4.3).

It is important to note that this model should not be understood as a hierarchic sequence of actions with strict interrelations. The order of the different features and their interrelations rather depend on the unique circumstances of the information-seeking activities (Ellis, 1989). Nevertheless, Wilson (1999) developed a representation that relates the different features to each other into a specific sequence.
In 2003, the model developed by Ellis was revised by Meho & Tibbo (2003). They found four additional features besides the ones proposed by Ellis: *accessing*, *networking*, *verifying*, and *information managing*.

2.2.2 Kuhlthau's Model of the Information Search Process

Kuhlthau (1991, 1993) provided another process-oriented model. Based on a series of five studies, she developed a model of the *information search process* (ISP) displayed in Figure 2.3. The ISP model consists of six stages: *initiation*, *selection*, *exploration*, *formulation*, *collection*, and *presentation*.

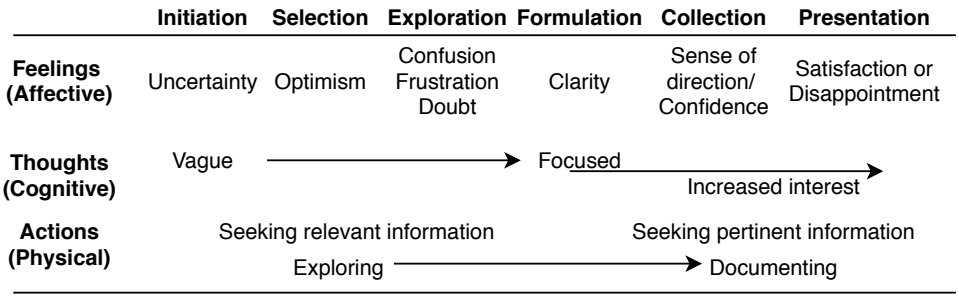


Figure 2.3: Model of the information search process. Figure adapted from Kuhlthau (1993).

The novelty of this model, in comparison to previous approaches, is that it incorporates three realms of human experience: *affective* (feelings), *cognitive* (thoughts), and *physical* (actions), which are now briefly described along the six stages of the ISP.

During the **initiation** stage, the user identifies a lack of knowledge which leads to a feeling of uncertainty. On the **selection** stage, a user has to identify the general topic to investigate and select the approach to be followed. The feeling of uncertainty is often replaced by a brief feeling of optimism after the selection has been made. The task during the **exploration** stage is to extend the personal knowledge about the topic that is being investigated. Actions during this stage involve locating information, reading about the topic, and connecting new information to existing knowledge. This stage is often characterised by a feeling of confusion, frustration, and doubt. The inability to formulate the information need makes the communication between the system and the user complicated. During the **formulation** stage, a focus is formed based on the information encountered so far, which is comparable to a hypothesis that is being constructed. During this stage, a feeling of increased confidence and a sense of clarity can be observed. In the **collection** stage, the process of interacting with the system is most effective and efficient. During this stage, the user is collecting information that is relevant to the topic of interest. The user is capable of formulating his or her information need to the system and thus, facilitates a more comprehensive search. During the final stage **presentation**, the search is completed and the task is to prepare, present or otherwise use the information gathered. At this stage, a feeling of relief and a sense of satisfaction can be observed.

A unique feature of this model is the incorporation of the affective experience of a user during different stages of the ISP. White (2016) also argues that emphasising the different feelings and actions through different stages of the ISP can help to provide phase-appropriate support like user guidance. One shortcoming of Kuhlthau's model of the ISP is discussed by Ingwersen & Järvelin (2005b) in which it is ar-

gued that her model claims to be applicable over a range of empirical domains but that it may depend on the kind of work task on which it is based. Thus, certain work tasks may not be represented by her model.

2.3 Saracevic's Stratified Model of Interaction

Saracevic (1997) introduced a stratified model of interaction, which is displayed in Figure 2.4. By borrowing concepts from Human-Computer Interaction, the stratified model separates levels of interaction in IR between the computer and the user, connected through the interface on the surface level. The strata involved on the user side comprise a cognitive, affective, and situational level, and on the other hand engineering (hardware), processing, and content level on the system side. One of the key aspects of his model is that the effectiveness of an information-seeking episode is not exclusively dependent on either, the user or the system side. Having a highly efficient retrieval model, for instance, may not be sufficient if the user interface is incapable of providing this information to the user. Similarly, having a user who is not capable of expressing his or her information need will most likely not lead to an effective information-seeking episode, no matter the quality of the system side.

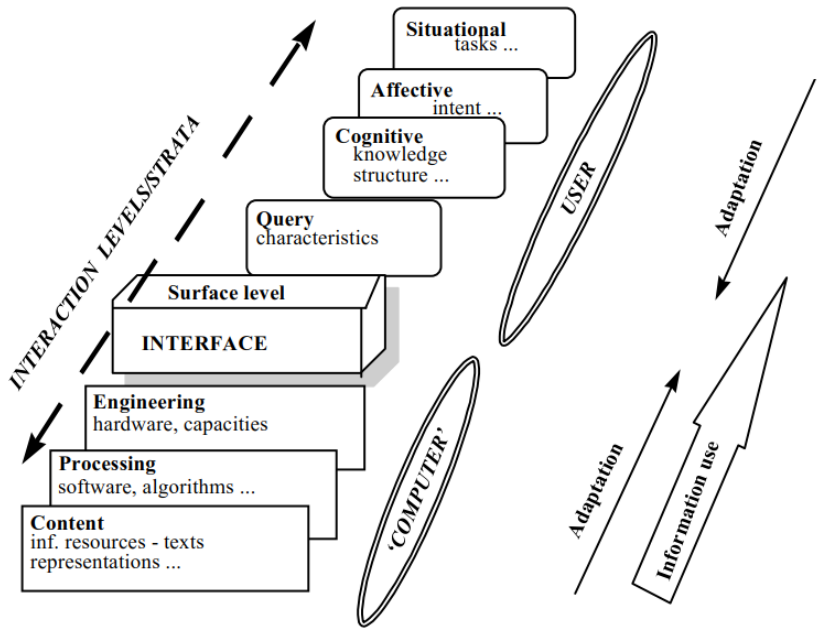


Figure 2.4: Saracevic's stratified model of interaction (Saracevic, 1997).

In line with this, Saracevic (1996) provided a reconsideration of a so-called "system of relevances" which, similar to the stratified model, operates and interacts on different strata. The different manifestations of relevance are: *system or algorithmic*, *topical or subject*, *cognitive relevance or pertinence*, *situational relevance or*

utility and *motivational or affective* relevance. Each of these manifestations considers relevance from a different perspective: situational relevance, for instance, is the relation between the task, situation or problem currently experienced and the texts retrieved. Topical or subject relevance, on the other hand, is the relation between the subject expressed in a query and the topic/subject covered by the retrieved text. These manifestations are not separate from each other, but rather a dynamic interdependent set of relevances that interact within and between themselves. The topical relevance, for instance, is usually derived from the set of retrieved items and thus, based on the system or algorithmic relevance.

2.4 Bates' Strategic Levels of Information Seeking

An approach to model the different search strategies that users employ during information seeking has been proposed by Bates (1990). Based on empirical studies of the information-seeking behaviour of experienced users of a physical library, Bates identified four levels of search activities that, amongst others, differ in their complexity: moves, tactics, stratagems, and strategies.
A **move** is the lowest unit of search activities and usually encompasses simple operations such as entering a query term or selecting a certain document. **Tactics** are described as a combination of many moves. A move could be the selection of a broader search term, or breaking down complex search queries into sub-problems. Bates defines a **stratagem** as follows: "..a stratagem is a complex of a number of moves and/or tactics, and generally involves both a particular identified information search domain anticipated to be productive by the searcher, and a mode of tackling the particular file organization of that domain." (Bates, 1990, p. 6).
Finally, a **strategy** is a combination of moves, tactics, and stratagems that satisfies an information need like, for instance, searching for related work in a specific research area. In Bates' model, strategies account for the highest level of search activities and often span over a longer period of time.

Especially the considerations regarding stratagems have been highly influential for this thesis making these types of searches the theoretical foundation for all three main chapters of the present work. Although Bates already provided an informal definition of stratagems, it is useful to explain these types of searches in more detail.

The two main components of Bates' definition are: 1) that a stratagem consists of an identified search domain anticipated to be productive for a searcher and 2) that a stratagem involves a mode of tackling the particular file organisation of that domain. The first part of her definition involves a domain that is expected to be productive. This could be, for instance, a particular journal which according to the user's experience is known to provide relevant content. The second part of her definition involves the tackling of a particular file organisation of that domain.

This could involve to browse and read through the articles of the identified journal. This implies that a searcher is familiar with the structure of the underlying domain. Other examples of stratagems that were provided by Bates are as follows:

- *Footnote chasing or backward chaining*
  This technique uses references and footnotes found in books or articles moving backwards in successive trails of interest. A user having identified a relevant article may use the list of references to find other articles related to the current topic. These types of search activities have been subject to research in recent years. One example is the *ISearch test collection* (Lykke et al., 2010) which covers articles from the preprint repository *arXiv.*[4]ISearch is a classic IR test collection that contains topics, a document corpus, and graded relevance judgements which are provided in combination with references for each article.

- *Citation search or forward chaining*
  In this technique, a user discovers a relevant article and looks up the list of articles citing that particular one. In contrast to footnote chasing, this technique leaps forward when looking for related articles. Both, forward and backward chaining are also found in Ellis' behavioural model on the chaining level (see 2.2.1).

- *Journal run*
  One identifies a journal of interest and browses through volumes and issues. Bates argues that a journal run has the advantage of usually covering a large number of publications to a particular subject, and thus provides good precision and in many cases, a good recall. As the number of publications and journals has increased in recent years, recall and precision may suffer from the now increased amount of content.

- *Area scanning*
  Given an area of interest, a user can look up other articles that are indexed in the same general area. Today, this type of search technique often includes the usage of a thesaurus (e.g. TheSoz see Section 4.4.1).

- *Subject searches in bibliographies, abstracting and indexing services*
  Topical descriptions of an article are often provided by an indexing subject and usually referred to as keywords. Keywords can be provided in a controlled way (e.g. the *Unified Medical Language System* indexed in PubMed[5]) or by free descriptors chosen by the authors of an article. Given this kind of representation, a user can employ keyword searches to retrieve those documents that share the same attribute.

[4] `https://www.arxiv.org/`, last accessed April 15th, 2020.
[5] `https://www.ncbi.nlm.nih.gov/pubmed/`, last accessed April 15th, 2020.

- *Author search*
  Given an article on a particular subject, a user can employ an author search to look for other articles of that particular author on the same topic. The assumption is that an author may have contributed to the same field in the past.

One may recognise that several of her search activities on the stratagem level were influenced by the work of Ellis (1989) (see Section 2.2.1).
Today, these search activities are widely supported in modern Digital Libraries (see Section 4) such as Daffodil, one of the first implementations of the ideas suggested by Bates (Fuhr et al., 2002).

2.5 Bates' Berrypicking Model

The earlier introduced model by Bates emphasised the strategic nature of information seeking. In 1989, Bates introduced the so-called *berrypicking* model, which emphasises the dynamic nature of information seeking and information gathering (Bates, 1989). The berrypicking model advances from traditional IR models by considering the process of information seeking as dynamic and evolving with each interaction leading to potentially new relevant results, ideas, and directions. An illustration of a user involved in a berrypicking search is displayed in Figure 2.5.

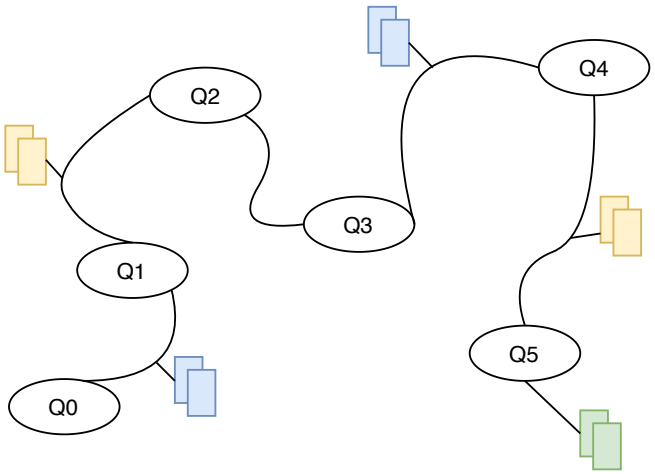


Figure 2.5: Schematic visualisation of a user engaging in a berrypicking search. Figure adapted from Hearst (2009).

The underlying assumption of this model is that a query is not a single solitary conception but rather an evolving one that changes throughout the course of interaction with each new piece of information potentially altering the conception of the query and redefining what is considered as relevant. In Figure 2.5, this is depicted by the query variations (Q0 to Q5) and the documents encountered along the path.

Bates developed this model with respect to the classic IR model accentuating two major distinctions: a) that the nature of a query is an evolving one instead of a static as suggested in the classic IR model and b) that the nature of the search process is following a pattern of berrypicking, rather than leading to the retrieval of a single best set (Bates, 1989). Following these considerations, one can conclude that information seeking is an evolving process in which each action has the potential to change the user's understanding of a particular information need which may even shift entirely during information seeking.

Although the berrypicking model seems reasonable to describe the way how people interact with search results in a metaphorical way, it does not provide deeper insights into the reasons why people behave like this, or for how long a searcher would stay on one path (Azzopardi, 2014).

Bates is not the only one who used an analogy such as berrypicking to describe information seeking. The information foraging theory, for instance, proposed by Pirolli and Card, is an attempt to analyse the trade-offs in the value of information gained against the costs of performing a certain search activity (Pirolli & Card, 1995). The assumptions of the information foraging theory are derived from the optimal foraging theory, which analyses the behaviour of animals when searching for food. Animals that forage for food are optimising their behaviour in order to reduce the cost of hunting while fitting their needs to survive.

2.6 Exploratory Search

Some of the most renowned models of information seeking have been introduced in the previous sections emphasising the theoretical aspects of a user-oriented perspective on information retrieval. Exploratory search represents another type of information seeking and a type of sense-making (White & Roth, 2009). According to Marchionini, searches considered as exploratory search go beyond simple lookup tasks such as known-item searches and usually operate on the level of learning and investigating (Marchionini, 2006). A general classification of exploratory search, its main goals of learning and investigating, and its relation to lookup searches are depicted in Figure 2.6.

Lookup tasks are frequent search activities in today's life and include short and well-supported tasks such as known-item searches (lookup a document for which the title is known) or question answering (e.g. "what is the population of germany" which is supported by for instance the Google fact box[6]).
More complex tasks, like **learning**, involve search activities that require a much longer period of time and involve cognitive processing such as comparing and

[6] `https://www.google.com/search?q=what%20is%20the%20population%20of%20germany`, last accessed August 29th, 2019.

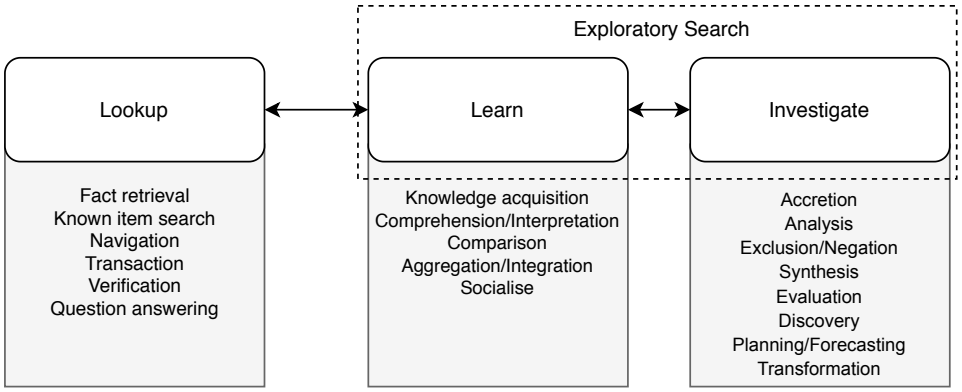


Figure 2.6: Types of exploratory search activities. Figure adapted from Marchionini (2006).

making qualitative judgments, in order to acquire knowledge about a certain topic. According to Marchionini, learning is best achieved by a combination of browsing and analytic strategies including lookup searches to "get one into the correct neighborhood for exploratory browsing" (Marchionini, 2006, p. 43).
Search activities that involve **investigation** require substantial knowledge in order to achieve high-level objectives such as analysis, synthesis, and evaluation. Investigative searches may require long periods of time. They may be done in order to support planning and prediction or to transform existing data into new data or knowledge. From these explanations, it is concluded that learning and investigating require strong human participation in a continuous and exploratory way (Marchionini, 2006).

Although the three categories, lookup, learning, and investigating in Figure 2.6 are represented on distinct levels, they are to be understood as interplaying, which is depicted by the arrows. An exploratory search that is aimed to learn about a particular topic may involve several lookup activities, as well as investigation.

White et al. (2006) introduced exploratory search as a specialised form of searching in which a user has only little knowledge about a certain field of interest and after an initial search (e.g. via an initial tentative query) starts exploring the retrieved documents passively, in order to obtain cues about next steps. Moreover, exploratory search must include complex cognitive activities that are associated with the acquisition of knowledge and the development of intellectual skills (White & Roth, 2009).

Two main activities can be identified during exploratory search episodes: exploratory browsing and focused searching. **Exploratory browsing** within document collections is performed, to better define the information need of a user and to promote new directions and cognition on the basis of encountered con-

tent. Browsing activities on the web usually occur between hyperlinked pages and may support the generation of a hypothesis (White & Roth, 2009). During **focused searches**, users formulate queries to explore the document collections. These types of searchers usually require an understanding of the terminology and a rather well-defined information need.

Although exploratory search can be considered a subdiscipline of information seeking, White & Roth (2009) demonstrated six characteristics of exploratory search that differ from other types of information seeking and related disciplines. First and foremost, exploratory searchers are often driven more by things like curiosity and learning rather than by a concrete information need (White, 2016). Additional attributes differentiating exploratory search from other types of information seeking can be found in White & Roth (2009). Amongst other, exploratory search

- may involve multiple query iterations possibly, over multiple search sessions,
- are driven by an open-ended information need which relates to uncertainty about the information available or incomplete information on the search task itself,
- have a goal that goes beyond simple lookup tasks towards activities that are associated with learning and investigating,
- involve a combination of browsing and focused searching with a stronger focus on the former,
- may involve a collaborative setting in which multiple people interact synchronous or asynchronous,
- are using an evaluation methodology that targets learning and insights.

At a workshop held at the University of Maryland, researchers from the fields of Information Retrieval, Human-Computer Interaction and Information Visualisation came together to discuss issues of exploratory search. One of the outcomes of this workshop was that exploratory search needs to be contextualised based on the users' search activities and contextual information available to aid user understanding and reduce uncertainty about the nature of the problem (White et al., 2005).

2.6.1 Simulated Work Tasks in Exploratory Search Studies

The design of user studies in the area of exploratory search is a challenging task due to the goal of inducing an exploratory and undirected search in which the participants assess the task, relevance, and the results from their own individual

perspective. At the same time, the study should maintain some form of experimental control (Kules & Capra, 2012).

One suitable approach for designing studies on exploratory searches is to employ so-called *simulated work tasks*, which were introduced by Borlund & Ingwersen (1997). Simulated work tasks are short cover stories that aim to trigger and develop a simulated information need that provides the participants of a study with the given purpose and goal of the retrieval. The cover story is described in a somewhat open manner that helps to describe to the test persons: the source of the information need, the environment of the situation, the problem to be solved, and generally serves to provide the test person with an understanding of the objective of the search (Borlund, 2003).

An example of a simulated work task, that was taken from (Borlund, 2000, p. 78) could be as follows:

> **Simulated situation:** After your graduation you will be looking for a job in industry. You want information to help you focus your future job seeking. You know it pays to know the market. You would like to find some information about employment patterns in industry and what kind of qualifications employers will be looking for from future employees.
>
> **Indicative request**: Find, for instance, something about future employment trends in industry, i.e., areas of growth and decline.

Simulated work tasks are described in a way that frames the situation of the search task with sufficient clarification. Furthermore, they provide experimental control by being unique across all participants, while at the same time, over-clarification is avoided. Thus, simulated work tasks offer enough space for an individual to adjust to the given situation, in order to derive their own individual information need from that particular scenario, which is comparable to a real-life situation.

Although simulated work tasks were not originally developed with respect to exploratory search, they are well suited to be applied in this area of research. Exploratory search tasks are usually ill-structured and open-ended involving uncertainty. Thus, an over clarified situation, which is explicitly avoided through simulated work tasks, would neither be appropriate nor applicable in the context of exploratory search tasks.

2.7 Advances in Information Seeking

A considerable amount of research has been carried out, in order to better understand information-seeking behaviour. Studies of this kind are often rather focused on exploring and/or describing users' natural search behaviours and interactions, than on demonstrating the performance of a feature or a system.[7] The methods employed to study information-seeking behaviour often follow a multi-method design that, amongst others, involves: questionnaires, semi-structured interviews, transaction logs, and (unobtrusive) observations of search behaviour. For a comprehensive overview of methods that are applicable for studies on information seeking, one can refer to Case (2002).

A selection of research related to the models presented earlier in this section and the upcoming empirical studies (see Section 5 and Section 6) is discussed in the following. First, studies are presented that deal with information-seeking behaviour in the area of academic literature search. The second set of related work is concerned with studies on exploratory search.

Academic Literature Search

Academic literature searches represent a frequently explored field of research on information-seeking behaviour (e.g. Niu et al., 2010; Pontis & Blandford, 2015; Pontis et al.). The search for academic literature usually goes beyond simple lookup tasks and is often driven by complex information needs, such as learning about a new topic or looking for opportunities for collaboration.

To investigate the information-seeking behaviour of academics from the computer science domain, Athukorala et al. (2013) conducted a mixed-method study involving interviews, diary logs, user observations, and a web survey. The study was divided into two phases: 1) a case study phase, and 2) a web survey aiming to generalise the results of the case study. The case study involved six participants with different academic degrees: Ph.D. students, postdoctoral, and senior researchers. The case study consisted of three parts: 1) an interview about search methods and search strategies, 2) an observation phase in which the participants were observed during information-seeking for a real purpose work task, and 3) a longitudinal diary study, in which the participants were asked to keep record of their search behaviour during information search. The second phase was designed to validate the results of the case study by conducting a web survey with a larger population of 76 computer scientists. From the case study, Athukorala et al. identified six purposes that motivate literature search: keeping up to date, exploring new and unfamiliar topics, reviewing literature, collaboration, preparing lectures, and recommending material for students. The web survey was designed along these six purposes to

[7] A distinction of different research goals can be found in (Kelly, 2009, p.25).

determine their respective frequency and difficulty. Athukorala et al. showed that keeping up to date is the most frequent purpose of searching for computer scientists. The participants considered the exploration of an unfamiliar topic as the most challenging search task. Furthermore, they showed that backward-chaining is the most frequently used literature review technique.

Hemminger et al. (2007), conducted a census survey to investigate the information-seeking behaviour of academic scientists. A total of 902 subjects from general science and medical science departments took part in the survey. Amongst others, they studied researchers sources of information, and how frequently they are used. The results showed that the most frequently used sources of information were academic journals, web pages, and online database. Considering that the work was published in 2007, some of these results may well have changed by today.

A more recent work investigating academic search behaviour was published by Li & de Rijke (2019). The focus of this work was to determine interaction behaviour that led to conversions, which is an indicator of the performance of a platform frequently applied in online shopping and other e-commerce domains. In the context of academic searches, they understood conversion as the download of a paper. Among others, they studied sequences of user actions that led to the download of an academic paper using transaction log data provided by ScienceDirect.[8] They observed that the most common sequence of interaction prior to the first download was a single query which made up for 30.3% of all sequences. Furthermore, behavioural differences with respect to the article topics (health sciences, life sciences, physical sciences and social sciences) were examined. The results of this analysis showed that users interested in social sciences, in comparison to the other disciplines, perform the least number of downloads while, at the same time, they account for more clicks and spend more time in sessions. Behavioural differences across different academic (sub-)disciplines were also observed in (Jamali & Nicholas, 2010; Tenopir et al., 2009).

Wildemuth (2004) examined the search tactics of medical students searching a factual database in microbiology. The students searched a medical database at three occasions over a period of nine months. Each of the three occasions represents different levels of knowledge: the first occasion, before the students received any instruction in microbiology, the second occasion just after a course on microbiology, and the third occasion six months after the end of the course. On each occasion, the participants were asked to respond to six clinical problems in microbiology. Subsequently, the participants were asked to respond to those questions that could not be answered by searching a factual database. The transaction log of these searches was analysed. The results show that the most common search tactic was the specification of a concept followed by extending one or more concepts and gradually narrowing the retrieved result set. Furthermore, they showed that domain knowledge affects search behaviour. Low domain knowledge was associated with less efficient selection of search terms.

[8] `http://www.sciencedirect.com/`, last accessed April 5th, 2020.

Exploratory Search Tasks

Athukorala et al. (2016) conducted a user study with 32 participants from the computer sciences domain to distinguish between exploratory and lookup search tasks. The main objective of the study was to collect and examine search behaviour from lookup and exploratory search tasks, in order to investigate how well these two types can be distinguished. To characterise a search session, different features were used, like, for instance, task completion time, maximum scroll-depth and length of the first query. Athukorala et al. (2016) showed that exploratory searches take longer time to complete, had a higher scroll-depth and involved shorter queries than in lookup tasks. Using an eye-tracking device, they analysed the participants gaze data, which showed only minor differences between lookup and exploratory search tasks.

Kules et al. (2009) examined how users interact with an online public access catalogue when conducting exploratory searches. Their method involved eye-tracking, stimulated recall interviews, and observations, to learn what parts of the faceted interface searchers attend to, for how long, and in what order. The authors defined *areas of interest* (AOI), which included search result pages, facets, and queries. They showed that the most time is spent on inspecting the search result pages (50 seconds per task), followed by facets (25 seconds), and queries (6 seconds). The results suggest that facets play an important role during exploratory search. However, the authors acknowledge that the exploratory tasks provided involved topics that were well-matched by the facets. Hence, the results of the AOI analysis might differ for tasks in which the facets are not well-matched.

Kang & Fu (2010) compared the search behaviour of experts and novices when searching in a traditional search engine and a social tagging system. They recruited 48 participants who were asked to solve an exploratory search task by using Google and the social bookmarking service Delicious.[9] They showed that the group of experts used queries more often than novices. At the same time, novices employed tag-based queries more frequently. The authors assume that experts are more likely to conduct queries from existing knowledge. The group of novices, on the other hand, rely on existing information in the environment that facilitates browsing on a predefined terminology comparable to stratagem browsing.

2.8 Conclusion

This section introduced some of the most renowned models of information seeking, sub-divided by their respective attributions such as process-oriented, strategic, and exploratory models. Even though each of the models plays a vital role in establishing the theoretical foundations, three models are of particular relevance for

[9] This service was not available anymore by the time of writing.

the present thesis. First and foremost, the strategic model developed by Bates is highly relevant, because it outlines the theoretical consideration of *stratagems* (see Section 2.4).
The process-oriented behavioural model of information seeking strategies by Ellis (2.2.1) determines different stages of the information-seeking process. In particular, the initial stages: *starting*, *chaining*, and *browsing*. These stages describe the phases of the information-seeking process that are addressed in this thesis.
The concept of exploratory search (see Section 2.6) introduces another direction of information seeking that puts more complex search activities into the centre of attention: learning and investigating both incorporating two elementary search activities *searching* and *browsing*. The latter is usually aiming to resolve uncertainty and confusion, which often occurs when new information is encountered (White, 2016).
From these three models, it is concluded that stratagems represent a variant of exploratory browsing allocated in the area of academic searches whose exploratory nature is reflected in the initial stages, starting, chaining, and browsing, of Ellis' model.

3

Contextualisation

The notion of *context* is used in various disciplines with different meanings and characteristics. Thus, studying the numerous definitions of what is understood as context can be exceedingly difficult. In an attempt to capture the different existing interpretations, Bazire & Brézillon (2005) extracted 150 definitions of context that came mainly from web sources in different domains of cognitive sciences and related disciplines. Although their study was not designed to serve as a literature review, it still demonstrates that context is an ill-defined concept for which it is difficult to identify a definition that is generalisable across all disciplines. Dimensions that represent different shapes of context are for instance: spatial and temporal context which are commonly employed in mobile recommender systems (e.g. Aliannejadi & Crestani, 2018), the task context (e.g. Li & Belkin, 2008), or the interaction context which is determined by the actions a user has performed.

In the following section, the role of context in (I)IR is introduced based on a nested model of context stratification by Ingwersen & Järvelin (2005b).

3.1 Nested Model of Context Stratification

In research on information, one can distinguish between the social context of a searcher (e.g. cultural and organisational features), which is increasingly important for research in the area of human-computer interaction, and a more algorithmic, computer science driven view on context, which refers to information objects, their contents, and interrelationships (Ingwersen, 2006).

The importance of context in (I)IR has already been recognised in previous theoretical models, such as Saracevics' stratified model of interaction (see Section 2.3), or Belkins' hypotheses on the anomalous state of knowledge (Belkin, 1980). Though, these models make the concept of context sometimes interchangeable with the concept of situation, which is one central component in the analysis of the IR process.

It is generally accepted that IR is an interactive process that occurs in multiple, overlapping contexts that inform, direct, or shape the interactions (Cool & Spink, 2002). Thus, context plays a fundamental role and has a substantial impact on nearly all components involved in the IR process.

Recognising the importance of context for IR, a nested model of context stratification is presented by Ingwersen & Järvelin (2005b) that incorporates multiple dimensions of context. A representation of the model is displayed in Figure 3.1.

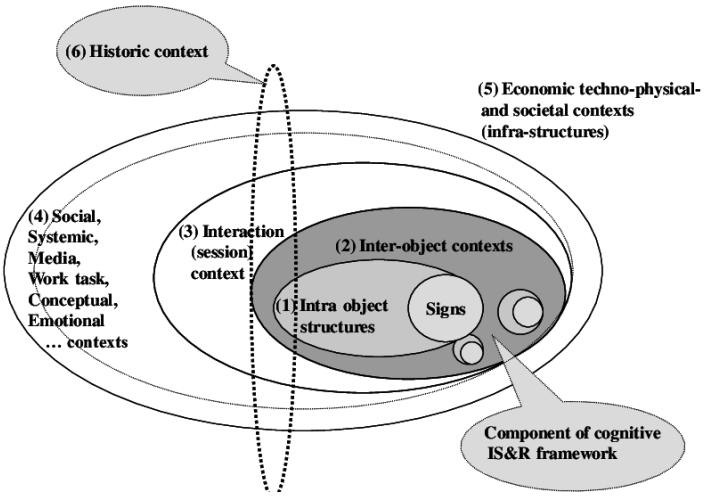


Figure 3.1: Nested model of context stratification (Ingwersen & Järvelin, 2005b).

The model of context stratification comprises the following dimensions[10]:
1) *Intra-object* refers to contextual elements that exist within range. This could be, for instance, a text surrounding an image. 2) *Inter-object* represents between-object relations such as hyperlinks or citations, which connect objects to one another. 3) *Interaction context* represents the interaction behaviour during a search session. It can form a rich network of potential information regarding preferences, style, experience, and knowledge, as well as interests. As such, it can be applied in recommender systems and personalised information systems. 4) *Social, Systemic, Work task* context incorporates, for instance, the social context (e.g. peer group), systemic (e.g. search engine), and task context (e.g. work task). 5) *Societal infrastructures* influence all actors, components and interactive sessions. 6) *Historic* context operates across each stratification. It is the history of all participating actors experiences affecting their expectations.

From this model, they derived the hypothesis (and belief) that "..by taking account of context the next generation of retrieval engines dependent on models of context can be created, designed and developed delivering performance exceeding that of out-of-context engines" (Ingwersen & Järvelin, 2005a, p. 32).

Each of the dimensions in the nested model of context stratification encourages different aspects of research. The first dimension, inter-object relations, can, for instance, be utilised to develop contextual article rankings in which the context of

[10]For a more detailed discuss see Ingwersen & Järvelin (2005b); Ingwersen (2006).

a citation, i.e. the text around a citation mark, can be employed in order to rank articles concerning a given topic (e.g. Doslu & Bingol, 2016).

The configuration of the stratified model of context is dependent on the component under consideration. The innermost-circle of the model represents the core component whose nature determines the characteristics of all remaining contexts in the model. Hence, more focused representations can be developed by putting concrete use cases or actors into the centre of attention. An example of a modified version of the model applied to a recommender scenario was developed by White et al. (2009). Given a user on a web page u_t, they employ the inter-object context (u_t), interaction context (u_t), and the historic context(u_t) to model user interest from contextual information.

Another distinction of different context levels is found in the introduction to the special issue concerning *issues of context in information retrieval* by Cool & Spink (2002). They defined context to be relevant on four levels: 1) *the information environment level* within which information behaviour takes place. For instance, organisational or work task settings. 2) *The information-seeking level*, which includes the goal(s) a person is trying to achieve. 3) *The IR interaction level* that explores the user–system interaction within search sessions. 4) *The query level* which explores IR system performance on user queries. An example of context on the query level is adding further information (the context) to determine the scope of the request in more detail.

3.2 Relevance Feedback

A valuable source for contextual information, especially regarding the interaction context is the so-called *relevance feedback*. The core principle of relevance feedback is to choose important concepts from previously retrieved documents that were identified as relevant and enhancing the importance of these concepts in a new query formulation (Salton & Buckley, 1990). The main goal of relevance feedback is to support the user during the initial phase of retrieval, in which a user often encounters difficulties in formulating appropriate queries. Instead, it is assumed that it is easier to judge particular documents as relevant and engage in iterative query refinement.

Relevance feedback can be mined from three sources: explicit, implicit, and via pseudo relevance feedback. The former two variants are discussed in the subsequent section in more detail.

3.2.1 Explicit Relevance Feedback

Explicit relevance feedback collects information about relevant content directly from the user. Given a set of retrieved documents, a user explicitly marks or selects those documents that are relevant for his or her search task. A retrieval system can then exploit the information concerning what the user is interested in and what not in the subsequent search activities.

According to Manning et al. (2008), the basic procedures are: 1) a user initialises a search on the basis of a query, 2) the system returns a set of retrieved documents, 3) the user marks relevant documents from the initial set of retrieved documents, 4) the system determines a better representation of the user's information need on the basis of the provided feedback, and 5) returns a revised and more accurate set of retrieved documents. This process can go through multiple iterations until, eventually, the information need of a user is satisfied.

Explicit relevance feedback can be very effective but has a rather obvious drawback: The effectiveness of relevance feedback depends on additional activities resulting in higher costs to the user. At the same time, the benefits of these activities are not always apparent to the user, which is why users might be reluctant to provide additional information. Thus, it can be difficult to collect the necessary data and consequently the effectiveness of explicit techniques can be limited (Kelly & Teevan, 2003).

3.2.2 Implicit Relevance Feedback

One way to overcome the downsides of collecting explicit relevance feedback is to utilise implicit relevance feedback. In contrast to explicit feedback, information about the preferences of a user are collected unobtrusively during one or multiple search session(s). The general assumption during the collection of implicit relevance feedback is that a user is implicitly providing traces of relevance information during a search session which helps a system to approximate the user's information need.

There exist various features of interaction that imply relevance and which in combination or separately act as a surrogate for explicit relevance feedback. Some categories of observable implicit relevance feedback are listed below:

- **Click-through data** cover information about the documents on which a user clicked. The main assumption is that clicked documents are more relevant than the ones that are not clicked, for example, on a result page. A frequent interpretation of click-through data is that each click represents an endorsement of that page (Joachims et al., 2017).

- **Dwell time** indicates the time that a user spent reading a document. The main assumption here is that a user would engage more intensively with a relevant document and therefore, spend a longer time span with a relevant document than with a non-relevant document.
- **Mouse movements / Text selections:** The movement of the cursor and the selection of text snippets on a document potentially indicate user interest with the particular item.
- **Save:** This category acts as an umbrella concept for interactions that involve further activities with a document such as saving, forwarding, bookmarking, printing, citing, or exporting.

The main advantage of implicit relevance feedback compared to explicit feedback is that the cost of providing additional information is avoided. Furthermore, implicit relevance feedback allows a combination of signals, e.g. click-through data and dwell time. On the other hand, implicit relevance feedback might suffer from noisy data as the relevance of a document is just an approximation that might not be accurate enough and thus, have a negative impact on the effectiveness of the retrieval.

3.2.3 Query Expansion

In order to operationalise relevance feedback, whether implicit or explicit, it is a common practice to expand the user query by adding useful terms, to better reflect the information need (Buckley et al., 1995). This process is called *query expansion*.

Query expansion (QE) has been well known for decades, and the advantages of this approach have been discussed in detail throughout the literature. Amongst others, QE has the potential to clearly specify or narrow the query when ambiguous terms are involved (e.g. python the animal vs python the programming language), enrich an imperfect query with additional concepts potentially leading to better results and thus, potentially reducing uncertainty at early stages of the search process (see Section 2.2.2).

Assume a user has submitted a query Q which consists of a set of n terms $Q = \{t1, t2, \cdots tn\}$. A generalised query expansion could now be formalised as depicted in Equation 3.1:

$$Q^* = \{\texttt{expand}(t_i) | \forall t \in Q\} \tag{3.1}$$

The set of expanded query terms Q^* is provided by an expansion function *expand*, which represents a general expansion function that determines associated terms to a given term t.

One of the most prominent methods for query expansion is known as the *Rocchio Algorithm* (Rocchio, 1971), which operates within the *Vector Space Model*.[11] Given an initial query Q, the Rocchio Algorithm is mathematically defined in Equation 3.2:

$$\vec{q_m} = \alpha\vec{q} + \frac{\beta}{|D_r|}\sum_{\forall\vec{d_j}\in D_r}\vec{d_j} - \frac{\gamma}{|D_n|}\sum_{\forall\vec{d_j}\in D_n}\vec{d_j} \tag{3.2}$$

In Equation 3.2, D_r denotes the set of relevant documents, determined either via explicit or implicit feedback by a user. Accordingly, D_n represents the set of non-relevant documents. The parameters α, β, γ represent the associated weights of the initial query (α), the relevant documents (β), and the non-relevant documents (γ). The goal of the Rocchio Algorithm is to expand the original query q to a new query q_m, which moves closer towards the centroid of the relevant documents (D_r) and away from the set of non-relevant documents (D_n).
An optimal modification of the user's query is, in reality, rather unlikely, as relevance feedback usually only provides an incomprehensive view of what is considered as relevant by the user.

Although the present section focused strongly on relevance feedback, the methods to address query expansion are not limited to those. Other approaches utilise, for instance, a thesaurus to find appropriate expansion terms e.g. (e.g. Aronson & Rindflesch, 1997).
More recent methods employ word embedding techniques that represent terms as a low-dimensional vector in a vector space. Terms that are semantically or syntactically related tend to be close in the semantic space (Kenter & de Rijke, 2015). Word embeddings, which have been effectively applied to NLP tasks, are recently used for many IR related tasks, such as document ranking and query expansion (e.g. Ganguly et al., 2015; Zamani & Croft, 2016; Roy et al., 2018).
For a comprehensive overview of query expansion methods in IR one can refer to Azad & Deepak (2019).

3.2.4 Applicability of Implicit Relevance Feedback

The following section reviews related work with respect to the effectiveness and applicability of implicit relevance feedback.

A comparison of implicit and explicit feedback is presented by White et al. (2001). They examined to what extent implicit feedback can act as a substitute for explicit feedback. Their experiment comprised 16 participants who were asked to solve

[11] The vector space is a classic model in IR. In this model, documents and queries are represented as vectors in a t-dimensional space (Baeza-Yates et al., 1999). As IR models, in general, are not in the original scope of this thesis, they are not further introduced. However, for a thorough overview of the different models used in IR, one can refer to Croft et al. (2009).

predefined tasks, that were taken from the TREC-10 interactive track. Two interfaces were developed that connect to the Google search engine: one that uses implicit, and one that uses explicit relevance feedback. Both variants were compared with regards to viewed result pages, task completion, and task time. The results showed no significant difference in terms of search effectiveness on any of the investigated dimensions.

Joachims et al. (2017) presented an examination of the reliability of click-through data as implicit relevance feedback. The focus of their study was two-fold: a) to investigate the users' decision process when interacting with result lists by using an eye-tracking device, and b) to evaluate the degree with which implicit signals indicate relevance. The latter was done by comparing the implicit feedback signals, such as click-through data with explicit feedback, that was collected manually. Their result showed that the users' click decisions are influenced by the relevance of the results. At the same time, however, they noticed that the interpretation of click decisions was problematic because of two reasons. First, they observed a trust bias by comparing clicks at the two top-ranked documents. The document on position one was more frequently clicked despite that an analysis of gaze data showed that both documents are viewed much more equally. Second, they observed a quality bias which was exposed by: a) swapping documents at rank one and rank two in the result list and b) by reversing the order of the documents in the result lists. Their experiment showed that the quality of the ranking influences the users' clicking behaviour. With decreasing quality of the result sets (e.g. reversed order), users click on results that are on average less relevant. From these observations, they argue that the interpretation of clicks as relevance feedback should not be on an absolute scale but rather take into account the trust bias and the quality of the ranking.

Another study comparing individual and combinations of implicit relevance feedback to explicit ratings is presented by Claypool et al. (2001). In their study, they examined individual implicit relevance feedback which involved the time a user spends on a page, the time spent moving the mouse, the amount of scrolling, and the number of mouse clicks on a page. The authors concluded that: a) the time spent on a page, b) the so-called "amount of scrolling" which is a combination of different scrolling measures, and c) the combination of time and scrolling were good indicators of interest. The number of clicks on a page, however, turned out to not represent an indicator for interest.

Morita & Shinoda (1994) investigated to what extent the preference of a user is reflected in the time spent reading a news article. Their experiment involved eight participants reading a total of 8,000 news articles. They concluded that the preference of a user was the main factor that affects the time spent reading an article and thus, that it is possible, at least to a certain extent, to derive relevance feedback from the time reading an article. Kelly & Belkin (2001) conducted a replication study to reproduce the results of this experiment in another IR context. The results of this replication study, however, showed that the time viewing a document was not significantly related to the user's subsequent relevance judgement. Kelly &

Belkin (2001) concluded that aspects such as the tasks, the underlying document collection, and search environment might affect the generalisability of the original study.
The applicability of text selection events on search result pages as implicit relevance feedback was investigated by White & Buscher (2012). Their central hypothesis was that text selections provide implicit information about a user's search interest which could be employed on future queries for that particular user to increase search performance. The study involved volunteers from the Microsoft Corporation whose text selection events were recorded when using Microsoft Bing.[12] On the basis of text selections, they developed an interest model, which contained non-stopword terms appearing during a text selection event. The top-10 results of a near-future query B were then re-ranked according to the cosine similarity between the interest model and each result snippet. Their experimental results showed significant improvements over two baselines: the search engines' original result ranking and a re-ranking by query similarity. Hence, their findings suggest that text selections could be used effectively as implicit relevance feedback.

The contributions discussed reflect only a few, highly relevant examples from the literature available on implicit relevance feedback.[13] Nevertheless, one can observe that the features applicable to determine implicit relevance feedback are manifold. The literature discussed suggests that implicit relevance feedback can be employed as an approximation for relevance. However, it has to be stressed that studies on implicit relevance feedback are often difficult to generalise, due to factors like the given task, underlying document collection etc., which have an impact on the results (e.g. Kelly & Belkin, 2001). Furthermore, it is important to emphasise that interactions, such as the click-through rate, are not interpreted on an absolute scale (Joachims et al., 2017).

3.2.5 Conclusion

In the present section, the notion of context was introduced. First, the nested model of context stratification was discussed, which highlights the importance and the different manifestations of context during information seeking. Of particular relevance is the interaction context which provides the theoretical basis for the development of contextualised stratagem browsing. Subsequently, relevance feedback and query expansion were introduced. They provide the algorithmic methods to implement contextualised stratagem browsing (see Section 7). Finally, the applicability of different sources of implicit relevance feedback was discussed.

[12] `https://www.bing.com/`, last accessed April 15th, 2020.

[13] For a more comprehensive view on implicit relevance feedback one can refer to Kelly & Teevan (2003).

4

Digital Libraries

The previous two sections introduced and discussed models of information seeking and contextualisation in conjunction with their applicability, primarily under the consideration of an abstract entity denoted as an information system. Concrete instances of such an information system are Digital Libraries (DLs). Consequently, this section is dedicated to the introduction of DLs, in order to illustrate how the different concepts that contribute to DLs interplay.

DLs have been subject to research for more than two decades,[14] with contributions from international conferences (e.g. JCDL[15], TPDL[16]), as well as journals (e.g. IJDL[17]) dedicated specifically to the domain of DLs. One essential and often mentioned aspect of DLs is that they represent an interdisciplinary research domain in which several disciplines (e.g. computer science, library science) and groups of practitioners (e.g. publishers and librarians) are involved (Fuhr et al., 2001).
Providing an adequate and generally accepted definition of DLs, however, is a non-trivial and challenging task and, as of today, there exist numerous definitions of DLs each emphasising different aspects (e.g. Borgman, 1999), (Lesk, 1997), (Arms, 2001). One reason for this is the aforementioned interdisciplinarity of the field. A computer scientist, for instance, might highlight the task of indexing and retrieving information, which consequently associates to the traditional competences of information retrieval. A librarian, on the other hand, might focus on the representation, curation, and structuring of the available information in a DL.

One prominent and often referenced definition of DLs is provided by Arms (2001), who states that

> ".. 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, 2001, p. 2).

[14] Although other terms have been used previously, such as electronic or virtual library, the term Digital Library got widely accepted around 1991.

[15] Joint Conference on Digital Libraries (`https://www.jcdl.org`, last accessed April 15th, 2020).

[16] Theory and Practice of Digital Libraries (`http://www.tpdl.eu`, last accessed April 15th, 2020).

[17] International Journal on Digital Libraries (`https://www.springer.com/journal/799`, last accessed April 15th, 2020).

Although Arms understood this as an informal definition, he emphasised one key aspect of DLs which determines a necessary scope and serves as a distinction from, for instance, web search which is: that DLs are "managed collections of information". A DL is thus, not only understood as a set of indexed information, but rather as a controlled environment in which information is collected and organised.

A similar distinction between DLs and the WWW was expressed by Fox et al.: "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)." (Fox et al., 2012, p. 6).

Despite several other existing definitions, the ones proposed by Arms and Fox et al. are well suited in the context of this thesis as they offer a certain degree of generality, while at the same time a differentiation between DLs and web search engines is drawn.

4.1 Components of Digital Libraries

In order to develop and, equally important, to evaluate DLs, it is necessary to gain a better understanding of the theoretical foundations and the different stakeholders involved in the DL field. Formal models and theories can help to provide such an understanding as they are crucial to specify and understand the characteristics, structure, and behaviour of complex information systems (Goncalves, 2004).

Two prominent examples that aim to lay these foundations are the *DELOS Reference Model* (Candela et al., 2007) and the *5S framework* (Streams, Structures, Spaces, Scenarios, Societies) (Gonçalves et al., 2004), which are both discussed in the following.

4.1.1 DELOS Reference Model

In Figure 4.1, the DELOS Reference Model is displayed. In the centre of this model resides the "Digital Library Universe", which is an aggregation for a three-tier framework. The components of this framework are briefly explained below following Candela et al. (2007):

- **Digital Library** represents an organisation that comprehensively collects, manages, and preserves digital content that is offered to its users in conjunction with specialised functionality on that content.
- **Digital Library System** represents the deployed and running software that implements the DL facilities and provides all functionalities required.

- **Digital Library Management System** represents a generic software component that provides the appropriate software infrastructure to produce and administer a Digital Library System.

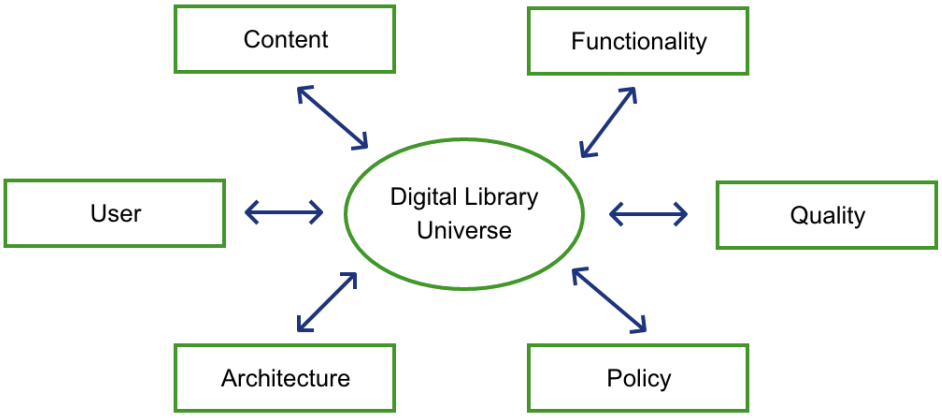


Figure 4.1: The main concepts of the Digital Library Universe according to the DELOS Reference Model. Figure adapted from Candela et al. (2007).

Despite the wide range of existing DL applications, one can identify a common set of concepts that are available in almost any of the systems. These are represented by the six core concepts, which surround the Digital Library Universe displayed in Figure 4.1. Following Candela et al. (2007), the involved components are:

- **Content**: This concept covers all data and information that is managed by the DL. It covers all forms of information objects that a DL collects, manages, and delivers (e.g. metadata, annotations).
- **User:** The concept "user" encompasses all the different actors that are entitled to interact with a DL. Amongst others, the concept user represents authors, publisher, librarians. As an umbrella concept, it covers any other aspect that is related to a user, e.g. the storage of interaction behaviour profiles which are often used to contextualise a retrieval session.
- **Functionality**: This concept describes all services that a DL offers to its users. The concept "functionality" reflects the particular needs of the digital library's community.
- **Quality**: This concept represents the parameters that can be used to evaluate the quality of a DL. It is a strongly inter-related concept that applies to multiple components such as the content or the functionality of a DL.
- **Policy**: This concept covers the conditions, rules, and regulations governing the interactions between the user and a DL. Policies deal, for instance, with digital rights management, pricing, privacy.

- **Architecture**: This concept represents the mapping between the functionality and content offered by a DL onto the hardware and software components.

Although the six core concepts of the DELOS Reference Model are represented separately, they are by no means to be understood independently but much rather as strongly inter-related: one entity within a concept is often related to or influenced by other entities from different core concepts (Agosti et al., 2007). This is particularly visible for the concept quality, which is strongly inter-related to the concepts content and functionality.

4.1.2 5S Framework

Another model that aims to provide the theoretical foundations for DLs is the *5S framework* whose name is a result of the five components: Streams, Structures, Spaces, Scenarios, and Societies which are introduced below following Gonçalves et al. (2004).

- **Streams** represent the elements of an arbitrary type which is either of static (e.g. textual material) or of dynamic (e.g. a digital video) nature.
- **Structures** represent the way in which parts of a whole are organised and arranged. A typical example of a structured organisation is a book that can be separated into chapters, sections, or subsections.
- **Spaces** refer to a set of objects, together with operations on those objects, that obey certain constraints. Spaces account, for instance, for presentation (e.g. virtual reality spaces) or indexing (e.g. estimating the similarity of two concepts by measuring their distance in a vector space). Spaces are a highly generic component that can be applicable when a concept can not be matched to any of the other Ss.
- **Scenarios** cover possible ways to use a system and are suitable to describe external system behaviour from the users' point of view.
- **Societies** cover a set of entities and the relationships among them. Society in the context of DLs includes, for instance: authors, publishers, editors, or developers.

The components of the 5S framework are split into two parts: content-related, and people and service. Streams, Structures and Spaces are closely connected components that relate to all matters with respect to the content of a DL. Scenarios and Societies, which are also strongly connected, relate to people and services of a DL.

Both, the DELOS Reference Model and the 5S framework, aim to provide a theoretical framework for DLs. The 5S framework applies a rigorous definition of various DL concepts, while the DELOS reference model is more concerned with

identifying the main concepts and relationships encompassing the entire DL (Isah et al., 2013). The DELOS reference model was developed mostly top-down, in order to cover up for existing and planned DL. The 5S framework, on the other hand, was launched earlier and was developed largely bottom-up, starting with key definitions and with an elucidation of the DL concept from a minimalist and formal approach (Agosti et al., 2007).
Both models represent a comprehensive conceptualisation of DLs, but the pragmatic nature of the DELOS reference model is more suitable for this thesis in comparison to the rather abstract formalised 5S model.

4.1.3 The Concept of Quality in Digital Libraries

Even though only briefly introduced, the two models show that DLs are undoubtedly complex systems, which amongst others, involve different actors, concepts, and organisational units. In order to build an effective and usable DL, one has to consider each of the different stakeholders by taking into account the various needs and goals. Consequently, the assessment of a DLs quality depends on the particular scope and the underlying need; or as Fuhr et al. recognised:

> "What is a digital library?" The answer depends upon whom you are asking. This is even truer if you ask, "What is a *good* digital library?" (Fuhr et al., 2001, p. 187).

The evaluation of DLs is a non-trivial task and various researchers have proposed suitable evaluation frameworks. Gonçalves et al. (2007) proposed a quality model for DLs which is build upon the 5S framework aiming to define critical dimensions and indicators of quality in a DL. A comparison of the quality concept in the DELOS reference model and the quality dimension in the 5S quality model can be found in Agosti et al. (2007).

For a more comprehensive discussion of the complexity of evaluating DLs one can refer to Saracevic (2000) who highlights the conceptual and pragmatic challenges that are encountered when evaluating DLs and a more recent publication by Fuhr et al. (2007).

4.2 Metadata in Digital Libraries

An essential component of DLs is *metadata*, which are often loosely described as "data about data". In the context of DLs, metadata are considered structured information that ".. describes the content and attributes of any particular item in a digital library." (Cleveland, 1998, p. 5). Moreover, metadata play an essential role by enabling users to search or browse for resources of interest (Riley, 2017; Cleveland, 1998). Metadata in the DELOS Reference Model are incorporated in

Table 4.1: Six types of metadata according to the National Information Standards Organization (NISO) with shortened examples taken from Riley (2017).

Type	Description
Descriptive metadata	Finding or understanding a resource
Structural metadata	Relationships of parts of resources to one another
Administrative: • Technical metadata • Preservation metadata • Rights metadata	 Decoding and rendering files Long-term management of files Property rights attached to content
Markup languages	Integrates metadata and flags for other structural or semantic features within content

the concept "content" and take up ".. a central role in the handling and use of information objects, as they provide information critical to its syntactical, semantic, and contextual interpretation" (Candela et al., 2007, p. 19).

Today, metadata are pervasive in everyday online life and used in social media, webshops, or even in the infamous exposure of metadata collection by the National Security Agency (Greenwald, 2013).

According to Riley (2017), one can distinguish six types of metadata which are displayed in Table 4.1. Each of the different types of metadata is of relevance for the design of DLs. In the context of this thesis however, it is sufficient to concentrate on two of the given types: descriptive and structural metadata.

Descriptive metadata provide fundamental information about a record in DLs and are arguably the most important metadata type for users of a DL. Descriptive metadata in DLs are analogous to a traditional catalogue record and describe the intellectual content of a record, enabling its retrieval and assessment. Examples of descriptive metadata are the title or the author of a record.
Structural metadata, on the other hand, describe the relationships of parts of resources to one another (Riley, 2017). The primary use of structural metadata is to enable the navigation between related resources and as such support the process of browsing.

The opportunities of browsing on the basis of metadata have also been discussed by Witten et al. (2009), who understands the structure that is implicit in metadata as the key to providing browsing facilities. Some of today's browsing capabilities in DLs are presented in the following section.

4.3 Browsing in Digital Libraries

Many modern DLs today offer some variation of browsing capabilities. What can be explored, however, is dependent on the richness of the underlying metadata, and of course, on the providers' design choices. Figure 4.2 shows two adjusted[18] example records that were taken from the *Web of Science*[19] and *SSOAR (Social Science Open Access Repository)*[20].

Both examples display descriptive metadata which provide the users of a DL with information necessary for the assessment of the record. The descriptive metadata cover information about the records title, authors, abstract, keywords, classifications, and publication information such as the publishing journal, volume, published date, digital object identifier (DOI). Additionally, both examples facilitate browsing by implementing keywords, classifications (SSOAR only), and author information as hyperlinks. This represents today's implementations of Bates' stratagems as introduced in Section 2.4.

The implementation of browsing functionalities and its subsequent document ranking is, of course, DL dependent. Many DLs, however, facilitate browsing based on a filtered query retrieving those documents that share the particular filter criterion. This will be explained in more detail in Section 4.4.3.
Figure 4.2 shows that the two displayed DL records share certain commonalities, considering the representation of descriptive metadata. The access to the content, however, is very different. SSOAR is an open access repository for the social sciences that provides full-text access to the users. SSOAR covers 56,000 open access articles mainly from the social sciences. Full texts in SSOAR are published either by authors themselves (self-archiving) or by research institutes, universities, and discipline-specific infrastructure providers (SSOAR, 2020). The Web of Science, on the other hand, is a commercial citation index that covers more than 72 million bibliographic records together with references for many of the records (Web of Science Core Collection) (Mangan, 2018).
While these are only two examples of DLs, they represent a very common implementation design that can be found in many other systems such as PubMed[21], ACM Digital Library[22], or Springer Link[23]. In the following section *Sowiport*, a DL for the social sciences, will be introduced.

[18] The screenshots were adjusted to demonstrate the use case. Unnecessary elements such as translated abstracts were removed to fit the page.

[19] `http://apps.webofknowledge.com/`, last accessed April 15th, 2020.

[20] `https://www.ssoar.info`, last accessed April 15th, 2020.

[21] `https://www.ncbi.nlm.nih.gov/pubmed/`, last accessed April 15th, 2020.

[22] `https://dl.acm.org/`, last accessed April 15th, 2020.

[23] `https://link.springer.com`, last accessed April 15th, 2020.

(a) Web of Science

(b) SSOAR

Figure 4.2: Two modified example documents taken from *Web of Science* and *SSOAR* . Both records were retrieved by submitting the query "violence and sports". Screenshots taken April 4th, 2020

4.4 The Digital Library Sowiport

To implement and evaluate the concepts of contextualised stratagem browsing *Sowiport*, a Digital Library for the Social Sciences is employed.[24] To allow for a better understanding of the upcoming experiments presented in this thesis, Sowiport, its constituting document corpus, and some of the most relevant features are now described briefly.

By the time the experiments presented in this thesis were conducted, Sowiport comprised about 9.7 million literature references covering topics from the social sciences. On a weekly basis, Sowiport reached around 20,000 unique users. The records indexed in Sowiport were integrated from 23 different databases, that came primarily from collaborating institutes and libraries.

An example document indexed in Sowiport is displayed in Figure 4.3. One can see a composition very similar to the DLs Web of Science and SSOAR (see Figure 4.2). Figure 4.3 is organised into three sections. In Section A, one can see that the example document comprises various descriptive metadata with essential information about that particular document, such as the title, author information, classifications, and topics. Furthermore, one can see Sowiports implementation of stratagem browsing. In the depicted example, specific fields of descriptive metadata are represented as hyperlinks which allow further exploration of documents indexed in Sowiport. Metadata implemented as hyperlinks are available for the following fields: keywords, classifications, the journal, and author information. A detailed example of this implementation follows in Section 4.4.3. A list of descriptive metadata, along with corresponding browsing facilities, can be found in Appendix A.8.

The documents' abstract, citation, and reference information are located in Section B of Figure 4.3, which is organised in a tab view. References and citations account for a special case: first of all, not all documents contain information about citations and references. In total, only 5.2% of all documents contain citation information, and 2.3% of the documents contain information about referenced articles. A further restriction is that browsing is only supported if the reference or citation resolves to a document that is also indexed in Sowiport. As illustrated in Figure 4.3, only the references five, six, and seven are linked to content in Sowiport.
In addition, Sowiport offers a recommendation service which displays documents related to the currently examined one. These recommendations are shown in Section C of Figure 4.3.

[24] In 2018, Sowiport was replaced by an integrated search system `http://search.gesis.org`, that covers several object types such as bibliographic records, datasets, and information about researchers. Sowiport was last accessed in December 2017.

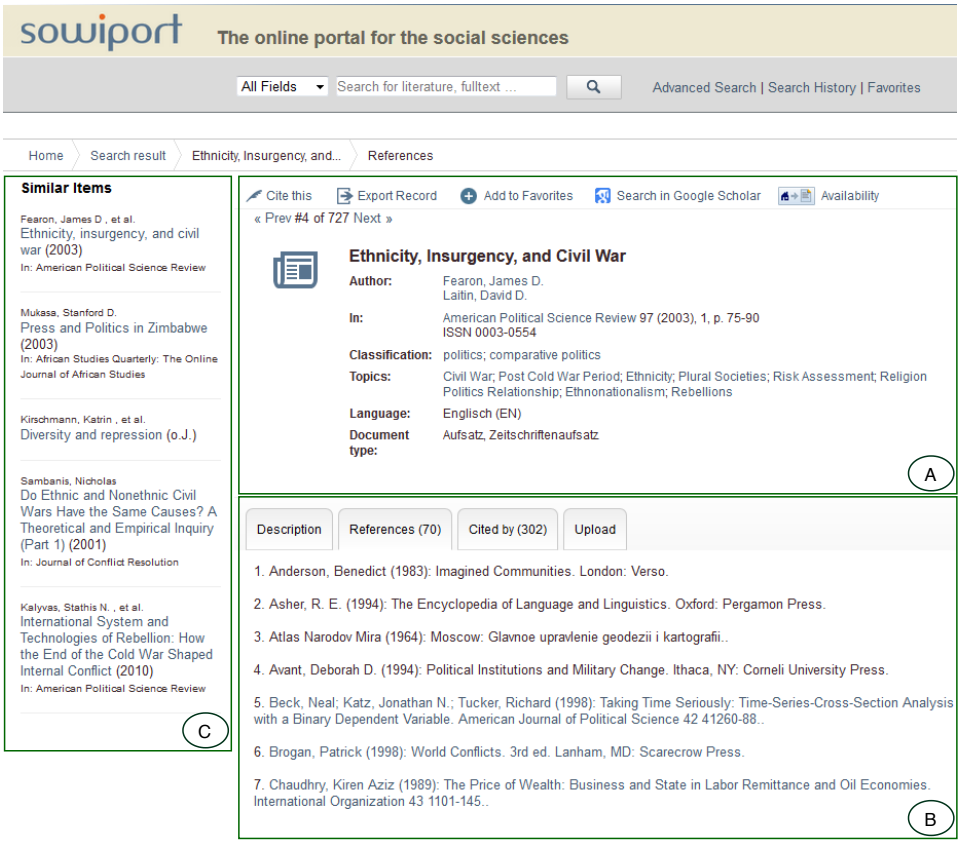


Figure 4.3: Example document from Sowiport that shows descriptive metadata (A), a tab organised view (B) containing the abstract, references and citations, and (C) recommendations related to that particular document.

4.4.1 Architecture

The architecture of Sowiport is based on a *VuFind 2* framework.[25] The VuFind discovery software is a customisable open-source library portal that provides many of the basic functionalities available in Sowiport, such as browsing, simple, and extended searching, and faceted searching. Sowiport records are indexed in an Apache Solr 5.3 index.[26]

An overview of the Sowiport architecture incorporated in the conceptualisations of the DELOS reference model is displayed in Figure 4.4. Applying the conceptualisation from the DELOS reference model (see Section 4.1.1), one can identify a four-level architecture that involves the concepts: content, functionality, user (interaction log), and the digital library system.

[25] `https://vufind.org`, last accessed April 15th, 2020
[26] `https://lucene.apache.org/solr/`, last accessed April 15th, 2020

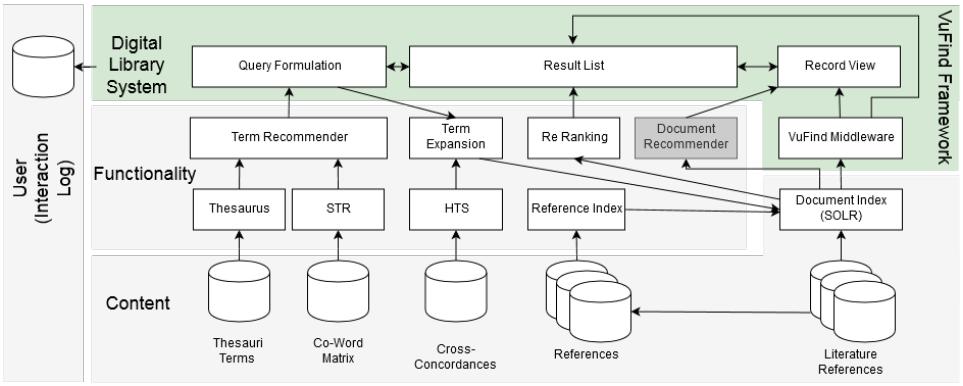


Figure 4.4: Overview of the Sowiport architecture, adapted towards the DELOS reference model on the basis of Hienert et al. (2015a).

The **content** level encompasses all data that are made available to the users. This covers, amongst others, the documents (depicted as literature references) that are indexed and retrieved, a thesaurus and a reference index.
The **user** level involves a logging framework in which each interaction of a user is stored in a MySQL database. The logging framework utilises a javascript based logging mechanism that stores information about the users' actions. These include, for example, a session identifier, a timestamp, the identifier of each viewed record, and each submitted query.
The **Digital Library System** is an instantiation of the *Digital Library Management System* VuFind which involves the VuFind middleware and the user interface. The latter includes essential components such as templates for result lists and record views.
The **functionality** level incorporates tools developed to support the users of Sowiport during their information-seeking task. These involve a term recommender, a heterogeneity service, and a recommender system. These three features are now described in more detail.

- **Term Recommender:** In order to support the user during the query formulation process, a *term recommender* is implemented, which supports the user in finding alternative concepts to a given query. The term recommender utilises two services: the *Thesaurus for the Social Science* (TheSoz) and the *Search Term Recommender* (STR). The TheSoz covers topics from the Social Sciences with basic semantic relations like broader, narrower, and related (Zapilko et al., 2013). The STR maps query terms to indexing terms at search time in order to recommend more appropriate query terms to the user (Hienert et al., 2011).

- **HTS:** Due to the heterogeneity that results from integrating content from 23 different databases, a web service called *heterogeneity service (HTS)* was de-

veloped to support cross-concordance searches. According to Mayr & Petras (2008), cross-concordances are intellectually created crosswalks that determine equivalence, hierarchy, and association relations between terms from two controlled vocabularies. The underlying cross-concordance database contains controlled terms from 25 different thesauri with about 513,000 entries. This web service plays a vital role during the ranking of documents as it is employed during query expansion. A query that was formulated by a user is transmitted to the HTS. The query is then expanded on the basis of equivalence relations from all cross-concordances where the TheSoz is the source or target database (Hienert et al., 2015a). This query expansion serves as a baseline in the evaluation of the contextualised ranking features which will be introduced in Section 7.2.

- **Recommender Systems:** Recommendations in Sowiport are provided to the user as part of the detailed view of a document (see Figure 4.3 C). The way these are implemented, however, differs from other services like the HTS and STR in a way that these are provided by a cooperation partner in the form of a Web Service (Beel et al., 2017a). The implementation of these *recommendations as a service* enables other researchers to develop own recommendation algorithms that can be evaluated in experiments using real-life users and data and is thus serving as a living lab. The experiments conducted include, for instance, a comparison of stereotype and most-popular recommendations on the basis of click-through data (Beel et al., 2017c). They have been made publicly available for other researchers for reproduction and improvement (Beel et al., 2017b).

4.4.2 Interaction Pattern

Although Sowiport comprises a large number of documents and attracts a high number of online users, one can observe a very common interaction pattern visible in a large proportion of sessions. A majority of the users access Sowiport via search engines that crawl and index the documents that are covered by Sowiport. Another large proportion of Sowiport users come from university libraries that link to Sowiport content. Both groups usually visit Sowiport by immediately accessing the detailed view of a document, called *seed document*, as these are usually indexed by search engines and university libraries. A visualisation of interaction sequences that show this access pattern is depicted in Figure 4.5. One can see a Sankey diagram that has been created using *WHOSE* - a whole-session analysis tool in IIR (Hienert et al., 2015b). The diagram shows the distribution of interactions per session step. Each action of a user (e.g. view record, query, browse) is depicted as "Step" in a chronologically ordered sequence of interactions. Of particular interest is the distribution along Step 1. One can see that the majority of initial interactions in Sowiport are on the action level *view record*, which classifies this interaction as a user looking at the detailed view of a record (see Figure 4.3). These sessions

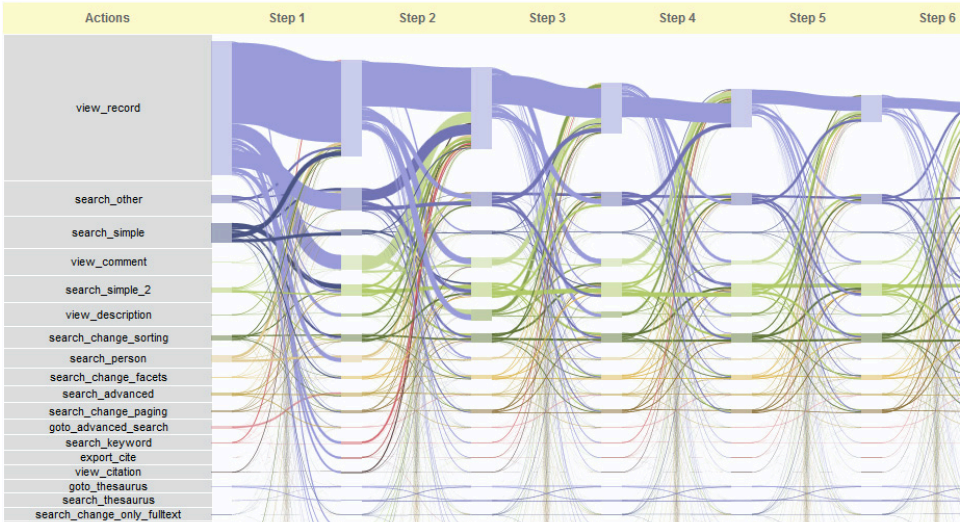
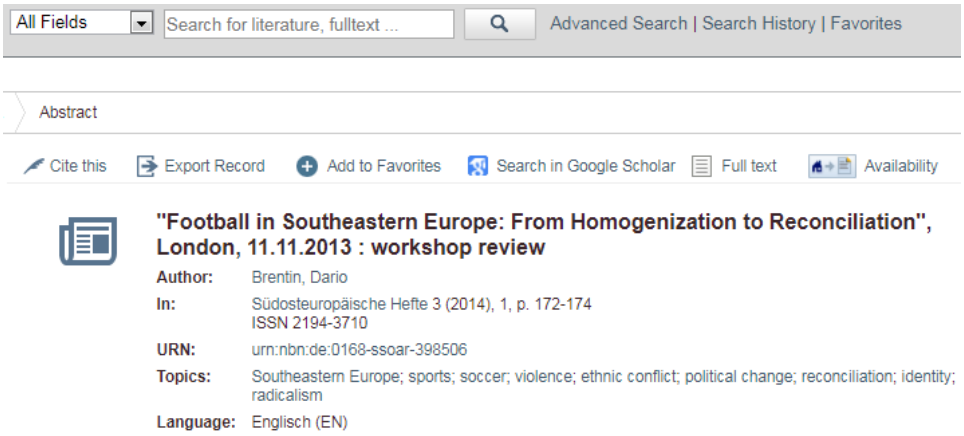


Figure 4.5: Sankey diagram of Sowiport interaction sequences. The diagram was provided by the WHOSE tool developed by Hienert et al. (2015b).

presumably came from a web search engine given that no initial search within Sowiport has been performed. On the opposite, only a comparably low amount of sessions start on the action level *search_simple*, which is the leading search form of Sowiport.
The interaction patterns depicted in Figure 4.5 have an essential influence on the evaluation of contextualised stratagem browsing and will be discussed in detail in Sections 8 and 9.3.

4.4.3 Stratagem Browsing in Sowiport

A practical example of stratagem browsing in Sowiport is displayed in Figure 4.6. In this example, a seed document with the title "*Football in Southeastern Europe:* ..." has been retrieved via the query for "violence and sports". Stratagem browsing can be employed by selecting a keyword, the name of the author, or the journal (*Südosteuropäische Hefte – Southeastern Europe Magazine*). Each of these interactions would lead to a result list of documents containing that particular filter criterion. Such a result list is displayed in Figure 4.6 (b), which was generated by selecting the keyword *sports* from the seed document. In this example, the ranking has not been contextualised and thus, represents the default behaviour of Sowiport and many other modern DLs.
The result list displayed in Figure 4.6 (b) contains various characteristics worth noticing. First of all, one can see that the top-ranked documents are not related to the seed documents topic about football in the former Republic of Yugoslavia. Instead, the top-ranked documents cover various topics related to the field "sports". Furthermore, one can see that the top-ranked documents in the result list are pub-



(a) Seed document

(b) Result list

Figure 4.6: Seed document retrieved via a query search for "violence and sports" (a) and the corresponding result list after stratagem browsing via the keyword "sports" (b).

lished in German, which again is in contrast to the seed document, which was published in English.
In total, more than 250,000 records were retrieved, which makes an effective assessment of the result list nearly impossible.

In many modern DLs, the most common feature to narrow down search results based on certain document features are facets, which are widely implemented nowadays. Facets are described as a "set of meaningful labels organized in such a way as to reflect the concepts relevant to a domain" (Hearst, 2006, p. 60). Although empirical studies identified various beneficial aspects of faceted browsing (e.g. Fagan, 2010), they usually operate on the level of simple Boolean retrieval. Furthermore, facets require the user to interact with the result lists and to select each filter criterion one after another. In contrast to facets, the ideas of contextualised stratagem browsing, presented in this thesis, tailor search results on the level of stratagems based on the users' interaction context without additional effort of the user. In other words: contextualised browsing aims to re-rank documents that result from stratagem browsing with respect to an approximation of users' search interests, while the advantages of faceted browsing remain unaffected and can still be utilised by the user.

4.5 Conclusion

Digital Libraries are the primary use case for contextualised stratagem browsing. The present section aimed at providing an understanding of DLs, which is a non-trivial task. Furthermore, the most important components and characteristics of a DL were described and interrelated. Descriptive metadata were identified as a core component that represent both, an important source that enables users of a DL to assess the relevance of a document, and at the same time represents the operationalisation of stratagem browsing on the user interface.

Sowiport, a specific instance of a DL, was described in detail. Although there exist various other aspects of Sowiport that could be of interest, the present section has focused only on the most crucial aspects that are necessary for a better understanding of the upcoming experiments. These include the user interface, the technology stack, and the implementation of stratagem browsing in Sowiport.

Part II

Empirical Studies on Browsing in Digital Libraries

Introduction to Empirical Studies on Stratagem Browsing

Although modern DLs widely facilitate browsing, especially on the level of stratagems, the attention by the scientific community on their empirical evaluation is comparatively small. Thus, it can be challenging to assess the usefulness of these types of search activities without additional research. While today, descriptive studies exist that measure the quantitative usage of certain stratagems (e.g. Kacem & Mayr, 2018; Hemminger et al., 2007), these studies do not reflect potential benefits or flaws of stratagem browsing.
The present thesis is devoted to investigating and implementing contextual stratagem browsing by following the principles of Interactive Information Retrieval (IIR), as described in Section 2. Therefore, in line with the user-centred paradigm of IIR, it is crucial to gather a deeper understanding of stratagem browsing and to address the lack of research in this area. In order to achieve this, two empirical studies were conducted which are presented in the following two sections.

- The first study was conducted in the form of an online survey. Overall, 204 respondents were recruited to answer questions regarding the usefulness of different stratagem browsing variants. The results of this survey are presented in the following section.

- The second empirical investigation was conducted in the form of a multimethod study on stratagem browsing behaviour in which 32 participants took part. The participants were asked to solve a task on searching related documents to a given topic. The task was embedded in a scenario in which the participants had already found a relevant document and wanted to find other documents which are related to the one provided. The main objective of this observational study was to determine the actual use of stratagem browsing during a search task. Besides observing the participant's information-seeking behaviour during the experiment, an eye-tracking device was used to gain a better understanding of the perceived usefulness of certain stratagems. The results of this study are presented in Section 6.

The two studies presented in this chapter are motivated by the belief that the design of effective IR systems is only possible with a better understanding of how users interact with them (Robins, 2000).

5

Survey on Search Activities at the Stratagem Level

The first step towards a better understanding of stratagem browsing was done by conducting an online survey. The goal of this survey was two-fold: 1) to gain knowledge about the use of stratagem variants during information seeking and 2) to preliminary assess the usefulness of contextualised stratagem browsing. The stratagem variants in the survey were derived and adapted from the examples presented by Bates (1990).

The survey addressed the following research question:

> **RQ:** What kind of stratagem browsing variants do users employ during information seeking, and how do the respondents assess their usefulness?

The usefulness of stratagem variants was studied with respect to two scenarios: a) the user has found a relevant document and wants to find similar documents and b) the user employs a stratagem variant without a preceding document. The motivation for using different scenarios was driven by the assumption that the usefulness of a certain stratagem depends on the users' state of search. If, for instance, a user has found a relevant document he/she may be eager to find other relevant documents and therefore, be more encouraged to employ a certain stratagem that might lead to similar documents. On the other hand, a state of search in which a present document is not considered relevant might lead to a different assessment of a stratagems' usefulness.

The present section is an extended version of Carevic & Mayr (2016) that has been published and presented at the *International Conference on Theory and Practice of Digital Libraries (TPDL)*.

5.1 Survey Design

The survey was available for two weeks during August 2015 and was primarily designed for researchers and postgraduate students but not limited to a particular field of study. The respondents were recruited via collaborating universities and institutes, mailing lists, and social media (Twitter, Facebook). To keep the survey

maintainable for the respondents, it emphasised on only two of the six stratagems proposed by Bates: a) journal and conference run and b) citation and reference search. In total, the survey consisted of 28 questions, that were divided into four parts[27].

The first two parts were primarily concerned with the general usage of journal and conference runs as well as citations and references. The third part of the survey concerned the general usage of stratagems. To this end, the respondents were presented with six stratagems, which were all derived from Bates (1990) and slightly modified (see Table 5.3). The respondents were asked to rank these activities by their usage frequency when searching for relevant documents.

The fourth part of the survey was designed to gain information about the potential benefits of contextualised stratagem browsing. To this end, the respondents were presented with a use case in which they were asked to organise the content of a journal based on six ranking options: four generic options such as date or in alphabetic order and two novel options that were based on a contextualised ranking. The survey concluded with nine socio-demographic questions and two optional questions regarding feedback and contact information.[28]

5.2 Results of the Online Survey

In total, there were 204 respondents, of which 129 completed the survey. The results of the survey that are presented in this section are reported on all available responses, even if the survey was not completed by the respondent.

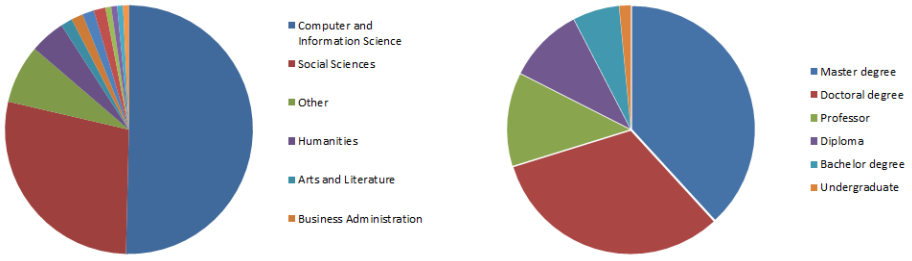
5.2.1 Socio-Demographics and Search Experience

62.6% of the respondents were male, and the age of the respondents ranged from 23 to 79 years (mean=40.3, sd=12.1, N=128).

Information about the respondents' field of study and academic degree can be found in Figures 5.1 (a) and 5.1 (b). The respondents were asked to choose their field of study from a set of 26 options. In total, 12 fields were chosen with the majority of the respondents coming from the field of "Computer and Information Science" (50.4%) and the "Social Sciences" (28.2%). Regarding the academic degree of the respondents, 54.2% replied to have a master's, diploma or bachelor's degree, 32.1% obtained a doctoral degree, 12.2% were professors, and 1.5% of the respondents were undergraduates.

[27] The questionnaire of the survey can be found in Appendix B.

[28] Two pretests with colleagues of the department Knowledge Technologies for the Social Sciences were carried out.



(a) Field of study (b) Academic degree

Figure 5.1: Distribution of the respondents field of study and academic degree (N=131).

When asked to rate their experience in the usage of DLs, 24.4% considered their experience as expert, 42.7% as high, 21.4% as moderate, 10.7% as little and 0.8% had no experience at all.

Alongside their experience, the respondents were asked how often they used DLs and Google Scholar using a five-point Likert scale ranging from very rarely to very often. The results are displayed in Figure 5.2. 59.6% of the respondents use DLs ”often” or ”very often” (median=4, mode=4, N=129) and 71.8% use Google Scholar ”often” or ”very often” (median=4, mode=5, N=129).

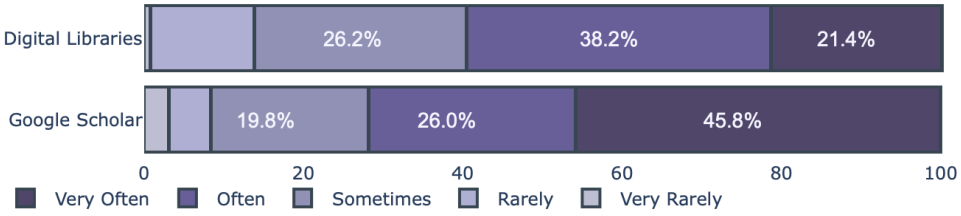


Figure 5.2: Percentaged usage of Digital Libraries and Google Scholar (N=129).

5.2.2 Journal and Conference Run

The first part of the survey addressed general questions concerning journal and conference runs. A five-point Likert scale with different item labels for each question was employed. The following example shows the item labels regarding a question about usefulness in which the labels range from: not at all useful (1), rather not useful (2), neither useful nor not useful (3), rather useful (4), very useful (5). For each of the questions, the negative item was left-aligned.

The questions and the corresponding results regarding a journal and conference run are displayed in Table 5.1 (each item-label is highlighted).

Table 5.1: General questions concerning a journal and conference run (N=156).

Task	Mdn	Mode	M	SD
How **useful** are conference proceedings or journals as a source for relevant documents during your search task?	5 (very useful)	5	4.31	0.89
How **satisfied** are you with the support of current Digital Libraries (e.g. ACM DL, Web of Science) browsing through conference proceedings or journals?	3 (neither satisfied nor unsatisfied)	4	3.27	0.9
How **important** is the quality of a conference (ranking) or a journal (e.g. the impact factor) for your confidence in the source?	4 (rather important)	4	3.44	1.06

Furthermore, it was investigated how frequently the respondents used journal or conference runs. Two questions with items ranging from "never" to "very often" were asked: a) "How often do you browse through conference proceedings or journals to find relevant documents?". The question was then repeated but from a modified situation: b) "After finding a document (e.g. ACM DL, Web of Science) that is relevant for your current search task: How often do you browse through the conference proceedings or journals the document was published in?".

The results for the usage of journal or conference runs are displayed in Figure 5.3. A journal run without preceding document (denoted as Journal Run) was selected "often" or "very often" by 54.9% (median=4, mode=4, N=142) of the respondents. Regarding a journal run with a preceding document (denoted as Journal Run 2), 35.2% replied to use this search activity often or very often (median=3, mode=3, N=142).

If the respondent replied that he/she never or rarely browsed through conference proceedings or journals, he/she was asked to further explain the decision using an open-ended question. 13 response were given for the variant without a preceding document (Journal Run), and 23 responses for the variant with a preceding document. The most frequently given responses concerning the journal run without a preceding document were a preference for query searches, predominantly mentioned in conjunction with Google, Google Scholar, or the Web of Science. The most frequent responses regarding the journal run with preceding document were a preference for searching and addressing the topic of a journal.

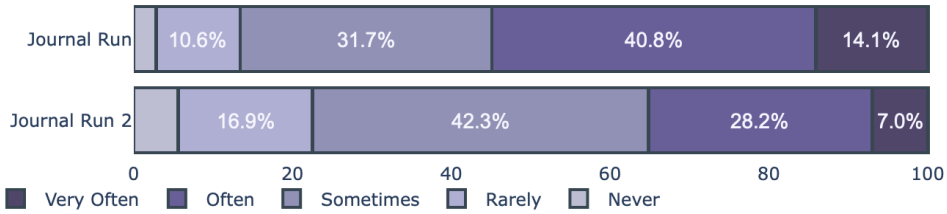


Figure 5.3: Usage frequency of a journal and conference run when looking for related documents without (Journal Run) or with preceding document (Journal Run 2) (N=142).

Some (shortened) examples that were provided by the respondents are presented in the following:

- "Often the content the journal is covering is broad compared to my search query."
- "Proceedings are usually not that close to a specific topic."
- "Papers in the same proceedings volume or the same journal issue are often merely related to each other in a very superficial way.."
- "Only very occasionally if the topic of the conference happens to be on something very relevant to me."
- "It usually takes too much time, browsing for additional related literature on a search engine is faster."
- "I rather search for documents citing the article/paper or its references. An exemption might be when the found article is published in a special issue on a certain topic."
- "I don't do it to keep focus. When I do it, it's out of simple curiosity or to find out if the article is part of a special issue."

Other reasons provided by the participants were concerned with orientation and learning:

- **Orientation**: "This is really something to do when one has no idea what one is looking for and needing hints.."
- **Learning (topic)**: "I only do that when I want to get new keywords to search for when I start a new topic.."
- **Learning (community)**: "I browse journals to get a feeling for a certain community or conference.."

Table 5.2: General questions concerning citations and references (N=140).

Task	Mdn	Mode	M	SD
How **important** is the number of citations a document has received to you?	3 (neither important nor unimportant)	4	3.33	0.91
How would you rate the **usefulness** of citation rankings (e.g. h-index) where documents are ranked by the number of received citations?	3 (neither useful nor not useful)	4	3.23	0.99
Assuming there is a key document in a particular field. How **important** is it to you to find central authors citing that particular document?	4 (rather important)	4	3.60	1.07

Overall, one can summarise that journals and conference proceedings are considered very useful sources (median=5, mode=5, N=156). The respondents often browse through a journal or conference proceedings in order to find relevant documents but less frequently as a follow-up search activity starting from a relevant document. It is assumed that journals and conference proceedings are not that valuable when looking for related material with respect to an already retrieved relevant document.

5.2.3 References and Citations

In the second part of the survey, general questions concerning the usage of citations and references were asked. Again a five-point Likert scale using different item labels was employed. The questions and the respective results are displayed in Table 5.2.

Furthermore, it was examined how frequently the respondents used references or citations after finding a relevant document.[29] Two questions with items ranging from never to very often were presented: a) "Starting from a relevant document: How often do you use **references** to find other relevant documents for your search task?" and b) "Starting from a relevant document: How often do you use **citations** to find other relevant documents for your search task?". The results of these two questions are displayed in Figure 5.4.

Regarding citations, the options "often" to "very often" were selected by 65.7%

[29] An example diagram was presented to the respondents to clarify the relationship between citations and references.

(median=4, mode=5, N=140) of the respondents and by 82.1% (median=4, mode=4, N=140) regarding references. If the respondent replied that he/she never or rarely used citations or references, he/she was asked to provide some additional information on why. For references, only one relevant response was available. He/she replied to prefer semantic tools. With respect to using citations, there was an overall agreement that they are more difficult to find and therefore not that commonly used.

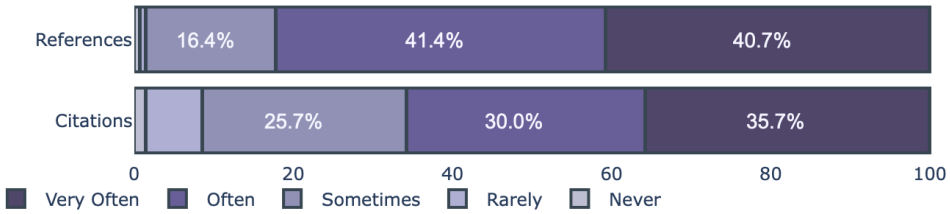


Figure 5.4: Results for the usage frequency of citations and references when looking for related documents. (N=140)

Overall, one can summarise that citations and references are commonly employed to find other relevant documents. 65.7% of the respondents used citations, and 82.1% used references often to very often. The number of citations a document has received has neither been important nor unimportant to the respondents. However, it is considered as rather important to find authors who cite a particular key document in a certain field. The central tendency regarding features like the h-index or the general citation count ranges between the mid-point and a rather positive tendency.

5.2.4 Stratagem Usage

In Bates (1990), six example stratagems were proposed (see Section 2.4). In order to study today's relevance of these search activities, the following search scenario was presented to the respondents:

> "Please consider the following scenario. You want to find out about the current state of the art in a particular field. You have already found one document that is useful to your current work task."
> Which of the following search activities do you perform to find other related materials? Please order the following options from best to worst.

The scenario was designed to put the user into a situation in which he or she utilises a stratagem as a follow-up search activity. Although this question clearly aimed at testing the applicability of the presented variants of stratagem browsing, it was decided not to include the term "browsing". The main reasons for this decision were

Table 5.3: Stratagem usage for the given scenario. Mean values range from lowest rank (6) to highest rank (1) (N≥125).

Ranking option	Mean	SD	Median	Mode
Follow references in the current document	2.38	1.24	2	2
Inspect the list of documents that cite the current document	2.79	1.50	2	2
Keywords that describe the current document as search terms	2.82	1.63	3	1
Look for papers the authors of the current document has/have published	3.46	1.21	3	3
Browse the conference/journal the current document was published in	4.10	1.53	4	5
Browse a thesaurus to find classification terms related to the current document	5.21	1.30	6	6

two-fold: a) to avoid over-specification of the situation, which would be necessary in order to introduce the term browsing and in turn b) to keep the scenario reasonably short. Alongside this scenario, the respondents were given the six example stratagems and then asked to order all these activities from best to worst regarding the given scenario using a drag and drop user interface. The list of stratagems, in conjunction with the corresponding results, are illustrated in Table 5.3.
References, citations, and keywords are the most commonly applied stratagem variants for finding relevant documents to an already retrieved seed document. Again one can see that citations and references are important stratagem variants. This is an observation that was evident throughout the survey so far. While journals and conference proceedings were considered as useful in order to find content that is relevant for a researcher in general (see Section 5.2.2), they do not appear to be considered useful as a follow-up search activity in which a relevant seed document is provided as it is the case in the given scenario.

In addition to the task described above, the respondents were given the non-mandatory opportunity to provide other search activities that they employ to find related material. Using an open-ended question, the respondents provided various other types of search activities.

Some of the responses are collected in the following:

- "I ask colleagues for further literature hints."
- "Sometimes, I check the corresponding Wikipedia article with its references - I use phrases for search input (e.g., for names) - I check number of citations, and year of publication..."
- "I look at its altmetric page to check what people say about it."

5.2.5 Organising Journal Articles

In addition to the goal of obtaining knowledge about the use and usefulness of different stratagem variants, the survey was designed to pretest the hypothesised need for contextualised stratagem browsing. This pretest was conducted by providing the respondents with a search task scenario along with six options to rank the content resulting from a journal run.
The respondents were given the scenario displayed in Figure 5.5, in which they are asked to look for related material in a journal named *Addiction*.

Please consider the following situation: You are about to write an essay about 'Alcohol Consumption in Germany and its Demographic Distribution'. You start your search by entering the search terms 'alcohol consumption germany'. You find a relevant document (see illustration) that was published in a journal named 'Addiction'. After reading the document you want to see more material from that particular journal.

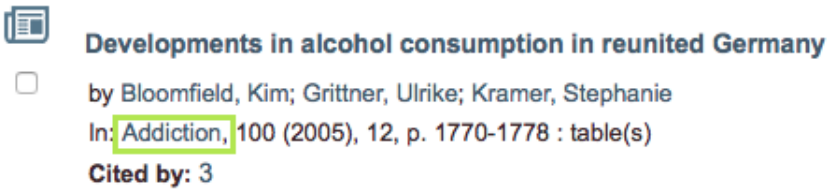


Figure 5.5: Scenario for the task on organising journal articles.

Alongside that journal run scenario, six randomly arranged options on how to rank articles from that particular journal were presented to the respondents. Four of the six options (issue date, title, author, and citation count) are well known and widely implemented in today's DLs. The two remaining options are contextualising the ranking of the articles from that journal: One option ranks the journal articles by the previously entered query ("alcohol consumption germany"), and the other option ranks the journal articles by similarity to the current relevant article the user inspected based on the title ("Developments in alcohol consumption..").[30]The respondents were then asked to order the six ranking variants from best to worst using a drag and drop user interface. The results of this task are displayed in Figure 5.6.[31]

[30]The idea behind the two contextualised ranking options had already been discussed in a position paper (Carevic & Mayr, 2015).
[31]A tabular representation of the results is appended in A.1.

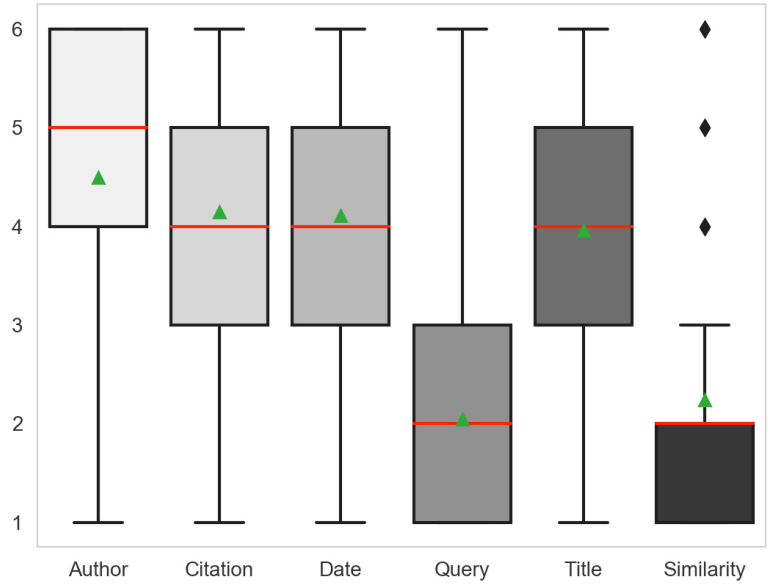


Figure 5.6: Results for the task on organising journal articles. Values ranging from the lowest rank (6) to highest rank (1) (N≥128).

The results of this task show a strong tendency towards a contextualised ranking of journal articles and are in line with the hypothesis of the thesis. Both ranking options, that are based on previous search activities (denoted as Query and Similarity), clearly outperform the other four ranking options, that are well known and commonly used in DLs. Overall, a ranking based on the previously entered query term was ranked slightly better (mean=2.08, sd=1.34, mode=1) compared to the ranking option based on similarity to the current relevant article (mean=2.23, sd=1.32, mode=2). The four generic ranking options, on the other hand, performed notably worse with a mean ranging between 3.95 and 4.42.

5.3 Conclusion

Primarily, the survey provided insights into the usefulness of stratagem browsing. The journal run and conference proceedings were considered as very useful sources by the majority of the respondents (median=5, mode=5, N=156). 54.9% of the respondents browse through a journal or conference proceedings often to very often, in order to find relevant documents, but only 35.2% use this stratagem as a follow-up search activity starting from a relevant document.
Particularly informative were the responses that were given in the open-ended questions, which offered valuable insights into potential reasons to reject a certain stratagem. A key observation that was made from these responses is the *topical broadness*. A very commonly given response was that the employment of a stratagem depends on the underlying content. Having very generic keywords or

a topically broad journal has a strong influence on the decisions whether or not a stratagem variant is employed.

The task on organising journal articles showed a strong tendency towards a need for contextualised ranking methods. The respondents considered the two contextualised ranking options notably higher in comparison to the rather traditional ranking options. This pretest is an indicator for the usefulness of contextualised stratagem browsing.

The results of the online survey provided a general overview of the usage of stratagem browsing across a broad range of disciplines and academic degree. To gain more in-depth knowledge of these types of searches, a user study was conducted in which the stratagem usage is observed in a controlled experiment. The setup and the results of this user study are presented in the following section.

6

User Study on Stratagem Browsing in Digital Libraries

The online survey was valuable to obtain first insights into the usefulness of stratagem browsing. The results of the survey suggest that there is a potential need for contextualised browsing. On the downside, the survey is missing a qualitative perspective that could lead to a more profound understanding of the users' intentions and expectations, while employing a stratagem. Furthermore, the survey does not provide information regarding information-seeking behaviour during an actual search task. Instead, the results rather represent a user's view on past search behaviour. A respondent of the survey might, for instance, agree that he/she is considering a journal run as useful, but insights regarding the operationalisation of such a stratagem can not be provided by the survey results.
Hence, to gain more advanced knowledge about stratagem browsing, it is necessary to move towards a qualitative perspective. One approach to obtain this kind of knowledge is to observe participants while solving a search task.

For this reason, a user study with 32 participants from the social sciences domain was conducted. The participants were asked to solve a predefined search task using the Digital Library Sowiport (see Section 4.4). The participants were provided with a topic on *educational inequality* and a seed document that was relevant for the given search task. The search task was then to look for content that is similar to the given seed document in a limited time slot of 10 minutes.
While solving this task, the search activities of the participants were observed to gain knowledge about the actual usage of stratagem browsing in a controlled environment. In addition to this, the gaze data of the participants were recorded using an eye-tracking device to gain insights on the perceived relevance of certain search activities. Different areas of interest within a document were defined in order to measure the dwell time and the number of fixations to determine the perceived relevance.

The participants were recruited from two levels of experience. The first 16 participants were students, and the other 16 participants were postdoctoral researchers from the social sciences. The intention of inviting two distinct groups of participants was to unfold potential differences between these two groups in terms of search behaviour and gaze data. An experienced researcher, for instance, may be aware of a renowned journal that is central to his/her research field and may be

encouraged to browse this journal more often while an undergraduate student may not be aware of these journals and thus, consider a journal run as less useful.
This study addresses the following research question:

> **RQ:** What are the most frequently applied stratagem browsing variants in a state-of-the-art DL, and how is their usage in comparison to other search strategies like, for example, query searches? How is the perceived relevance of stratagem browsing opportunities?

Besides analysing gaze data and measuring the usage frequency of stratagem browsing variants, a post-questionnaire was conducted in which the participants were asked to provide reasons for their use of a stratagem, rate its usefulness, and provide reasons for their usefulness rating.

The present section is based on Carevic et al. (2018a) that has been published in the *International Journal on Digital Libraries (IJDL)*.

6.1 Study Design and Setup

The user study took place in single sessions with a duration of about 30 minutes each. It was ensured that the conditions were the same in every session. The experiment was carried out on a laptop which was connected to an external 22”-monitor as the stimulus monitor for the participants. The display of the laptop was used for observation. An "SMI iView Remote Eye-tracking Device 250" installed at the bottom of the stimulus monitor was used to capture the gaze data. The screen activities, as well as the eye movements, were recorded by the corresponding software SMI Experiment Suite 360. A nine-point calibration with a visual and quantitative validation was used to ensure the quality of the gaze data, and a sampling frequency of 250Hz for recording the eye movements was defined. All participants used Mozilla Firefox to access Sowiport during the task.

6.1.1 Scenario

All participants had to accomplish the same task, which is shown in Figure 6.1. The task was developed to serve as a simplified version of a simulated work task (Borlund, 2003) (see Section 2.6.1). Accordingly, it was avoided to over-clarify the situation and to include information about the number of documents to retrieve. The task much rather aimed at providing guidance towards the problematic situation and the environment in which the scenario is placed.

The simulated work task was developed to be as close to a realistic work task as possible. However, two limitations were necessary: 1) The participants were required to use Sowiport for their literature search and thus, by the time the study took place, they had access to about 9 million documents from 18 different databases of

You are writing an essay about education inequality. You have already done a first literature search and found a document that you consider very relevant.

In order to find more literature for your essay, you are looking for similar documents now.

*You have **10 minutes** left before you have to pack your bags and go home. You decide to use the time for **finding similar documents**. For this purpose, you start in Sowiport with the relevant document that you have already found.*

Collect the documents that you consider relevant by opening them in new browser tabs.

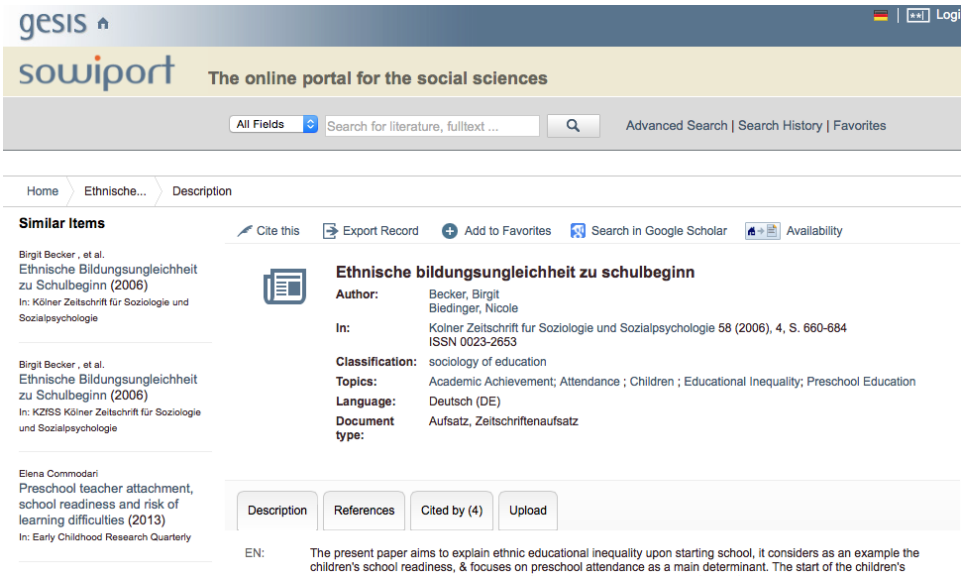
Figure 6.1: The scenario and search task for the user study.

which six are English-language ones, 2) the participants had a limit of ten minutes to solve the task.

At the beginning of the search task, each participant was provided with the same seed document titled *Ethnical educational inequality at the start of school* (German original title: *Ethnische Bildungsungleichheit zu Schulbeginn*). Details of the seed document and its descriptive metadata can be found in Figure 6.2 (a). The document was published by two authors, contained five keywords, one classification term, and was published in a German journal for sociology and social psychology. Each of these descriptive metadata were implemented as hyperlinks and could be utilised for further exploration (see Section 4.4.3).
Figure 6.2 (b) and (c) depict separate tab views which allowed the participants to browse through citations (four citations all accessible in Sowiport via hyperlinks), reference information (70 references of which 33 were accessible in Sowiport via hyperlinks) and read the abstract of the given document. Additionally, the participants had access to the full text of the seed document, which could be accessed via a hyperlink provided by Sowiport. Furthermore, the participants were provided with ten document recommendations (see Section 4.4.1): five of them were provided by the SOLR *more like this function*, and five were associated documents that were published in the same journal.

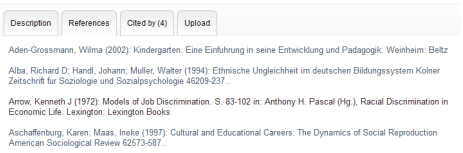
The document displayed in Figure 6.2 was chosen as a seed because: 1) it comprises a reasonable number of descriptive metadata, including citations and refer-

ences which facilitate further exploration, and 2) because it was assumed that the topic of the document was balanced to be neither too specific nor too generic for the participants coming from the social sciences.



(a) Seed document

(b) Citation tab



(c) Reference tab

Figure 6.2: The seed document used in the user study, including the available descriptive metadata (a), citations (b) and references (c) (shorted example).

6.1.2 Procedure

Every session followed a detailed protocol to ensure that the conditions were identical for each participant. After a short introduction to the background of the study and the procedure, each participant signed a consent form about the recording of the gaze data as well as the screen activities. The user study was divided into three parts: a pre-questionnaire, the search task, and a post-questionnaire. In the following, these are described in more detail.

Pre-questionnaire

In the first part of the study, the participants were handed out a pre-questionnaire, in which they were asked to provide socio-demographic information such as age and gender, as well as their search experience. Regarding their search experience, the participants were asked about the overall use of DLs, the use of Web of Science, Sowiport and additionally Google Scholar. In accordance with the online survey presented in Section 5, a five-point Likert scale was employed with items ranging from 1=very rarely to 5=very often.

Search task

During the second and main part of the user study, the participants were provided with the simulated work task that they were asked to solve. For this, the eye-tracking technology, as well as the task, were briefly explained. The participants were asked to take an upright and comfortable position in order to move as little as possible while solving the search task. Then, the monitor with the eye-tracker was aligned to their height and eye level, so that their eyes were positioned in the centre of the area captured by the eye-tracking device. Finally, the eye-tracker was calibrated.
First, the scenario was displayed to the participants. After reading the scenario, the participants proceeded to the seed document (see Figure 6.2) for the search task. The search activities and eye movements were followed on the observation screen while the participant was solving the task. At the same time, each stratagem usage was noted down as well as which information a participant looked at.
The focus was again on the six stratagems that were derived from Bates (1990) and that had already been part of the online survey (see Section 2.4). Those six stratagems were modified slightly to fit the use case of Sowiport. The list of investigated search activities comprised:

- Inspect the list of documents that the current document references.
- Inspect the list of documents that cite the current document.
- Follow keywords that describe the current document.
- Look for papers the author(s) of the current document has/have published.
- Browse the conference/journal the current document was published in.
- Follow classification terms related to the current document.

In addition to stratagem browsing, it was observed how often the participants clicked on recommended documents, and how often they submitted a query.
After ten minutes, the participants were asked to end their search and review the documents they had collected in browser tabs. They were instructed to close the documents they did not consider relevant anymore.

Post-questionnaire

In the third and last part of the session, the participants were asked to fill out a post-questionnaire, that was handed out after the simulated work task had been completed. The goal of the post-questionnaire was two-fold: First, it covered some general questions related to the difficulty of the task, the difficulty in finding relevant documents, the ability of the system in providing relevant documents, and the users' opinion of how successful they were in solving the task. The second goal of the post-questionnaire was concerned with the central idea for conducting a user study, which was to obtain more qualitative feedback regarding the usage of stratagems. This was approached by asking the participants to assess their stratagem usage. In order to avoid confusion and to eliminate potential bias, neither the simulated work task nor the questionnaire mentioned the term stratagem. Instead, the stratagems were paraphrased according to their characteristics. An example for a participant that followed references in a document is displayed below:

- You solved the task by following the references of the document.
  - Why did you choose this method?
  - How useful was this method (1=not at all useful, 5=very useful)?
  - Please give reasons for your choice of usefulness.

Information aimed at obtaining insights about why a participant had looked at a particular information but did not further employ that potential strategy was collected similarly:

- You looked at the classification terms of a document but did not follow them.
  - Why didn't you use this method?

6.1.3 Eye-Tracking Methodology

For each participant, full-screen records were taken during the search task showing their gaze data and capturing the navigation bar of the web browser as well as overlaying dynamic elements. These records were used to reproduce the user journey of each participant. The employed search activities (stratagems, queries, and recommendations) and the collected relevant documents per stratagem variant were counted.[32]

In order to analyse the participants' gaze data, certain pre-processing actions needed to be performed. First, following Reichle et al. (2012), a fixation time threshold of 104 ms was defined, which indicates the start of lexical processing. Second, for each participant, the stable eye was determined by comparing the scan paths of the

[32] If a participant combined two search activities like, for example, keywords and queries, only the first search activity is reported.

left and the right eye of a participant. Third, the data were reduced to three stimuli: 1) the seed document and its descriptive metadata, 2) the reference list of the seed document, and 3) the citation list of the seed document.
Restricting the results of the eye-tracking experiment to the seed document was necessary, since it is the only document that each participant visited before starting their explorative and individual search through Sowiport. Only the timespan between entering the document and the first interaction with a clickable object other than the citation and reference tab was taken into account. This way, overlaying elements, such as the search term recommender (see Section 4.4.1), that would have distorted the gaze data, were eliminated. An additional constraint is that gaze data is only considered during the first visit on each of the three stimuli.
Heat maps were employed to analyse the eye-tracking data, in order to obtain a visual representation of the participants' gaze data. Additionally, areas of interest (AOI) were defined to quantify the number of fixations and the dwell time on particular elements of the seed document.

6.2 Results of the User Study

The user study involved 32 participants, who were divided into two groups according to their academic degree. The first group included 16 students, and the second group included 16 postdoctoral researchers. All participants were social scientists from different fields of study that had been recruited via e-mail and personal recommendation. The students were aged between 22 and 35 years (m=26.38, sd=3.76). 75% were female. About 19% of the students had no academic degree yet, 69% held a bachelor's degree, 15% held a master's degree, and 6% had a diploma. The postdoctoral researchers' age ranged from 30 to 62 (m=40.19, sd=9.23). In this group, 50% were female.

6.2.1 Search Activities

The total usage frequency of the eight potential search activities is displayed in Figure 6.3.[33] The figure contains information about the usage frequency of a search activity (purple) and the number of participants that employed a particular search activity (light-grey).
In total, the participants utilised a stratagem variant 137 times. The most frequently used stratagems were: using keywords (50), following references (27), and following citations (26). The interest in utilising citations and references was even more evident when considering that not all documents encountered during the sessions supported these types of searches. A rather low frequency was found for the journal run, which has only been utilised by two participants.
The usage frequency of query searches and recommendations is displayed in the four rightmost bars in Figure 6.3. Queries and recommendations (n=111) were, in

[33] A tabular representation of this illustration can be found in the Appendix A.2.

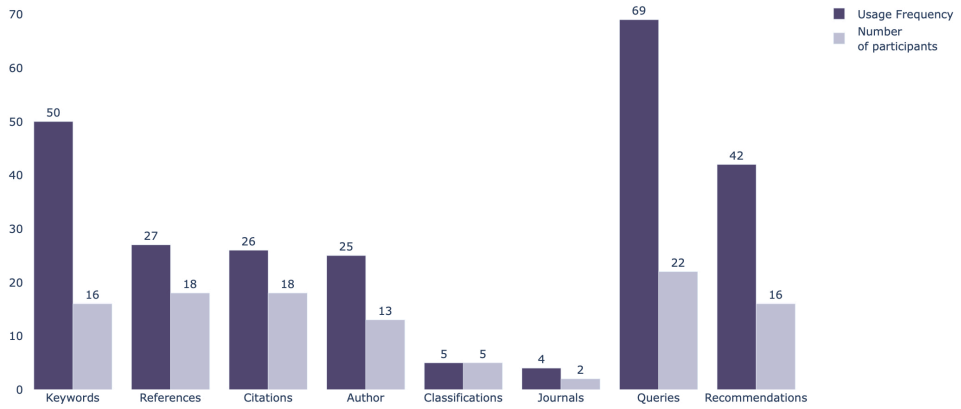


Figure 6.3: Total usage frequency of the six stratagems in comparison to queries and recommendations.

total, less often employed compared to stratagems, even though queries were the most frequent search activity. Surprisingly, only 22 participants used queries at all, which means that 10 participants solely relied on stratagems or recommendations to solve the task.

A possible explanation for the strong reliance on stratagems can be found in the qualitative feedback of the participants, presented in the subsequent section.

6.2.1.1 Feedback on Search Activities

Several participants named little effort as a criterion for using a stratagem and preferred quick and easy steps to find more relevant results during the task. Other frequently mentioned reasons were search habits and time restrictions as reasons for using or not using a stratagem. Participants also mentioned that inspecting the list of references and the employment of a journal run would have cost them too much time so that they decided not to use these stratagems during this particular search task. Several participants stated that they did not perform a journal run because the particular journal was too general for the task. Corresponding to that, participants frequently mentioned that the specific author was crucial for their decision to search for an author. If the participants surmise or know that an author focuses on the topic of interest, they are more likely to perform an author run. Regarding the classification terms and the keywords, the participants said that they only use them for further search if they are relevant for the topic and if the classification terms/keywords are neither too general nor too specific. Additionally, the participants named several document related factors as reasons for employing or rejecting a certain stratagem.

6.2.1.2 Usefulness of Search Activities

In the post-questionnaire, the participants were asked to assess the usefulness of the search activities employed during the search task.[34] The results of this inquiry are displayed in Table 6.1.

Table 6.1: Usefulness of the employed search activities. Results range from 1 (not at all useful) to 5 (very useful).

Type	Search activity	N	M	SD	Median	Mode
Stratagem	References	16	4.19	0.7	4	4
	Classifications	5	3.80	0.75	4	3
	Authors	12	3.75	0.92	4	4
	Citations	16	3.69	0.98	4	4
	Keywords	16	3.56	0.86	4	4
	Journal	2	3	0	3	3
Other	Queries	21	3.38	0.72	3	3
	Recommendations	16	3.19	1.01	3	4

The participants considered references, classifications, and author searches as the most useful stratagems for the given task, while the usefulness of journals, recommendations, and queries was comparably low. The reasons for the usefulness ratings correspond to the reasons for the usage or rejection of a search strategy. The participants explained that references gave them quick access to a large set of documents that were topically related. Critical aspects of references were that they often comprise large sets of referenced articles and thus, they were considered as time-consuming. Furthermore, some of the documents contained in the list of references were considered to be too specific.
Similar responses were provided for the citations which also provided quick access to related documents. A frequently stated critique was that the number of citations was rather low (four citations in the seed document). Having a seed document with a larger number of citations would certainly help to resolve this issue.
Most participants that employed an author search rated this activity as useful because authors had often published similar and more recent documents on the same topic. Frequent comments on the usefulness of keywords were a precise search on topically related documents and a good starting point to the given subject. Topical mismatches and broadness were mentioned as downsides by the participants. Similarly, the journal was considered to be too general for the topic at hand.
Positive aspects of the usage of queries were a feeling of control and a quick way to select topically adequate (neither too broad nor too narrow) documents. On the downside, it was mentioned that the number of results retrieved via query searches was too high.

[34] Not all participants responded to each search activity, which results in minor differences between the actual usage of a search activity (see Figure 6.3) and the usefulness rating.

Figure 6.4: Percentaged usage frequency of search activities within a session limit of ten steps.

6.2.1.3 Stratagem Usage Frequency by Session Step

The participants changed their search strategies several times throughout the study. In Figure 6.4, the percentaged usage frequency of each observed search activity is displayed, taking only the initial ten steps into account.[35] In this section, a *step* is referred to as an individual search activity performed by a participant on the stratagem, query, or recommendation level.

The two most frequent initial search activities starting from the seed document were keywords and citations, both applied by 21.8% of the participants. Journals (none of the participants), classifications, and authors (both 3.1% of the participants) were the least frequently applied search activities. The frequent usage of stratagems is particularly visible when considering just the initial search activities (Step 1). At this stage, stratagems account for nearly 70% of all initial searches.

With increasing session steps, one can observe a change in the participants' search strategy. Queries, which were only used by 15.6% of the participants in the first session step, were the most frequently applied search activity in the remaining three groups. The other search activities varied within the session steps like, for instance, the author run and the usage of keywords.

In the post-questionnaire, several participants stated that the success of their previous search influenced their search strategies. If they had not found a satisfying

[35] A tabular representation can be found in the Appendix A.3.

number of relevant results yet, they would have been more likely to use different strategies than they had before.

6.2.2 Eye-Tracking Results

The results presented in the previous section showed that stratagem searches were frequently employed in the context of this study. The results from the qualitative feedback, however, revealed that the employment of a stratagem strongly depends on the stratagems underlying content. A journal, for instance, that presumably covers a wide range of topics may be too broad to discover something similar. Therefore, even though the journal run itself is a legitimate and useful search activity, it may not be suitable for the given task.

In order to address this issue, it could be of benefit to study the so-called perceived relevance of the six stratagems by looking at gaze data which was provided by the eye-tracking device. The assumption is that relevant stratagem browsing opportunities attract a higher number of fixations and consequently result in longer fixation periods. Under this assumption, the analysis of gaze data has the potential to provide knowledge about stratagems that goes beyond studying only the actual stratagem usage and enables, to some extent, the approximation of a stratagems perceived relevance even if the stratagem was eventually disregarded.

Figure 6.5 shows a heat map of the gaze data and its distribution across the seed document for the first stimulus before any interaction with the system was performed. On the heatmap, one can see a focus along the descriptive metadata, most intensely in the area of the seeds authors and publishing information. Furthermore, one can recognise that the distribution is mostly focused on the first entries of each class of metadata. This is particularly visible for the keywords where only the first items appear to be fixated while the focus decreases for the remaining ones. Other areas with high fixations were: classifications, citations, references, access to full-text, and recommendations. Surprisingly, the distribution along the abstract of the seed document was rather low.

6.2.2.1 Areas of Interest

Heat maps provide a quick overview of gaze data for the entire document. A more detailed examination of gaze data is provided in the following section in which the number of fixations and the corresponding time spent were quantified by defining *areas of interest* (AOI). The defined AOI are displayed in Figure 6.6.
The main focus of this study is on stratagems and therefore, the AOI mainly concentrated on keywords, classifications, publishing information (journal or conference proceedings), authors, citations, and references. Alongside these stratagems, the fixations along the search bar (query formulation) and the recommendations were incorporated as well.

Figure 6.5: Heat map showing the distribution of gaze data among the seed document for all 32 participants.

In the AOI illustrated in Figure 6.6, the following two features were measured:

- Dwell time: The time spent on a certain AOI during the first visit.
- Number of fixations: The total number of times a certain AOI was examined by a participant.

In order to keep the analysis of the AOI reasonable, only the three stimuli of the seed document (descriptive metadata, citations, and references) were considered. Furthermore, only the timespan during the entry and the first interaction on each stimulus were taken into account.

The results of the AOI analysis are displayed in Table 6.2. The most frequently fixated AOI were references (m=178.76), recommendations (m=12.57), and citations (m=12.2). Other areas, like the search bar (m=5.11) or the classifications terms (m=3.44), showed a rather low number of fixations. A similar result was observed

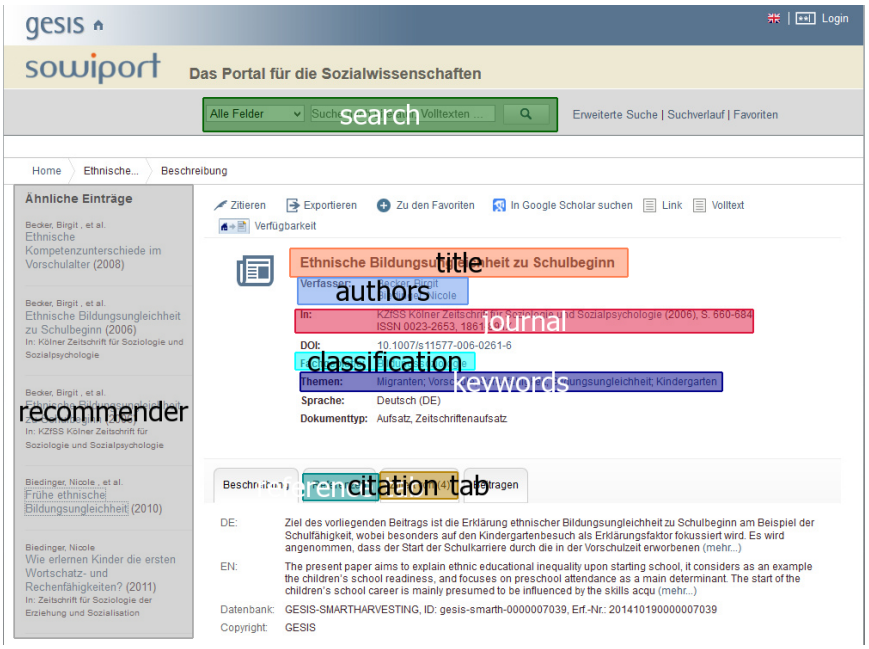


Figure 6.6: Areas of interest for the given seed document.

with respect to the dwell time. The participants spent the most time on inspecting the list of references (m=57.59s) and the recommendations (m=4.0s). However, it should be noted that citations and references were both displayed in a separate tab-view within Sowiport. Therefore, these two AOI were measured separately. One obvious reason for the increased dwell time and the high number of fixations for the references is the large size of this AOI comprising a total of 70 references.

Along with the gaze data, Table 6.2 additionally contains the total usage frequencies of the possible search activities. A comparison of the gaze data (columns four and five) and the actual search activity usage (see column six) indicates minor differences between the two groups. A noticeable difference, however, can be found for the journal. Although a journal run was employed by none of the participants, the number of fixations and the dwell time indicate a comparably intense examination which is even higher than the most frequently applied stratagem (keywords).

A possible explanation for this can be found in the qualitative answers from the post-questionnaire. Being asked, why they looked at descriptive metadata, but did not utilise it, several participants stated that metadata like authors, keywords, and classification terms are general information that is important for getting to know a document, regardless of their utilisation during the search session. Concerning the journal, as stated before, the majority of the participants explained that the seed documents journal was too general to be useful for the search.

Table 6.2: Number of fixations and dwell time on AOI in the seed document.

Type	Search Activity	N	Mean no. of fixations	Mean dwell time (s)	Utilised by
Stratagem	Keywords	16	4.50	1.38	7
	Classifications	18	3.44	0.99	1
	Journal	18	5.44	1.74	0
	Authors	29	5.68	1.58	1
	Citations	10	12.20	3.34	7
	References	13	178.76	57.59	6
Other	Recommendations	14	12.57	4.0	5
	Search	9	5.11	1.16	5
	Title	26	4.96	1.21	-

6.2.3 Diversity in Participants

The user study was made up of two groups of participants: students and postdoctoral researchers, both from the social sciences. The advantage of having a distinct set of experiences among the participants is that it enables one to look for differences in their information-seeking behaviour.
The results of the comparison between the students and the postdoctoral researches regarding all aspects of this study are displayed in Table 6.3.

In the columns two and three, one can see the mean dwell time (in seconds) that the participants spent on a specific AOI with respect to each of the two groups. The columns four and five represent a comparison of the two groups with respect to the mean number of fixations per AOI.
Overall, one can observe a general tendency towards a more intense focus on the AOI by the group of postdoctoral researchers. With the exception of citations and keywords, participants having a Ph.D. spent more time on the defined AOI and fixated these more often in comparison to the students. Differences between the two groups could be found in the AOI journal. The postdoctoral researchers spent noticeably more time on the journal AOI and fixated it more often than the group of students. The last two columns illustrate a comparison between the students and the postdoctoral researches regarding their usage of stratagems, queries, and recommendation. Postdoctoral researchers performed 128 search activities in total of which keywords (35) and queries (33) were the most frequent. Students, on the other hand, performed fewer search activities (120) most frequently in the form of queries searches (36) and by using recommendations (24), while a journal run was utilised by none of the students.
The most noticeable difference between the two groups was found for keyword searches which were utilised 15 times by the students and 35 times by postdoctoral researchers. Some of the participants indicated that prior knowledge and experience play a vital role when deciding for a certain search activity. If the

Table 6.3: Comparison of students and postdoctoral researchers with respect to search activities, number of fixations, and dwell time on AOI in the seed document.

	Avg. Dwell Time (s)		Avg. Fixations		Usage Freq.	
AOI	Students	Postdocs	Students	Postdocs	Students	Postdocs
Authors	1.05	2.01	4.92	6.31	9	16
Journal	0.63	2.85	2.55	8.33	0	4
Classifications	0.70	1.27	2.88	4	2	3
Citations	3.65	3.03	14	10.4	17	9
References	38.00	69.83	116.4	217.7	17	10
Keywords	1.27	1.45	5.33	4	15	35
Rec.	2.31	7.14	7.7	21.2	24	18
Queries	0.97	1.55	4.83	5.66	36	33
Title	0.90	1.57	4	6.08	-	-

participants surmised or knew that an author focused on the topic of interest, they will be likely to perform an author run. This could be one possible explanation for the more frequent usage of an author run within the group of postdoctoral researchers which is assumed to be more experienced than the group of students. Although differences in search behaviour between the two groups of participants were found, none of the results was of statistical significance, according to a non-parametric Mann-Whitney U-Test, and a Bonferroni corrected significance level of $p^* = 0.0056$, number of hypotheses tested $m = 9$.

6.2.4 Organising Journal Articles

During the online survey, the respondents were given a task on organising journal content (see Section 5.2.5). This task was repeated with the 32 participants of the user study, in order to collect more impressions on the potential usefulness of the contextualised browsing approach. As a brief recap: In accordance with the online survey, each participant was introduced with the scenario displayed in Figure 5.5, where they were looking for relevant material in a journal named *Addiction*. Alongside the scenario, six ranking options were presented. Four of the six options (issue date, title, author, and citation count) were well known and widely implemented in DLs. The two remaining options were contextualising the ranking of the articles from that journal. One option ranked the articles by the previously entered query term, and the other option ranked the articles by similarity to the current relevant article the user was examining. The results of this task are displayed in Figure 6.7.[36] The results of this second experiment on organising journal articles are in line with the observations made in the online survey. Both contextualised

[36] A tabular representation can be found in the Appendix A.4.

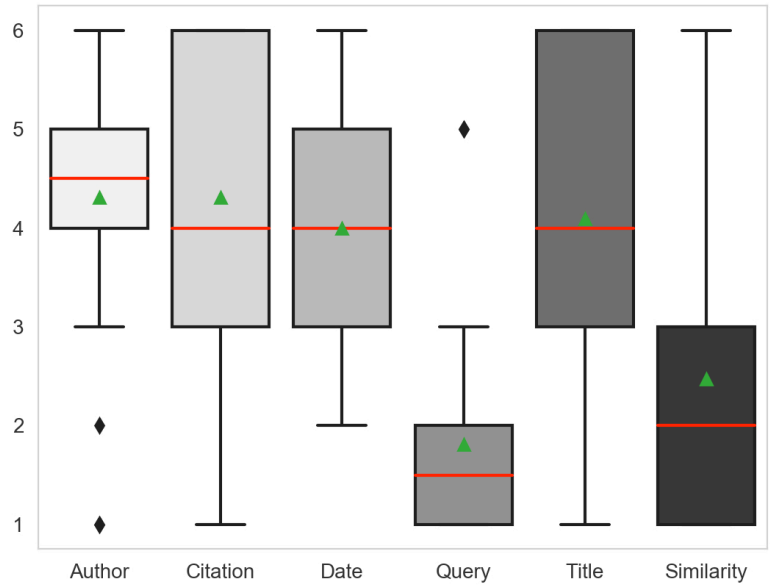


Figure 6.7: Task on organising journal articles in the post-questionnaire of the user study. Values ranging from the lowest rank (6) to highest rank (1).

ranking options (denoted as Query and Similarity) clearly outperformed the non-contextualised ranking variants. A ranking based on the previously entered query terms (m=1.81) is again ranked higher by the participants than the ranking based on document similarity (m=2.47). The results of this experiment again strengthen the hypothesis of the present contextualised stratagem browsing approach.

6.2.5 Post-Questionnaire

During the post-questionnaire, the participants were asked several questions concerning the user study, such as the difficulty of the task and the difficulty in finding relevant documents using five-point Likert scales ranging from (very easy=1 to very difficult=5). Overall the participants considered the task as rather easy (m=2.59) and had only little difficulty (m=2.38) in finding relevant documents.
Considerable diversity was observed regarding the number of retrieved relevant documents, which is displayed in Figure 6.8. On average, the participants retrieved 13.71 relevant documents ranging from 4 up to 41 retrieved documents.
One possible explanation for this rather huge disparity might lie in the description of the task itself. The participants were not provided with any guidelines that helped to determine relevance and were not given any information on the number of documents to retrieve. This is also reflected in the individual responses regarding their impression on difficult aspects of the task. Several participants responded that it is difficult to asses the relevance of a certain document, as many were unfamiliar with the given topic. Another frequent response was that they were unfamiliar with

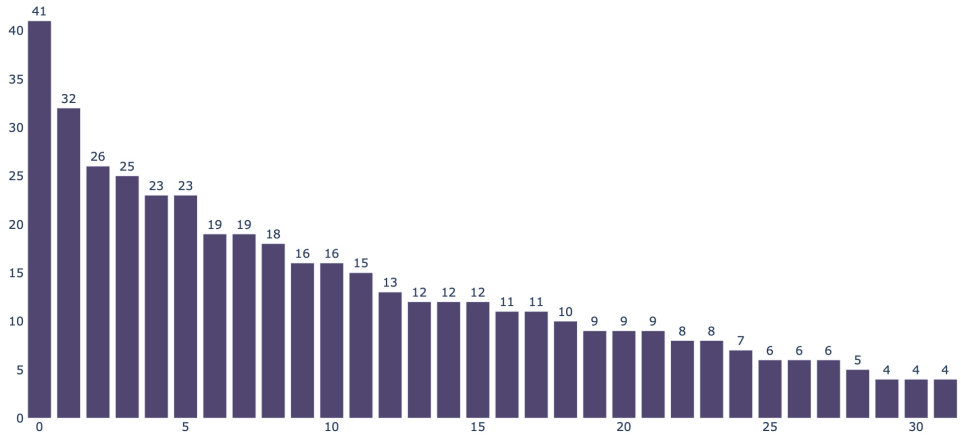


Figure 6.8: Distribution of retrieved relevant documents per participant of the user study.

Sowiport, which made the task difficult for the participants.
Nevertheless, the number of retrieved relevant documents was not part of the investigation, which in fact was focusing on observing the participants without incorporating any aspects regarding the performance of search activities or the quality of retrieved documents.

6.3 Conclusion

First and foremost, the user study presented in this section continued and enriched the findings of the online survey and provided more thorough insights into the search behaviour of social scientists in Digital Libraries. From a purely quantifying point of view, it was found that the participants strongly relied on stratagems (137 times) even more frequently than on query searches and recommendations (111 times). A strong reliance on stratagems was particularly visible for the initial search activities (Step 1). At this point, nearly 70% of all searches were stratagem-based.

Certain stratagem variants, for example, the journal run, were employed only a few times throughout the study. Thus, it may appear that those kind of searches are not applicable for certain search tasks, such as looking for related documents to a given seed. By analysing gaze data, the purely frequency-based observations were enriched, in order to identify a potentially perceived relevance. This additional level of studying the participants' search behaviour showed that although certain stratagems were rarely employed (e.g. the journal run), they do get considered by the participants. Thus, it can be concluded that the relevance of a stratagem can not be determined solely from studying its usage frequency. Other factors that

contribute to the decision whether to employ or reject a stratagem can be found in the post-questionnaire, which provided extraordinary valuable knowledge about stratagems and hence, is a key feature of the present user study.
The reason for the very limited usage of the journal run, for example, was a topical broadness that was not focused enough to find something related. This topical broadness was already mentioned as a downside in the online survey and was visible throughout the entire user study.
At the end of the user study, the participants were given the task on organising journal articles analogously to the online survey. The results of this task again clearly showed that contextual ranking features are considered as better suited in comparison to the traditional ranking features.

The identified topical broadness, along with the re-assessment of ranking journal articles are strong indicators for a need for contextualised browsing. An implementation and evaluation of such a contextualised browsing approach is the main goal of the present thesis and is presented in the upcoming section.

Further readings

The results presented in this section have been published in Carevic et al. (2018a). The published paper contains further investigations, such as a graph representation of search behaviour. These findings are not included in this work, as they are not directly related to the research questions of this thesis. In addition, the results of this user study complemented further research on the origin of search terms during retrieval sessions. The findings of this analysis can be found in Hienert & Lusky (2017).

Part III

Towards Contextual Browsing in Digital Libraries

7

Implementing Contextualised Browsing in Digital Libraries

The online survey and the user study provided valuable insights about stratagems and their usefulness during information seeking. While in general, the respondents and the participants considered those kinds of search activities as useful, one could identify a common downside in their operationalisation. The so-called topical broadness resulting from stratagem browsing was identified as a key factor that influences the decisions on whether or not a stratagem is utilised. A stratagem that presumably covers a very broad topic is less likely to be employed by the users. Contextualisation has the potential to overcome topical broadness by tailoring search results towards the users' search interests.
Another indicator for a potential benefit of contextualisation was derived from the task on organising journal articles. The participants were given a task on organising journal articles based on six predefined ranking options: four well-known ranking options and two contextualised ranking options. The results of this task showed that the participants of both studies favoured the contextualised variants substantially.

On the basis of these observations, the two contextualised ranking options, that were presented to the participants of the survey and the user study, have been implemented in Sowiport. One variant is based on document similarity, and one variant is based on the users' interaction context. This section describes the implementation of these two approaches in detail. Additionally, the non-contextual baseline of Sowiport is presented.
The effectiveness of the contextual variants described here was evaluated in Sowiport. The results of the evaluation are presented in Section 8.

7.1 Notation and Constraints

To understand the details behind the implementation of contextualised stratagem browsing, it is necessary to define the key concepts and the basic notation followed in this section.
A central component during contextualised browsing is the seed document. From a set of documents $D = \{d_1, \cdots, d_n\}$, the seed document d_s is defined as the document currently opened and examined by a user.

A seed document contains a set of potential browsing options referred to as *StratagemAct*, which are defined as follows:

$$StratagemAct(d_s) = \{Keywords_{d_s}, Authors_{d_s}, Classifications_{d_s}, Journal_{d_s}\}. \quad (7.1)$$

$Keywords_{d_s} = \{k_1, \cdots, k_j\}$ is a set of keywords, $Authors_{d_s} = \{a_1, \cdots, a_k\}$ is a set of authors, $Classifications_{d_s} = \{c_1, \cdots, c_k\}$ is a set of classifications in d_s, and $Journal_{d_s}$ is the journal in which the seed document d_s was published in. On the user interface of Sowiport these browsing options are represented by the descriptive metadata of the seed document (see Section 4.4.3).

Stratagem browsing in Sowiport operates on the basis of a filter query fq, which is defined as a tuple (q, ft) consisting of a query q and a field type ft. For a given seed document d_s, $q \in StratagemAct(d_s)$ and $ft \in \{keyword, classification, journal, author\}$.

Due to the heterogeneity of the documents in Sowiport, one can not ensure that every document contains each of the different sets given in $StratagemAct(d_s)$. A document might, for instance, only cover a set of $Keywords_{d_s}$. Hence, situations occur in which not all browsing options are available to a user.

Although the empirical results presented in Section 5 and Section 6 showed that citations and references are valuable search activities, they are not contextualised and thus, not part of the upcoming experiment. The reasons for excluding these from the implementation were two-fold: 1) citations and references are only contained in a comparably small amount of documents in Sowiport, 2) citations and references resolve to an associated record instead of a result list, and thus, a contextualisation is not reasonable in this experiment.

7.2 Non-contextual Baseline Ranking

In order to compare the performance of the two contextualisation approaches, the default implementation of stratagem browsing in Sowiport is employed as a baseline. The ranking of the baseline does not take previous interactions of the user into account and is thus, denoted as non-contextualised.

The default ranking of Sowiport during stratagem browsing is based on a filter query expansion which can be formalised as follows:

$$DR : Q_f \times D \rightarrow RL \quad (7.2)$$

Equation 7.2 denotes the default ranking function DR that takes an expanded filter query Q_f, searches in the set of documents D, and returns a ranked list of documents RL. In order to determine Q_f, a query expansion function is employed

which takes the initial filter query fq (selected by the user) as input and expands it on two levels: 1) it expands the initial query q and 2) it expands the field type ft.

- Query Expansion:
  On this level, the query q contained in fq is expanded with synonyms and different translations. The basis for this expansion is provided by Sowiports heterogeneity service (HTS see Section 4.4.1). The core principle of the HTS is to support cross-concordance searches that determine equivalence, hierarchy, and association relations between terms (Hienert et al., 2015a).[37]

- Fielded Expansion
  In contrast to the filter query expansion, the fielded expansion aims to broaden the field type (ft). Assuming a user has selected a keyword k_j from a seed document for browsing and hence, ft = *keyword*. The intuitive choice would now be a fielded search for all documents D where k_j in $Keywords_{Dn}$. Instead, the fielded expansion extends the filter type *keyword* to related fields. The actual scoring which determines the rank of the documents in the result set is then provided by boosting the results based on the contained field type. The order in which the boosting is performed is predefined by Sowiport following the assumption that certain fields are of greater relevance than others (e.g. *keywords* > *free_keywords*).

A simplified example query for the default ranking is displayed in Listing 7.1.

```
1 q => keyword:(("violence" OR "Gewalt")^400 OR
2 keyword_free:(("violence" OR "Gewalt"))^250
```

Listing 7.1: Example query for the default ranking.

In the example above, a user has selected the keyword violence for stratagem browsing. One can see that the baseline approach not only performs a Boolean query on the keyword level, but instead the query and the field type are expanded. First, q = *violence* is translated into the German word "Gewalt" (line 1), which is the result of the query expansion. The fielded expansion then extends ft = *keyword* to related fields, like in this case, the metadata field "*keyword_free*" (line 2), which is an alternative and less formal descriptor for keywords.

The ranking of the retrieved documents is based on $TF * IDF$, whereby the weighting of the fields is taken into account. Line one of the example query shows that the field type *keyword* is boosted by a higher factor ($boost = 400$) compared to the field type *keywords_free* ($boost = 250$) in line two.
The prioritisation and hence, the boosting factors are part of the Sowiport configuration. This configuration is not only used on the level of stratagem browsing but also for regular query searches and has been in productive use for several years.

[37] For a more comprehensive description of the HTS, one can refer to Hienert et al. (2015a); Mayr & Petras (2008).

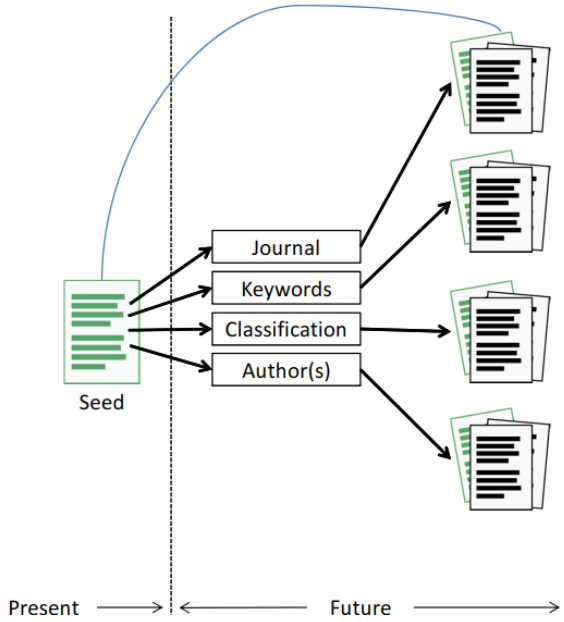


Figure 7.1: Similarity-based contextualisation based on one seed document.

Thus, the boosting factors were not adjusted for the upcoming experiment.

It was decided to utilise the described Sowiport default ranking as a baseline, although an out-of-the-box VuFind-Solr configuration, which performs a simple Boolean filter without query expansion, would have decreased the complexity and would have allowed a better reproducibility of the results. The main reason for the decision to use Sowiport's implementation is to compare the contextualisation approach to a realistic real-life DL ranking which also serves as a strong and established baseline.

7.3 Contextualisation based on Document Similarity

In this approach, a re-ranking of the result list is performed, which is based on the similarity of each document in the result set compared to the seed document.

The basic concept of the similarity-based approach is displayed in Figure 7.1. The figure describes a situation in which a user is inspecting a certain document, represented by the seed document D_s. This is depicted by the dimension "Present" in

Figure 7.1. Potential upcoming search activities are available in the form of stratagems which, when employed, lead to corresponding result lists. This is depicted by the dimension "Future" in Figure 7.1.
When employing the similarity-based contextualisation, the documents at the right-hand side are re-ranked according to their similarity to the seed document. Formally this approach is described as follows:

$$SR : Q_f \times D \times D_s \rightarrow RL_S \tag{7.3}$$

The ranking function SR is an extension to the default ranking DR in Section 7.2. The similarity-based ranking function (SR) uses the expanded filter query Q_f following Equation 7.2. Furthermore, SR employs the seed document (D_s) as an additional parameter to determine a ranked list of documents RL_S.

During the first step, the documents are ranked correspondingly to the default ranking. However, instead of only boosting documents based on field types (ft) that contain the query q, the ranking is extended by a similarity score of each document compared to the seed document. To determine the similarity score of each document compared to the seed document, the *MoreLikeThis (MLT)* query parser[38] built-in SOLR is utilised. Usually, this function is employed to provide related documents to a given seed document. Thus, the origins of this approach are borrowed from the field of recommender systems which are a part of Sowiport as well.

An example query for the re-ranking based on document similarity is provided in Listing 7.2.

```
1 q => {!mlt}DocID
2 keyword:((violence OR "Gewalt")^400 OR
3 keyword_free:((violence OR "Gewalt"))^250
4 qf=authors,keywords,journal,abstract
```

Listing 7.2: Example query for the re-ranking based on document similarity.

One can see that the query is based on the default ranking described in Section 7.2 (lines two and three), but additionally, the *MLT* query parser is supplemented in line one. The seed document is specified by the *DocID* parameter.

[38] http://archive.apache.org/dist/lucene/solr/ref-guide/apache-solr-ref-guide-5.3.pdf, last accessed April 15th, 2020.

In order to compute the similarity of each document to the seed document, the following list of attributes is used:

- Author names
- Keywords
- Journal Information
- Abstract (in different languages, if available)

Each of these attributes is specified in the *qf* parameter displayed in line four of Listing 7.2.

In most cases, the top-ranked document in the result list is equal to the seed document as it usually produces the highest similarity score. The similarity-based approach was part of a pretest phase in which Sowiport users were assigned either to the default baseline or to the similarity-based variant during stratagem browsing. During this phase, a high click-through rate on the previously encountered seed document contained in the result list was observed. It is assumed that the users confused this document with a newly discovered one and thus, were biased towards inspecting the same document again. To prevent a potential click bias towards the previously inspected document, the seed document was excluded from the result set ($d \neq d_s$).

7.4 Contextualisation based on Interaction Context

In this approach, the re-ranking is performed based on the user's interaction context, which is derived from a set of actions the user has performed prior to the encounter of the seed document. The basic concept of this approach is illustrated in Figure 7.2. One can see that an additional dimension (past) is introduced. The depicted example describes the interactions of a user seeking information on the topic "violence and sports". After entering the first query, "violence", the user inspects two documents and subsequently refines the query to "violence and sports". The seed document is selected from the corresponding result set. The knowledge about the user's previous interactions defines the interaction context, which is employed for the re-ranking of stratagem browsing content. Formally, this approach is described as follows:

$$CR : Q_f \times D \times U_c \rightarrow RL_C \tag{7.4}$$

where

$$U_c = \{Keywords(U), Classifications(U), Queries(U)\}.$$

The ranking function CR is again an extension to the default ranking DR. The function CR takes the filter query (Q_f), the user's interaction context (U_c), and the

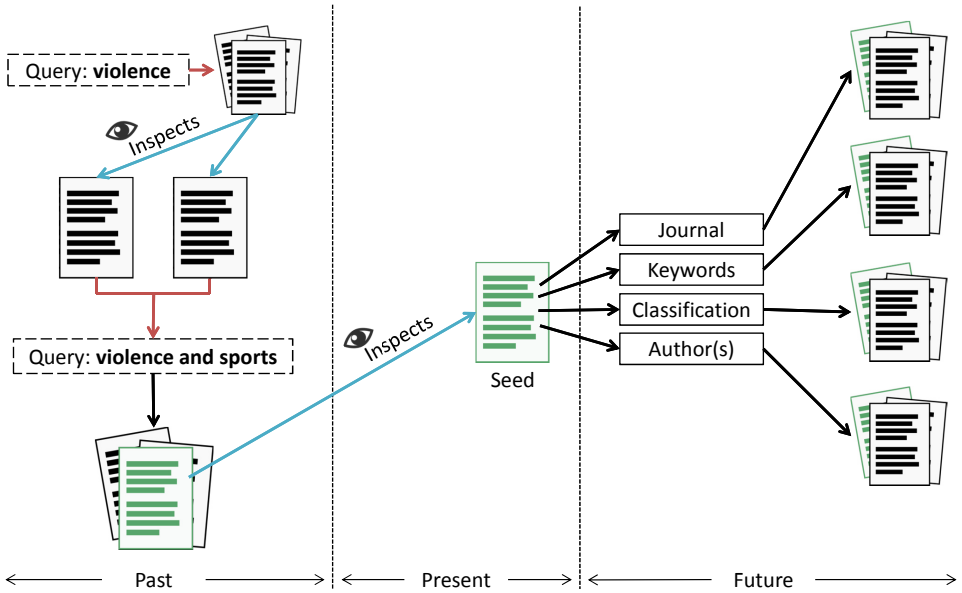


Figure 7.2: Contextualisation based on the interaction context.

set of documents D as input to produce a ranked list of documents RL_C.

Again, the first step is to filter all documents corresponding to the default ranking, which for a given filter query fq creates an expanded query representation Q_f. The re-ranking is then performed by boosting the filtered documents on the basis of the user's interaction context (U_c) which consists of the

- Keywords,
- Classifications,
- and Queries

that a user U has encountered/submitted.

The keyword and classification features are derived from two sources: 1) each keyword and classification is considered that was contained in the list of documents the user has seen, 2) each keyword and classification is considered that was contained in documents within a result set.
In the next step, both resulting sets of keywords and classifications are ordered according to the number of their occurrences. For example, if a certain keyword has appeared four times in the list of viewed documents and five times in documents within the result list, the number of occurrences for that particular keyword is nine.

In order to reduce noise, the number of keywords and classifications is reduced to the top three for each of the two features. This limitation is necessary because the

number of documents a user encounters could potentially be very high and thus, result in an excessive representation of numerous different keywords and classifications within the interaction context which is not beneficial.
In the final step, the frequencies of the top three keywords and classifications are normalised according to Equation 7.5.

$$\sum_{i=1}^{\mathrm{t}} K_{i_{\mathrm{norm}}} = \frac{K_{i_{\mathrm{freq}}}}{K_{\mathrm{max}}}, \qquad \sum_{i=1}^{\mathrm{t}} C_{i_{\mathrm{norm}}} = \frac{C_{i_{\mathrm{freq}}}}{C_{\mathrm{max}}} \tag{7.5}$$

First, for each feature (keywords and classification), the most frequent term K_{max} and C_{max} is determined. The normalised frequency of a feature $K_{i_{\mathrm{norm}}}$ and $C_{i_{\mathrm{norm}}}$ is then computed as the ratio between the frequency of a feature $K_{i_{\mathrm{freq}}}$ and the maximum frequency $K_{i_{\mathrm{max}}}$. The parameter t denotes the threshold ($t = 3$).
Regarding the queries, no threshold was applied as it is assumed that all queries are equally important to describe the user's information need and do not contain any noise. An example interaction context could be described as shown in Listing 7.3:

```
1 {"query":"violence sports","rank":1},
2 {"keyword":"Football","rank":1},
3 {"keyword":"Radicalism","rank":0.5},
4 {"keyword":"Ethnic Conflict","rank":0.5},
5 {"classification":"Political Sociology","rank":1},
6 {"classification":"Decision Making","rank":0.66},
7 {"classification":"Sociology","rank":0.66}}
```

Listing 7.3: Example interaction context.

In the example displayed in Listing 7.3, the user has submitted the query *violence sports*. The keywords and classifications (line 2 to line 7) are derived from those documents that were contained in the corresponding result set and from documents that a user has clicked (viewed). Furthermore, each keyword and classification is ranked according to its normalised frequency. It may happen that an overall frequency of 1 is determined for all features. According to Equation 7.5, this results in a rank of 1 for each keyword and classification. This is usually the case when a user enters Sowiport from a web search engine and navigates immediately to the detailed view of a document. In this case, each keyword and classification from the seed document is used for the interaction context.
The information covered in the interaction context can now be used to re-rank result lists coming from stratagem browsing. This is performed by boosting the score of documents that share attributes contained in the interaction context. The boosting of documents is performed in the following order: by the entered query terms, the keywords, and the classifications each multiplied by their normalised rank (see Listing 7.3).

An example query used for the re-ranking based on the interaction context is displayed in Listing 7.4. The interaction context in this example is derived from the example in Listing 7.3.

The first part of the query is based on the default ranking DR (lines two and three) which is extended by a boosting parameter *bq*. In line 5, it can be seen that the previously entered query term *violence sports* is matched against the title field, which is also the metadata field with the highest boosting value. In lines six to eight, documents are boosted that contain the most frequent keywords in the interaction context. It can be seen that each keyword is boosted with a decreasing factor which is depending on the keywords' frequency in the interaction context. In lines 9 to 11, those documents are boosted that contain the classification terms from the interaction context.

```
1  q =>
2  keyword:((violence OR "Gewalt")^400 OR
3  keyword_free:((violence OR "Gewalt"))^250
4  [bq]=
5  (title:violence sports^1700) OR
6  (keyword:Football^1200 OR
7  keyword:Radicalism^600 OR
8  keyword:Ethnic Conflict^600) OR
9  (classification:Political Sociology^800 OR
10 classification:Decision Making^530 OR
11 classification:Sociology^530 )
```

Listing 7.4: Example query for the re-ranking based on the interaction context.

7.5 Conclusion

The present section has introduced the implementation details behind two contextualised browsing variants together with a non-contextualised baseline that represents the default behaviour during stratagem browsing in Sowiport. Both described contextualisation variants were developed using well-established methods that are present in many modern Digital Libraries. This allows developers the incorporation of the described methods in their own DL with reasonable effort.
In the upcoming section, the effectiveness of these two approaches is evaluated and compared to the non-contextualised baseline of Sowiport.

8

Evaluation of Contextualised Browsing

The previous section has outlined the implementation details of the two contextualisation approaches and the non-contextual baseline. The present section is dedicated to the evaluation of these.
One of the major challenges in the evaluation of contextualisation approaches, in fact, of nearly any interactive system, is that large scale log data is usually hard to obtain. For this reason, the contextualisation approaches presented in the previous section were evaluated using the real-life Digital Library (DL) Sowiport, which was described in detail in Section 4.4. An A/B/C-testing was designed to compare the two contextual stratagem browsing variants against Sowiports' non-contextual baseline.

The study presented in this section is dedicated towards answering the following research question:

> **RQ:** Can the effectiveness of exploratory search on the level of browsing be improved by employing contextual ranking features in comparison to a non-contextual ranking feature?

The goal of this experiment is to obtain information about the effectiveness of contextualised browsing. The methods employed to determine the effectiveness are described in the following section.
The present section is an extended version of Carevic et al. (2018b) that has been published at the *Joint Conference on Digital Libraries*.

8.1 Evaluation Metrics

The experiment was running between 12th September 2017 and 20th December 2017. Following an A/B/C setting, each user was randomly assigned to one approach when visiting Sowiport. The assignment was carried out with an equal probability for each variant ($P = \frac{1}{3}$).

In order to measure the effectiveness of each variant, the following three measures were employed: 1) the click-through rate, 2) mean first relevant (MFR), and 3) the

usefulness as an implicit indicator for relevance. The click-through rate represents the number of record views per ranking variant. The last two measures are briefly introduced in the next two sections.

8.1.1 Mean First Relevant

The primary feature to determine the effectiveness of contextual stratagem browsing is to measure the rank of a clicked document in a result set. The results sets are, in this case, limited to those originating from stratagem browsing. Many Sowiport users only click on a single record in the result set, and thus, it was decided to use the rank position of the first clicked document, denoted as first relevant (FR), as a quality criterion. The most obvious metric to be used for this case would be the *mean reciprocal rank* (Kantor & Voorhees, 2000). But it has been argued by Fuhr (2018) that this metric is not on an interval scale, and thus the mean cannot be applied. Instead, for this experiment, the proposed alternative *mean first relevant (MFR)* is used. MFR takes the rank position of the first relevant document in a result set and computes the arithmetic mean for all result sets that were generated using stratagem browsing. This measure is proportional to the effort a user has to invest in finding the first relevant document in a result list. Correspondingly, an MFR value of *x* represents the *x-fold* effort in comparison to the ideal value of 1 (Fuhr, 2018). Hence, lower MFR values indicate better performance than higher values.

A detailed example of the MFR measure could be as follows. Suppose, there are two sessions U_1 and U_2, each representing an individual user *U*. The first user performs a journal run, and in the corresponding result set, he/she clicks on the document at rank 10 first and then scrolls back to the top of the result list and clicks at the documents at rank 3 and 7. Thus, the sequence of clicked documents in that particular result set is: $U_1 = \{\underline{10}, 3, 7\}$. The first relevant document clicked is underlined. The second user performs an author search and clicks on the documents at rank 2, 8, and 9, which results in the following sequence of clicked documents: $U_2 = \{\underline{2}, 8, 9\}$.
In order to compute the MFR, only the first clicked record (first relevant) during stratagem browsing is taken into account. Thus, the MFR of these two sessions is computed as the sum of the first clicked document for each stratagem browsing ($FR_{U_1} = 10$ and $FR_{U_2} = 2$), which is then divided by the total number of stratagem browsing interactions (N=2).

8.1.2 Mean First Relevant at 20

Given that the number of documents in a result set varies, it may happen that a result set only contains a small number of documents. Typical examples are the search for other documents of a certain author or the search for a highly specific keyword. In these cases, the MFR is usually rather low, which is a consequence

Table 8.1: List of implicit relevance signals indicating usefulness.

Short	Description
Add to favourites	Bookmark a single or multiple records to favourites
GoTo Google Scholar	Search a record in Google Scholar
GoTo Google Books	Search a record in Google Books
GoTo Fulltext	View the full text of the record
Check availability	Check availability in the local library
Export	Export a record in different citation styles or via e-mail

of having only a few documents in the result set and thus, might bias the results. Therefore, the MFR is additionally measured for all result sets that contain at least 20 documents. This is the default number of documents contained on the first result page in Sowiport. This is reported as the *MFR* $\geq$20.
As an additional restriction, a first relevant value is only considered if the clicked document was contained within the first two pages of a result set (*first relevant* $\leq$40), in order to exclude potential outliers that may distort the results when measuring the MFR. Situations in which stratagem browsing did not lead to any document click were not taking into account. Hence, the MFR and *MFR* $\geq$20 values only represent positive interactions with a result set. Adopting the measures by penalising stratagem browsing interactions that did not lead to a document click would also be a reasonable alternative.

8.1.3 Usefulness of Stratagem Browsing

Several session-related features, like the number of interactions, the number of document views, and the total duration of an individual session, were measured. However, session-related measures are primarily of descriptive nature and not suitable to provide conclusive insights into the benefits of each variant. By determining the session-related features mentioned above, it is not possible to tell whether a document was of relevance to the user or not. In fact, missing information about relevance is a prevalent problem in many studies that aim to evaluate a certain feature using transaction-log data.

In order to overcome this problem, a measure called *usefulness* was applied, which was previously described by Hienert & Mutschke (2016). The idea of this approach is to identify a set of interaction signals which are quantified, in order to approximate the success of a search session. For instance, if a user bookmarks a certain document one can consider that this document was of relevance to the user, which is contributing to the overall success of a session.
The ideas of the usefulness measure are similar to the principles of implicit relevance feedback. The main difference is that implicit relevance feedback is employed to estimate the users' search interests to improve future queries (see Section 3.2.2). Usefulness, on the other hand, is understood as a measure to determine the

effectiveness of a (interactive) feature. For a detailed analysis of usefulness, one can refer to Hienert & Mutschke (2016). The list of implicit relevance signals that were considered to measure usefulness is displayed in Table 8.1.

The usefulness was measured by quantifying the number of implicit relevance signals in the log file subsequently to a stratagem browsing action. This was performed on two levels:

- Local usefulness
  The local usefulness is the sum of implicit relevance signals on documents that were contained in the result sets **immediately** after stratagem browsing. For a visual example of implicit relevance feedback on a result set, see the right box in Figure 4.6 (b).

- Global usefulness
  The global usefulness represents the total number of implicit relevance signals contained in the entire session after the first employment of stratagem browsing.

The local usefulness is employed to determine the "immediate" implicit relevance of a document in a result set after stratagem browsing.
The global usefulness, on the other hand, measures the usefulness of stratagem browsing across the entire session. Consider the following example: a user clicks on the button "add to favourites" immediately after stratagem browsing. This action is considered to be a strong indicator for the usefulness of that particular document and thus, for the usefulness of the stratagem. This contributes to the local usefulness. The global usefulness, on the other hand, denotes the sum of implicit relevance signals across the entire session and is thus, an indicator for the success of the session in general.

It is important to notice that the interface was not modified for this study. Instead, only a re-ranking of the result lists from stratagem browsing is performed. Thus, users are still able to use features like facets to narrow down the search results or apply a filter to determine a subset of the result list based, for instance, on the year of publication. This is displayed in the left box of Figure 4.6 (b). The disadvantage of not modifying the user interface is that the user is not aware of the applied ranking. This downside is discussed in more detail in Section 9.3.

8.2 Results of the Transaction-Log Study

In total, 607,109 sessions were analysed during the study. Each session was studied anonymously, and only a session-id was used to identify and aggregate interaction data without referring to potential individual user data.

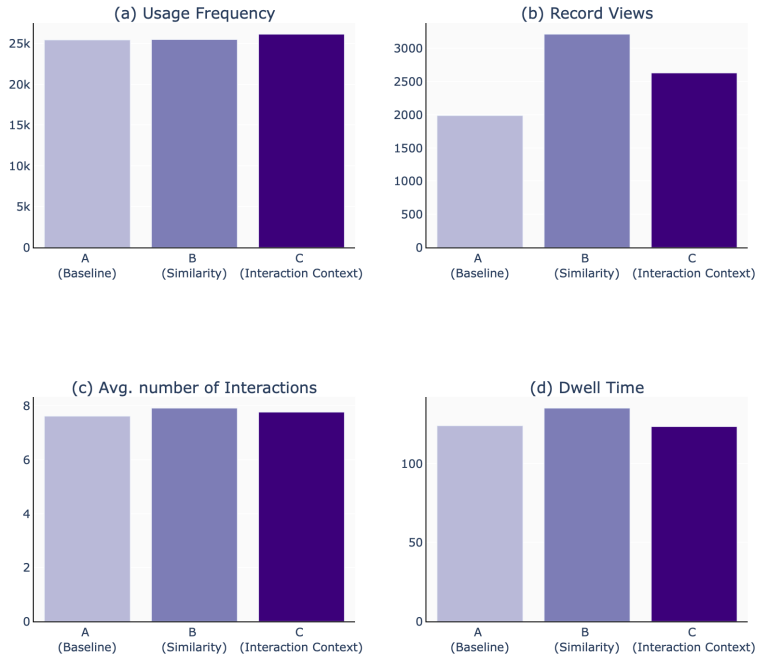


Figure 8.1: Descriptive statistics for the period of the study.

Descriptive statistics on the overall usage of stratagems can be found in Figure 8.1.[39]

During the experiment, all three approaches (A=Baseline, B=Similarity, and C=Interaction Context) were nearly equally distributed among the users. This is displayed in Figure 8.1 (a). Users conducted stratagem browsing 77,036 times. Figure 8.1 (d) illustrates the mean dwell time starting from the first usage of a stratagem until the end of the session. When looking at the dwell time, numerous outliers were identified. Therefore, all sessions that exceeded a dwell time of more than 20 minutes were excluded. On average, users continued their search for 2.1 minutes after the first stratagem usage. All three approaches show a similar dwell time. Users that were assigned approach A or C continued their search for 123 seconds while approach B was continued for 134 seconds. Figure 8.1 (c) displays the average number of interactions for all sessions containing stratagem browsing. It can be seen that the number of interactions does not differ substantially between the three approaches.

So far, the descriptive statistics show no substantial difference between the three variants. However, in Figure 8.1 (b), the number of document views resulting from stratagem browsing is displayed. This observation clearly showed a disparity between the variants. Both contextualised approaches have a considerably higher

[39] A tabular representation can be found in the Appendix A.5.

Table 8.2: MFR and MFR≥20 values for all three ranking variants.

Approach	MFR	SD	MFR≥20	SD
(A) Baseline	4.66 (N=1078)	6.45	6.47 (N=607)	7.74
(B) Similarity	3.10* (N=1999)	4.41	3.39* (N=1528)	4.81
(C) Interaction context	3.62* (N=1571)	4.74	4.30* (N=1097)	5.33

number of document views resulting from stratagem browsing in comparison to the non-contextualised baseline. Considering the baseline, only 1,985 documents were viewed while the contextualisation based on similarity received more than 3,200 document views. The contextualisation based on the interaction context also clearly outperforms the baseline with 2,627 document views. This tendency is also reflected in the click-through rates, which are a part of the upcoming section.

8.2.1 Mean First Relevant

To determine the effectiveness of the contextual ranking variants in comparison to the non-contextualised baseline, the mean first relevant metric (MFR), as described in Section 8.1, is utilised.

In Table 8.2, the MFR and MFR≥20 values for all sessions are displayed. The contextualisation based on similarity to the seed document performed best with an MFR value of 3.10. The contextualisation based on the interaction context received an MFR value of 3.62, while the non-contextualised ranking performed the worst with an MFR value of 4.66.
Besides the better results in terms of MFR, one can observe a higher click-through rate (denoted as N) for both contextualised variants. For the contextualisation based on similarity (B), nearly twice as many first relevant clicks (N=1999) could be observed compared to the baseline (A) which received only 1078 first relevant clicks. The contextualisation based on the interaction context (C) also clearly outperforms the baseline with 1,571 first relevant clicks.

Due to skewed data, a non-parametric Mann-Whitney U-Test was utilised, to test for significant differences in the results with a Bonferroni corrected $p^* = 0.016$. It was discovered that both contextualised ranking variants significantly outperform the non-contextualised baseline with A $>$ B, $p^* = 0.001, r = 0.13$ and A $>$ C, $p^* = 0.014, r = 0.04$. Furthermore, the contextualisation based on similarity significantly outperforms the variant using the interaction context with C $>$ B, $p^* = 0.001, r = 0.04$.

Users of Sowiport may encounter result set sizes of varying lengths with only a few records or up too many thousands of records. Thus, in addition, the MFR≥20 measure is applied to all result sets that contain at least 20 documents. These values are displayed in column four of Table 8.2.

Overall, the MFR≥20 values for all three approaches increased. The strongest impact can be observed for the non-contextualised approach, whose MFR increased from 4.66 to 6.47 (+1.81). The MFR values of the contextual variants increased moderately from 3.10 to 3.39 (+0.29) for the similarity variant and from 3.62 to 4.30 (+0.68) for the variant using the interaction context. Thus, the general tendency of a better performance for the contextualised variants remains the same. The contextualisation based on document similarity still performs best, and both contextualisation approaches still outperform the baseline significantly. The gap between the non-contextual baseline and the contextual approaches, however, increased substantially.

8.2.2 Mean First Relevant by History Size

The efficiency of the interaction context variant (C) strongly depends on how rich the interaction of the user was prior to the employment of stratagem browsing. Consequently, a contextualisation based on the interaction context suffers from a cold start problem that applies to users who just entered Sowiport with only a few or even no interaction prior to stratagem browsing.

Therefore, the MFR for different history sizes was determined to evaluate whether the employment of stratagem browsing later on in a session has any influence on the performance of the three variants. History sizes in this analysis are understood as the number of interactions prior to stratagem browsing. Three sets of history sizes were defined: MFR H $\in [2, 5]$ accounts for stratagem browsing with two to five interactions prior stratagem browsing. Likewise, an MFR H $\in [6, 10]$ represents stratagem browsing with six to ten interactions prior stratagem browsing, and an MFR H $\in [11, \infty]$ accounts for 11 and more interactions. In Figure 8.2, the MFR values for the three different history sizes are illustrated.[40]

The results in Figure 8.2 show that with a growing history size, the MFR values of all three approaches increased. Looking at the non-contextualised approach, the MFR at the early stages of the session is 4.52, while increasing to 5.06 after a history size > 5. For sessions with a history size H $\in [11, \infty]$, the MFR for all approaches further increases (A=5.49, B=3.59, C=3.80).
The differences between the MFR values are most notably by looking at the actual increase between the different history sizes. For a history size of H $\in [11, \infty]$, the MFR values of the baseline increased by 0.97 compared to a history size of H $\in [2, 5]$, while the contextualisation based on document similarity increased by 0.55.

An interesting effect of this analysis can be found for the contextualisation using

[40] A tabular representation can be found in the Appendix A.6.

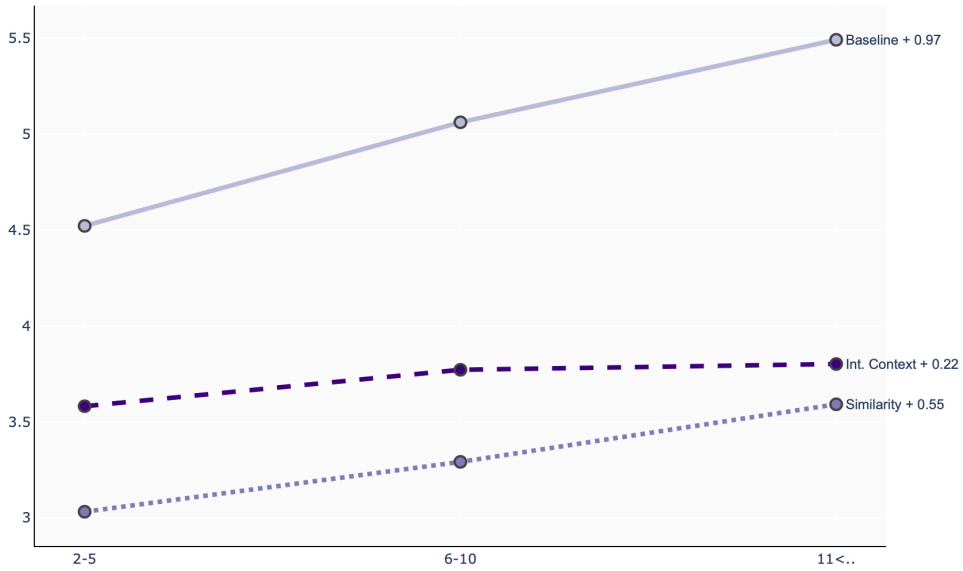


Figure 8.2: Mean first relevant for history sizes of H $\in [2, 5]$, H $\in [6, 10]$ and H $\in [11, \infty]$.

the interaction context. The variant employing the interaction context was least affected by the larger history sizes and increased only by 0.22. In general, one can observe that with an increased history size, the performance gap between the similarity variant and the interaction context is reduced notably. While this is no evidence for an overcoming of the cold start problem, it certainly has an impact on the effectiveness of the approach and serves as an indicator for an effect of the cold start problem.
Unfortunately, the number of sessions with a history size of H $\in [11, \infty]$ is comparably low, with only 487 sessions. Having a larger sample size would improve the reliability of these observations and might even lead to a better overall performance of the interaction context approach in comparison to the variant using document similarity.

8.2.3 Usefulness of the Contextualisation

To determine the usefulness of the contextualisation in comparison to the baseline, implicit relevance signals, which can be found in Table 8.1, were employed. In particular, the usefulness is determined by identifying the number of implicit relevance signals in the log file after the first usage of a stratagem on two levels: a) the local usefulness b) the global usefulness (see Section 8.1). To exclude outliers from this experiment, only those sessions were considered that contain a total of ≤ 10 implicit relevance signals.

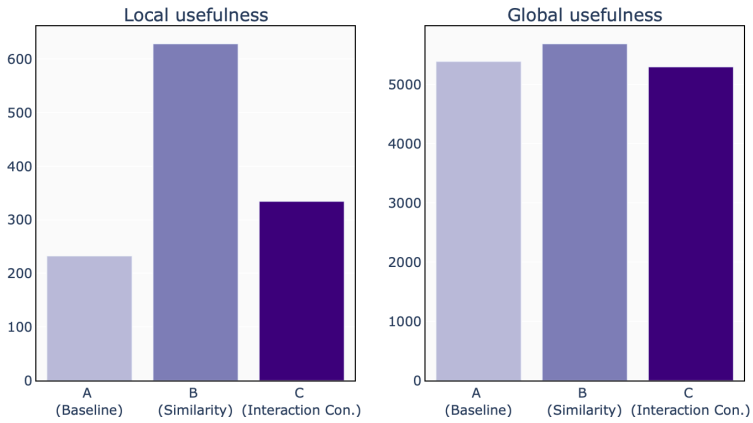


Figure 8.3: The local and global usefulness of the three ranking variants.

Figure 8.3 illustrates the results of this experiment.[41] Regarding the number of local usefulness signals which are displayed on the left side of the figure, one can observe that the contextual ranking variants again outperform the non-contextualised baseline. The similarity approached achieved the highest local usefulness (B=628), while the interaction context achieved C=334 relevance signals. Again, the non-contextualised baseline performed worst (A=232).
Regarding the global usefulness, displayed on the right side of the figure, the results show only marginal differences between the contextual approaches and the non-contextual baseline. The similarity approach again performs best with 5,684 implicit relevance signals in total after the first usage of a stratagem search. However, the results for the global usefulness show only marginal differences between the contextual approaches and the non-contextual baseline.

8.3 The Effects of Heterogeneous Data Sources

Sowiport offers the advantage that it allows an evaluation of contextualised browsing with real users. A downside of Sowiport, however, comes from its heterogeneous data sources. Having different data providers bears the risk of adding documents that had already been indexed by other data providers and thus, result in highly similar or even identical entries in the index. The degree of similarity, however, depends on the granularity of the descriptive metadata as discussed later on in this section.
Especially during contextualisation, this could have an impact on the ranking of documents. A seed document that is provided by more than one data source would naturally lead to top-ranked documents that represent alternative instances of the seed document when a contextualised re-ranking is performed. Despite that, re-

[41] A tabular representation can be found in the Appendix A.7.

moving documents from the results that exceed a degree of similarity is controversial and was not performed during the study. The main reason for this decision is that a record is rarely identical. In most cases, documents contain varying sets of descriptive metadata, which could be beneficial for users, such as the link to the full-text of a document. Applying a similarity threshold in order to remove documents from the result is a critical issue for users of a DL. Additionally, an identification of those document views was not possible during this study for various reasons which will be discussed in the following.

Different types of descriptive metadata

First and foremost, a duplicate entry is very difficult to detect in Sowiport. A record might intellectually be recognised as a duplicate, but from an algorithmic perspective, the solution is non-trivial. This difficulty is due to differing descriptive metadata between the records which are often of minor differences.

Avoiding false positives in duplicate detection

Records might be highly similar in terms of title and descriptive metadata, but a high degree of similarity is not always an indicator for a duplicate record. This is shown in the following example of two documents that are highly similar. The seed document was titled *Grundformen der Kartenarbeit (I)* and the first relevant record after stratagem browsing was titled *Grundformen der Kartenarbeit (II)*. These two documents share a high degree of similarity in terms of title and descriptive metadata, but still represent two distinct documents that are strongly related. After encountering the first relevant document (Grundformen der Kartenarbeit (II)), the user looked up this record in Google Scholar, which is an indicator for the local usefulness. This would not be possible if a document similarity threshold would have been applied.

Table 8.3: The connection between the seed document and the first relevant document.

Seed Document Title	First Relevant Title
Eingangsdiagnostik im Mathematikunterricht	Mathematisches Vorwissen im Anfangsunterricht
Das Gegenstromprinzip. Teil 1..	Das Gegenstromprinzip. Teil 3..
Politikvermittlung und Wahlen in den Medien	Politikvermittlung und Wahlen in den Medien: ARD-Forschungsdienst

Three examples extracted from the analysed logs are displayed in Table 8.3. The table shows the seed document (column one) along with the first relevant document retrieved from stratagem browsing (column two). All three examples show real

user behaviour, extracted from the transaction log and retrieved via the interaction-context variant. The first row contains an example of two records that are topically related. The first relevant document was contained on rank four in the result list. Employing the baseline for the same seed document would result in the first relevant document on a rank position of > 20 and thus, would not appear on the first result page of Sowiport. The second row shows an example of two documents that represent different parts of a series. These two documents can, therefore, be considered as highly related and provide a good example for the contextualisation of stratagem browsing. By applying the baseline ranking, the first relevant document would exceed the first result page of Sowiport. The third row contains an example of two records that share an identical title except that the first relevant document is extended by a sub-title. From an intellectual point of view, they are identical. However, only the first relevant document contains a link to the full-text of the record which was not available in the seed.

8.4 Conclusion

This section presented the evaluation of two contextualisation approaches on the level of stratagem browsing in a DL: a) a contextualisation based on document similarity and b) a contextualisation based on the users' interaction context. The results show that a contextualisation of browsing in DLs significantly outperforms the baseline in terms of MFR. Furthermore, an increase in the click-through rate for both contextualisation approaches was observed in comparison to the baseline. An analysis of the local usefulness, again, showed a better performance for the two contextualised browsing variants, while an effect on the overall session in terms of global usefulness could not be observed. Other session related factors, like dwell time or the number of interactions per session, did not differ between the approaches.

It has already been hypothesised by Ryen W. White that the incorporation of the interaction context can lead to a new generation of search systems that can be created, designed, and developed, to increase the performance of context-insensitive search systems (White, 2016). Even though White originally referred to contextualisation mainly on the query level, the results of this experiment suggest that this can also be transferred to the level of stratagem browsing.

Part IV

Discussion and Conclusion

9

Discussion

The following section is dedicated to the discussion of both core chapters: 1) the empirical studies covering the online survey and the user study as well as 2) the implementation and evaluation of contextualised browsing. The discussion is organised as follows: first, the results of each study are briefly recapped and interpreted in the light of the corresponding research questions. The key findings and contributions from each study are pointed out and put into context. This is followed by an analysis of the strengths and weaknesses of each study concerning study design, generalisability, and potential flaws that may bias the interpretation.

Implications for the design of Digital Libraries (DLs) are stressed with respect to the key findings of all three studies in Section 9.4. This chapter is finalised by a consideration of open questions that could not be addressed in this thesis. These open questions are discussed as potential areas of future work (see Section 9.5).

9.1 Survey on Search Activities at the Stratagem Level

The online survey (see Section 5) aimed to answer the following research question:

RQ 1 What kind of stratagem browsing variants do users employ during information seeking, and how do the respondents assess their usefulness?

With regard to the research question, the survey showed that stratagem browsing is a commonly used search activity across a wide range of respondents with an academic background. The results presented in Section 5.2.2 point out that journals and conference proceedings were considered as very useful sources by the majority of the respondents (mode=5, median=5, N=156). The employment of a stratagem, however, appears to be dependent on the user's state of search. The journal run, for instance, is less often utilised when the respondent is looking for related material to a given relevant document (often to very often by 35.2%) compared to a journal run that is performed without viewing a preceding document (often to very often by 54.9%).

Similar to the journal run, citations and references are commonly used. 65.7% of the respondents used citations, and 82.1% used references often to very often (see Section 5.2.3). However, the higher usage of references over citations does not necessarily reflect higher usefulness of these.

Using open-ended questions, several respondents stated that DLs rarely provide access to citations. This is a possible explanation for why respondents utilised citations searches less often than references. DLs that provide better access to citation information could balance this difference.

In Section 5.2.4, the respondents were asked to choose a stratagem as a follow-up search activity based on a given scenario description. The respondents preferred to follow references (mean=2.38), browse the list of documents citing the given paper (mean=2.79), and keywords (mean=2.82). The journal run, on the other hand, was ranked comparably low (mean=4.10).

Overall, the results concerning the usage of citations, references, and journals reflect the findings from Athukorala et al. (2013). The authors, amongst others, investigated the usage of citations, references, and authoritative forums (e.g. journals/conference proceedings) for different types of purpose, such as "Stay up to date with research" and "Explore unfamiliar research areas". The results showed a predominant usage of citations and references. Authoritative forums, on the other hand, were used, in order to stay up to date with research. In this case, the respondents were manually browsing through journals/conference proceedings of familiar and renowned sources. Additionally, Athukorala et al. investigated the preferred sorting options (citation count, forum rank, year) for search results. Their findings indicate that a sorting based on the rank of a forum is highly relevant for the ranking of search results. A ranking option based on contextualisation was not present in their study.

Key Findings

With respect to the aforementioned results, one can derive two key findings. First of all, the online survey demonstrates that stratagems are legitimate search activities. The findings yield to contemporary understanding of stratagems that indicate the legitimacy of Bates' conceptualisation, even though these were developed prior to the current advances of modern DLs purely from observing the information-seeking behaviour of experienced users of a physical library.

The second key finding from the online survey was the identification of an inadequateness in the operationalisation of stratagems, that became clearly visible from the open-ended questions. When asked for reasons to reject a certain stratagem, some of the respondents indicated that the underlying content was often too unspecific to be useful.
A similar argumentation is discussed in the well-known berry-picking model by Bates (see Section 2.5). She argues that a journal with a broad subject is unlikely to fulfil a user's information need, but more useful to monitor a certain area of research. In contrast, very specific journals are likely to meet a researchers interest (Bates, 1989). Considering her argumentation and the observations from the online

survey, one can identify a problematic situation of so-called *topical broadness* that results from certain stratagems.
From the motivational point of view, this topical broadness strengthens the assumption that stratagems, such as the journal run, need more advanced system support, to be truly applicable across a broader range of search tasks (e.g. looking for related material to an already retrieved relevant document). Modern DLs do in fact support stratagem browsing, but the implementation of these is, as of today, insufficient and therefore leads to a *contextual gap* between the users' information need and the ranking provided. A contextualisation of search activities, as proposed in this thesis, could help to increase the usefulness of these types of search activities by reorganising the content according to the users' preferences. This might eventually help to bridge the contextual gap.

These considerations are supported by the observations made in the task of organising the content of a journal (see Section 5.2.5). In this task, the respondents were asked to organise the content of a journal on the basis of six ranking options: four generic options, such as authors or title, and two contextualisation features (similarity and interaction context). The respondents considered the contextualised ranking options (median=2 for both contextual ranking options) as substantially more valuable than the four non-contextualised features (median=4 for the best ranked non-contextualised feature).

Strengths and Limitations

In total, 204 respondents took part in the online survey, of which 129 completed it. The respondents came from 12 different scientific disciplines, such as computer science and social sciences. This provides a solid foundation for a first overview of stratagem browsing from a diverse field of studies.

While the results of the online survey provided a general overview of the usage of stratagem browsing across a broad range of disciplines and academic degree, it is challenging to generalise the results. One factor that limits the generalisability is caused by a self-selection bias in web surveys, as discussed in Bethlehem (2010). Furthermore, an online survey can only help to determine the opinion of the respondents' based on their past experiences and preferences. It might be that the usefulness of a certain browsing strategy is assessed differently in a real usage scenario in comparison to an online survey. However, with respect to the present thesis, the results serve as an indicator of the usefulness of stratagem browsing and provoke further investigation of the topic.

The task of organising journal articles showed a need for contextual browsing. The results indicate that the ranking of documents during a journal run should stronger relate to the users' search activities (e.g. entered query term or the inspected doc-

ument). However, the scenario that was used to assess the users' opinion on organising journal content was composed using a relevant document as a starting point. It would be interesting to see whether the outcome of the task significantly changes when using a negative scenario, in which the respondents start from a non-relevant article. A negative scenario could possibly lead to a lower performance of context-based ranking options. A different survey design in which the scenario is alternated among the respondents with respect to a positive (relevant) and negative (non-relevant) scenario may have been more suitable to investigate contextualised ranking options.

With respect to the survey design, one could observe from the feedback of some of the respondents that the time for completing the survey was underestimated. During the pretest, a duration of five minutes was estimated, which was lower than the actual time to complete the survey (roughly 15 minutes). This might have resulted in certain respondents not completing the entire survey.

9.2 User Study on Stratagem Browsing in Digital Libraries

Given the discussed limitations of the online survey, the first research question is only suitable to provide an initial understanding of the usefulness of stratagem browsing. Gaining more in-depth insights into these types of searches requires a qualitative perspective. In order to extend the knowledge about stratagem browsing, a user study was designed in which the search activities of participants were observed during a given search task.

The user study (see Section 6) was dedicated towards answering the following research question:

RQ 2 What are the most frequently applied stratagem browsing variants in a state-of-the-art DL, and how is their usage in comparison to other search strategies like, for example, query searches? How is the perceived relevance of stratagem browsing opportunities?

In total, 137 search activities were performed on the stratagem level, of which keywords (50), references (27), citations (26), and author information (25) were most frequently used. A rather low usage was found for the journal run (4) and classifications (5). Queries and the usage of recommendations were observed 111 times.

Starting from the seed document, the participants most frequently utilised citations, keywords, and references as initial search activities. With an increasing number of interactions, the participants used queries and recommendations more frequently, while the usage of citations and references decreased. This may, however, be due to missing citation and reference information in the documents discovered by the participants. Several participants mentioned that the continuous usage of a stratagem is influenced by the success of their previous search. One of the participants

explicitly stated in the post-questionnaire that he/she avoids the mixing of search strategies.
When asked about the usefulness of different stratagem browsing variants, the participants considered references, classifications, and author information the most useful ones for the given task (see Section 6.2.1.2).

The gaze data showed a strong fixation along the descriptive metadata of a record, most intensely in the area of authors and publishing information of the seed document. Furthermore, it was shown that the distribution is mostly focused on the first entries of the metadata, e.g. the first two keywords contained in the seed document. Other areas with high fixations were: classifications, citations, references, access to full-text, and recommendations. Surprisingly, the distribution along the abstract of the seed document was rather low.
The gaze data, furthermore, showed notable differences between the two groups of participants for the seed document. Postdoctoral researchers spent more time inspecting the journal information of the seed document. The group of students spent considerably less time on this AOI (m=0.63 seconds) compared to the senior researchers (m=2.85 seconds). Likewise, the number of fixations on the journal was lower for the students (m=2.55 compared to m=8.33). Both groups of participants spent the most time on inspecting the list of references.
The mean dwell time spent on inspecting the list of references was 69 seconds for postdoctoral researchers and 38 seconds for the group of students. Overall, postdoctoral researchers spent more time on nearly all metadata information of the seed document. The only exception were citations, which were more intensively inspected by the group of students. It is assumed that certain descriptive metadata information is more valuable to experienced researchers and thus, inspected more intensively. A similar argumentation can be found in Mayr (2016).

Key Findings

One of the most important observations from this study is that stratagem browsing was a frequently applied search strategy across all participants. From a quantifying point of view, stratagem browsing was more often utilised in comparison to queries and the usage of recommendations (see Section 6.2.1). This is a rather surprising result since a lot of today's research focuses on query searches and recommender systems. The attention on stratagem browsing is comparably low. Additionally, it was observed that stratagem browsing was most frequently utilised as an initial search activity (see Section 6.2.1.3). With an increased number of interactions, the participants more frequently employed query searches or clicked on recommended articles. Following these observations, it is assumable that stratagems are particularly important during the early stages of the information-seeking process, in Ellis model described as the *starting* phase (see Section 2.2.1).[42]

[42]The reference to "starting" is particularly conclusive in this case because Ellis' model stressed the employment of an already retrieved relevant document as a starting point during this stage.

A unique feature of the user study is the analysis of user feedback concerning stratagem browsing provided in the post-questionnaire. By employing item-scale questions and open feedback, the user study provided valuable insights into the reasons why participants decided to use a stratagem or not. Various answers were, again, directed towards imprecision and topical broadness. Reasons for the little usage of a journal run, for example, were given in the post-questionnaire. Several participants responded that they did not perform a journal run, because the particular journal was too general for the task. Another reason to reject a certain stratagem was, for instance, that the employment would have cost them too much time. Hence, the reasons for rejecting a certain stratagem browsing variant are in line with observations from the online survey, namely that certain stratagems provide content that is too broad to be inspected in a reasonable time, in order to find related material.
This, along with the reassessment of organising journal articles (see Section 6.2.4), once more supports the hypothesis of contextualised browsing.

Another key finding that can be derived from the user study is provided by the analysis of gaze data. The gaze data indicate that the decision on utilising a stratagem strongly depends on the underlying content and its opportunities for the information seeker. The discrepancy between the actual employment of a stratagem and the perceived relevance, in terms of the number of fixations and the dwell time, strengthens this predication (see Section 6.2.2.1). The comparison of the actual usage of search activities and gaze data in the seed document showed that certain areas of interest (AOI) along the descriptive metadata were assessed intensively by the participants. It is, thus, assumed that certain search activities were in fact of relevance despite eventually not being employed. Nevertheless, it is important to note that one should not confuse perceived relevance with the relevance of a search activity in general. The results of the eye-tracker, as interpreted in this study, just provide an approximation of relevance which is not conclusive. However, a more profound and generalisable conclusion of this approximation would require a controlled experiment on its own. Overall, it is also possible that the descriptive metadata are the reason for the intense fixation counts.

Strengths and Limitations

Although it had been pretested, the design of the user study showed some limitations. The qualitative feedback of the participants indicates that 10 minutes were not enough time to complete the task given in the scenario. Apparently, more time or no time constraints at all might have been more appropriate for the given task.

Furthermore, it turned out that some participants were not familiar with Sowiport and the opportunities it offers, for which reason they lost some time during the task. A tutorial introducing Sowiport, e.g. a short demonstration or explanation

followed by a few minutes for getting to know the system before the scenario was handed out, might have increased the quality of the results. However, the decision not to introduce Sowiport made sure that the participants were not biased when focusing on certain areas of the interface.

One could also assume that the seed document for the task was not adequately chosen because of its generic scope and therefore, influenced the users' behaviour. The journal in which the seed document was published covered a very broad topic (*Cologne Journal of Sociology and Social Psychology*). This broad topic might have fueled the identification of topical broadness. On the other hand, a different seed document with a more specialised journal might have influenced the participants' behaviour in the opposite way.

A potential flaw in the design of the user study is that an additional task is missing which might have revealed effects that are caused by the task provided. Having an additional task would have increased the reliability of the results and allowed a comparison between the two tasks. Despite this, the results are still suitable to serve as an empirical foundation for the validity and the frequent employment of stratagem browsing during an actual search task, which was the primary focus of this user study.

For this study, Sowiport was utilised, which makes the present study reasonably close to a real search task and search behaviour. Also, the participants were allowed to navigate Sowiport without any restriction. The goal of such an open experiment design is to reflect information-seeking behaviour as realistic a possible.

The conclusions drawn in this section are based on a behavioural observation of 32 participants. While this is a reasonable number of participants for a controlled observational study, it is difficult to generalise the results presented in this section. A different task with a different sample of participants might lead to a different outcome. Nevertheless, the results do contribute to a better understanding of stratagem browsing and furthermore, provoke a need for contextualised browsing.

9.3 Evaluation of Stratagem Browsing

The first two research questions followed the principles of Interactive Information Retrieval. They were designed to obtain a better understanding of stratagem browsing. One of the main outcomes of these two studies was the identification of topical broadness. In order to address this problem, two contextualised browsing approaches were implemented (see Section 7).

The following study is dedicated to the evaluation of the effectiveness of contex-

tualised stratagem browsing. Its underlying research question, represents the core investigation of the present thesis.

RQ 3 Can the effectiveness of exploratory search on the level of browsing be improved by employing contextual ranking features in comparison to a non-contextual ranking feature?

By measuring the position of the first relevant document in the result set (MFR), it was revealed that both contextualisation approaches led to significantly improved results in comparison to the baseline. The similarity-based contextualisation (variant B) performed best (MFR=3.10), and the contextualisation using the interaction context (variant C) reached an MFR of 3.62. The baseline (variant A) performed worst, achieving an MFR of 4.66.
Besides the better performance with respect to MFR values, one can see a considerably higher click-through rate of the contextualised variants (B=1999, C=1571) compared to the baseline (A=1078).

When limiting the MFR measure to all result sets that contain at least 20 documents, one can observe a negative effect on the performance of all three approaches. The MFR score of the non-contextualised baseline, for instance, increased from an MFR of 4.66 to 6.47 (+1.81). The contextualised approaches were less affected and increased from an MFR of 3.10 to 3.39 (+0.29) for the similarity variant and from an MFR of 3.62 to 4.30 (+0.68) for the interaction context variant.

A similar observation was made, when the MFR values were analysed with respect to different history sizes. All three variants performed worst with increasing history size. Comparing the MFR values for a history size of $H \in [2.5]$ with larger history sizes (H $\in [11, \infty]$) showed that the non-contextualised approach had the largest increase in terms of MFR by 0.97. The MFR of the contextualisation based on similarity increased moderately by 0.55. The contextualisation based on the interaction context was least affected by larger history sizes and increased only by 0.22. With increased history sizes, the gap between the similarity approach and the interaction context variant is lowered (B=3.59, C=3.80). Following these trends, it is plausible to assume that the interaction context would outperform the similarity context for sessions with history sizes above a certain threshold.

It is assumed that the performance of the interaction context depends on the complexity of the search task. If a session of a user contains a larger set of interactions, then this could be understood as an indicator for the complexity of the user's search task. Likewise, it is assumed that sessions with a higher number of interactions, also have greater demands on the quality of a result list. This could be an explanation of why all three approaches performed worse when the number of interactions increased.

The comparably low increase in terms of MFR for the interaction context could also be explained by a cold start problem. A large fraction of users visit Sowiport via a web search engine, which indexes the detail view of a document (see Section 4.4.2). Thus, users are enabled to employ stratagem browsing at an early stage in a session. However, in this case, the interaction context contains only little information about the user's information need, and therefore, only few information can be used for contextualisation. The more interactions a user performs, the more detailed the interaction context of the user can be modelled. This might be one explanation for the good performance of the interaction context for larger history sizes.

Unfortunately, a large fraction of search sessions analysed throughout the experiment comprises only 1 to 5 interactions, and thus, it is difficult to generalise the results of the interaction context approach (C). One can only assume that for highly interactive search sessions, the results of the re-ranking will differ and lead to better performance. Having a larger sample size would improve the reliability of these observations. The low number of interactions in Sowiport is, however, not a unique phenomenon but rather a common pattern that occurs in other DL systems, too. This has also been observed in Jones et al. (2000).

The analysis of the MFR values provided valuable insights about the users' click behaviour when interacting with result lists. On the opposite, the better performance in terms of MFR and the higher click-through rate does not include information about the relevance of the clicked documents but rather about a topical relatedness to the user's search interests. Thus, similarly to Joachims et al. (2017), one should not regard the increased click-through rate as relevance on an absolute scale. Certain aspects, such as trust bias in ranked result lists, also play a vital role, but could not be uncovered by the MFR analysis.

To overcome the absence of relevance information and to get a deeper insight into the relevance of the clicked documents, additionally, the usefulness metric was employed. The intention behind this metric is to take implicit relevance signals into account that serve as an approximation of relevance.
This analysis was carried out on two levels: 1) the local usefulness of a search result after a stratagem run representing the immediate relevance and 2) the global usefulness which measures implicit relevance signals for the entire session. The results of the local usefulness are in line with the results of the MFR. Both contextualised approaches outperform the baseline. The contextualisation based on document similarity gathered more than twice as many immediate relevance signals than the baseline. The contextualisation based on the interaction context was also considerably higher (A=232, B=628, C=334).
This supports the hypothesis that the contextualised approaches are more effective in ranking documents that are of relevance to the user's search interests. By utilising the global usefulness, the number of implicit relevance signals after the first

stratagem usage for the entire session was measured. For the global usefulness, only minor differences in the overall success of the sessions could be observed. This may be an indicator for a rather equal performance of the three approaches in terms of session-wide effectiveness.

Similarly to the global usefulness, other session-related measures, like the dwell time or the number of interactions, did not differ substantially between the approaches.

Key Findings

From this experiment, one can derive three key findings:

The most important observation is that contextualised ranking clearly outperforms the non-contextual baseline. Both contextual ranking options perform significantly better in terms of MFR compared to the non-contextual baseline. In addition, by increasing the users' effort in assessing a result list that contains at least 20 documents, one can identify that the MFR scores for all three variants increase. The increased MFR scores for the larger result sets are not surprising and could be expected. But the effect on the baseline is substantial in comparison to the contextual approaches. The MFR values for the baseline increased by 1.81 while the MFR values of the contextual variants increased moderately (0.29 for the similarity variant and 0.68 for the variant using the interaction context).

The second key finding addresses the click-behaviour observed in the experiment. While the MFR values reflect the rank of the first clicked record in a result list, the click-through rate provides information regarding the total number of MFR clicks. From these results, one can see a considerably higher click-through rate of the contextualised variants compared to the baseline.

The third observation stems from the usefulness analysis. In particular, the results of the local usefulness stress that the immediate relevance signals are notably higher for the contextualised variants. This indicates that the quality of the result sets determined by the contextual variants is considerably higher.

All three findings described here are strong indicators for the effectiveness of the contextualised approaches in comparison to the non-contextualised baseline. Certain aspects that lie in the nature of a transaction log study might, however, have an impact on the results. These are discussed in the following section.

Strengths and Limitations

One downside of this experiment stems from the nature of transaction log studies. The results show a significantly better performance in terms of MFR on the level

of contextual ranking features. However, an interpretation of the results in light of quality and relevance is not possible solely by using the MFR measure. To gain a more qualitative view on the usage of contextual ranking features, it is necessary to conduct a user study that has the potential to provide insights into the relevance of documents and the users' particular reasons for regarding or disregarding certain documents.

Considering usability aspects and transparency, one can identify another downside. The users are not aware of the contextual ranking features. In a pretest, the users were provided with an information box, which aimed at providing transparency about the ranking mechanism. However, after conducting the pretest, it was decided to remove all information about the ranking features on the user interface as this led to confusion and distorted results.

In order to measure the overall success of each session in terms of usefulness, implicit relevance ratings (see Section 8.2.3) were employed. Although the implicit relevance ratings had previously been applied to measure the usefulness of Sowiports *Search Term Recommender (STR)* (Hienert & Mutschke, 2016), they are not evaluated entirely. Thus, the results for this observation are influenced by a subjective list of relevance ratings that were predefined in our research group at GESIS. Nevertheless, certain features are undoubtedly reasonable, such as adding a document to the users' personal favourites by clicking on "add to favourites", which indicates the usefulness of that particular document.

A further limitation of the transaction log study is the heterogeneity of data sources in Sowiport. Having multiple data sources incorporated in Sowiport leads to a situation in which identical variants of one document may exist. This is, of course, an issue for contextualised ranking features that re-rank documents based on content similarity. However, it was not practical to remove documents that exceed a given similarity threshold for numerous reasons that had already been discussed in Section 8.3. In order to gain knowledge about the degree of document similarity during contextualisation, a random sampling of 20 record views per contextualised stratagem browsing variant was selected. The results of this random sampling showed that users had a rather strong tendency towards clicking on similar records. The strongest bias was observed for the contextualisation operating on document similarity. Roughly 35% of the samples contained indications for a click on identical records. This is, of course, not too surprising, considering the employment of a similarity-based ranking on potentially heterogeneous data sources. For the contextualisation based on the interaction context, this number is lowered to 22.5%. Additionally, the sampling was repeated for longer sessions as it is assumed that users who have a more complex information need are more demanding when interacting with a result list. In fact, only 15% of the sample contained indications for identical records for the similarity variant and only 10% for the variant using the interaction context.

Given that the experiment was placed in a real-life DL with heterogeneous data sources such downsides are inevitable. Even outside of a contextual ranking, the behaviour of Sowiport in terms of identical records remains comparable. Hence, similar records would still appear close to each other in a result list of a non-stratagem search situation, such as ad-hoc querying.

9.4 Implications for the Design of Digital Libraries

The results of the present thesis have several implications for the design of DLs that are discussed with respect to the two core chapters: empirical studies and the transaction log study.

Implications resulting from the empirical studies

Considering the results of the empirical studies, two aspects for the improvement of DLs become apparent:

- Linking content: The frequent usage of stratagem browsing suggests that the participants consider structured and interlinked information as potentially useful when exploring a DL. These types of searches could be further supported by interconnecting related content in a more intense way. Possible interconnections could be revealed by applying bibliometric features like co-citations or author networks (Carevic & Schaer, 2014; Mutschke & Mayr, 2015), which are then added to the documents descriptive metadata.

- Extracting references and citations: Citations and references were both considered as highly useful and were frequently employed during the search task. However, the number of documents that contain citations and references is still comparably low. This issue has also been mentioned by the participants of the user study. Further research should be invested to increase the interconnectivity via citations and references, in order to better support users in exploring DLs.

Implications resulting from the transaction log study

The results of the transaction log study provided valuable insights into the potential of contextualised browsing features in modern DLs. The effects of contextualisation on the click position of documents resulting from stratagem browsing showed that users could be supported in discovering content that has the potential to be of relevance to their current search task. Not much effort is necessary to transfer the present approaches into a productive large scale DL. Both contextualisation variants are built upon existing standards such as the *More Like This* query parser and well-known query expansion methods.

To provide more user-oriented systems, that support the user during exploratory search tasks, designers of DLs should consider contextualisation as a potential enhancement that might help to improve their respective systems. Additionally, the response time for calculating the contextualised ranking was negligible, and usually, a response could be provided in a reasonable time. As performance has not been of particular interest in the experiment, there could still be room for improvement.

9.5 Future Work

In the following, open questions and potential future research directions are described.

Qualitative Investigation of Contextualised Browsing

The most important aspect of a follow-up experiment would be to evaluate the contextualisation in a controlled environment, such as a lab study. This would provide a more accurate and thorough impression of the effects of contextualised browsing. Given the results from the transaction log experiment, it is only possible to assess the effects from a quantitative perspective. The quantitative perspective, however, cannot provide insights regarding the actual benefits of contextualisation for the success of a search session.[43] The values from the MFR analysis show a higher click-through rate on documents that are further on top of the result list, but this can only be used as an indicator for effectiveness. Therefore, a user study should be conducted in which the participants can provide immediate feedback on the ranking. In order to conduct these types of experiments, a lot of effort has to be invested in the design of the study and the recruitment of participants.

Transparent User Interface

Providing a transparent user interface would allow users to control the contextualisation and to understand its impact. However, this is a challenging task that demands more research on the level of human-computer interaction. A good example of a transparent user interface can be found in di Sciascio et al. (2018). They developed a social exploratory search system that allows users to understand and control the ranking features. Any DL designed today should enable its users to gain control of the ranking methods applied. Furthermore, a transparent user interface could provide valuable feedback to determine in what situation a contextual ranking is demanded.

[43] The success of a search session could be, for instance, to satisfy an information need.

Diversity and Topical Narrowing

One aspect that was not investigated during the experiments is the balance between diversity and topical narrowing. A non-contextualised baseline offers more diversity by arranging content in a more or less ambiguous order. The contextualised ranking, on the other hand, is better suited for providing content in a topically narrow manor. Both approaches, however, may be adequate in different situations. It is possible that a user performs a journal run, in order to get a broad overview of the content that a particular journal provides. These types of searches can be characterised as serendipitous browsing. In those situations, the content is best presented using the non-contextualised baseline, in order to avoid a topical bias that would undoubtedly be present in a contextualised setting. A contextualised ranking, on the other hand, is better suited in situations in which a user is engaged in stratagem browsing to find related documents to an already retrieved relevant document.
These considerations, even more, suggest that a transparent user interface is necessary because it provides information about the present ranking with the ability to tune the ranking according to the users' needs.

Extending the Contextualisation Methods

With respect to the two contextualisation approaches, at least four potential areas for future work can be provided:

- The set of document features (see Section 7.3) that were employed to determine the similarity between two documents have not been evaluated separately. Judging whether the set of features used was adequate or not is not possible so far. Although the configuration has performed well, it may be necessary to adjust the similarity function to a more systematic and sophisticated approach that needs to be evaluated separately.

- The interaction context approach is developed from a limited set of features. Including additional parameters, that might contribute to a more precise interaction context, could improve the effectiveness of this approach. Examples for additional features could be the employment of mouse tracking positions, text selections (e.g. White & Buscher, 2012), or demographics (e.g. the language of the user).

- The query expansion on the basis of the interaction context employed a fielded weighting. This weighting is still at an early stage that was adapted from Sowiport's default configuration (see Section 7.4). A different weighting would undoubtedly have an effect on the performance. Evaluating different weighting parameters has not been tested and could be part of a follow-up investigation.

- The contextualisation based on the interaction context suffered from a cold-start problem. At the beginning of a session, only little information about

the user is available, which makes it difficult to estimate the user's search interest. Each interaction of the user with the system, however, increases the interaction context and in turn, the systems knowledge about the user. To overcome the cold-start problem, one could implement a hybrid contextualisation approach. A hybrid approach could utilise the similarity-based approach at the early stages of the session, while richer sessions that developed more knowledge about the user could be contextualised based on the interaction context.

The applicability of the present contextualisation concepts is not limited to DLs. Any other system that provides browsable content might benefit from a contextualisation. Thus, it would be of interest to transfer the present contextualisation concepts to a different use case. Social tagging systems, for instance, would provide an interesting use case. Transferring the present approaches would allow a comparison between contextualisation in DLs and, for example, social tagging systems.

Follow-up investigation

Based on the concepts presented in this thesis, a follow-up project on *Contextualised Dataset Retrieval (ConDATA)* has been granted for two years of research by the *German Research Foundation* (DFG). The goal of this project is to investigate contextualised browsing and its effects from a qualitative and user-centred design perspective. The two main objectives of this project are: 1) the development of a user interface that provides the users with information about the underlying contextual retrieval method, and 2) the extension of the present contextualisation methods.

In addition to the document retrieval perspective, this project will be focused on the dataset retrieval use case. This project not only provides more time for additional qualitative research, but it also shows that there is a lack of research in this area which is valuable and worth further investigation.

10

Conclusion

Exploratory search tasks that go beyond simple lookup searches, such as learning, are highly interactive and cognitively demanding tasks that involve various search tactics that occur iteratively, often across multiple search sessions. The information need of an exploratory search is usually ill-defined, and users seeking for information are accompanied by a feeling of uncertainty and confusion about what strategy to pursue, what information is available, and what terminology to use for searching. These characteristics are particularly observable during the initial stages of the information-seeking process, in which an information seeker aims to identify a suitable foundation, often in the form of a relevant document, that could expose possible directions to pursue the search. During these initial stages, information seekers regularly employ exploratory browsing strategies, in order to familiarise themselves with a particular topic, examine the information available, and to form a hypothesis. Today, exploratory browsing is recognised as a fundamental strategy during information seeking. Especially, academic search engines, such as, Digital Libraries, are designed to facilitate these types of searches in the form of stratagem browsing. Stratagem browsing enables the exploitation of structured metadata annotating the content of a document for further exploration. While the importance of stratagem browsing is undeniable, the system support on this level is comparably low. Many modern Digital Libraries only support the technical means of stratagem browsing, without any additional system support. Instead, stratagem browsing is realised from a solely system-oriented perspective that disregards essential user-centred aspects such as the user's information need, goals, and context of interaction.
The present thesis set out to study the applicability of contextualised stratagem browsing in Digital Libraries. Stratagem browsing aims at tailoring search results towards the user's search interests. In order to contribute to the development of contextualised stratagem browsing, three studies were conducted throughout this thesis. The results of these studies were presented in two distinct core chapters:

- **Empirical Studies:**
  The first chapter follows the principles of Interactive Information Retrieval, aiming to gain a thorough understanding of the information-seeking process with respect to stratagem browsing. Two studies, that analyse the behavioural aspects of stratagem browsing in Digital Libraries, were conducted: an online survey and a user study. During the online survey, a total of 204 respondents provided feedback on stratagem browsing and their usefulness for

finding related material. The online survey was followed up by a user study involving 32 participants that were asked to solve a pre-defined search task. This study was conducted in a controlled environment aiming to observe the information-seeking behaviour of participants, while looking for documents that were relevant to a given search task.
Contribution: Both, the online survey and the user study, contribute to a thorough understanding of stratagem browsing. The result of these two studies show that stratagem browsing is a highly relevant search activity, that is more frequently employed than query searches and the usage of recommended documents. The more frequent usage of stratagems is in particular visible during the early stages of the information-seeking process. A key observation made during these studies was that the employment of a stratagem is depended on the underlying content. Having very generic keywords or a topically broad journal has a strong influence on the decisions whether or not a stratagem variant is employed. This *topical broadness* is identified as a primary issue during stratagem browsing. This might be contra-intuitive to the objectives of browsing and therefore, indicates a need for a contextualised browsing approach that might help to overcome the identified issue of topical broadness.

- **Contextualised Stratagem Browsing in Digital Libraries:**
The second core chapter of the thesis presented the implementation and evaluation of contextualised stratagem browsing in a Digital Library. In order to address the problem of topical broadness, two approaches were developed that contextualise stratagem browsing on the level of descriptive metadata in a Digital Library: one variant which is based on document similarity, and one variant utilises the users' interaction context comprised of features such as queries, and different document attributes encountered during the users' session. To determine the effectiveness of the two contextualised stratagem browsing variants, a transaction log study was conducted using the real-life Digital Library Sowiport. An A/B/C-testing was designed to compare the two contextual stratagem browsing variants against Sowiports non-contextual baseline.
Contribution: The evaluation of contextualised stratagem browsing shows that contextualisation is more effective than the non-contextualised baseline. The results indicate a higher click-through rate and a significantly better performance in terms of the first relevant rank position, concerning the contextualised browsing variants. Furthermore, the contextualised variants achieve a higher usefulness measured by the number of implicit relevance signals compared to the baseline. Overall, the results of this experiment suggest that contextualised stratagem browsing can help to overcome the disadvantages of topical broadness. This might lead to a new generation of contextualised Digital Libraries.

The following hypothesis was presented in the introduction of this thesis:

The employment of the user's interaction context on the level of stratagem browsing will lead to the development of more effective retrieval systems.

The results presented in this thesis provided experimental evidence that contextualisation is more effective than non-contextualised stratagem browsing. Both contextual browsing variants were developed using well-established methods that are widely known. The use of well-established methods creates the opportunity for the designers of Digital Library to incorporate and adapt the proposed methods. Furthermore, this thesis contributed to research by providing a more profound understanding of stratagem browsing and exposed an inadequateness in the operationalisation of stratagems described as topical broadness. Ultimately, this thesis laid the foundations for contextualised stratagem browsing in Digital Libraries and created opportunities for follow-up research directions.

Bibliography

Adomavicius, Gediminas and Tuzhilin, Alexander. Context-aware recommender systems. In *Recommender systems handbook*, pages 217–253. Springer, 2011.

Agosti, Maristella, Ferro, Nicola, Fox, Edward A, and Gonçalves, Marcos A. Modelling dl quality: a comparison between approaches: the delos reference model and the 5s model. In *Second DELOS Conference on Digital Libraries Working Notes*, 2007.

Aliannejadi, Mohammad and Crestani, Fabio. Personalized context-aware point of interest recommendation. *ACM Transactions on Information Systems (TOIS)*, 36 (4):1–28, 2018.

Arms, William Y. *Digital libraries*. MIT press, 2001. ISBN 9780262255653. doi: 10.7551/mitpress/2423.001.0001.

Aronson, Alan R and Rindflesch, Thomas C. Query expansion using the umls metathesaurus. In *Proceedings of the AMIA Annual Fall Symposium*, pages 485–489. American Medical Informatics Association, 1997.

Athukorala, Kumaripaba, Hoggan, Eve, Lehtiö, Anu, Ruotsalo, Tuukka, and Jacucci, Giulio. Information-seeking behaviors of computer scientists: Challenges for electronic literature search tools. *Proceedings of the American Society for Information Science and Technology*, 50(1):1–11, 2013. doi: 10.1002/meet.14505001041.

Athukorala, Kumaripaba, Głowacka, Dorota, Jacucci, Giulio, Oulasvirta, Antti, and Vreeken, Jilles. Is exploratory search different? a comparison of information search behavior for exploratory and lookup tasks. *Journal of the Association for Information Science and Technology*, 67(11):2635–2651, 2016. doi: 10.1002/asi.23617.

Azad, Hiteshwar Kumar and Deepak, Akshay. Query expansion techniques for information retrieval: A survey. *Information Processing & Management*, 56(5): 1698–1735, 2019. doi: 10.1016/j.ipm.2019.05.009.

Azzopardi, Leif. Modelling interaction with economic models of search. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 3–12. ACM, 2014. doi: 10.1145/2600428.2609574.

Baeza-Yates, Ricardo, Ribeiro-Neto, Berthier, et al. *Modern information retrieval*, volume 463. ACM press New York, 1999.

Bates, Marcia J. The design of browsing and berrypicking techniques for the online search interface. *Online review*, 13(5):407–424, 1989. doi: 10.1108/eb024320.

Bates, Marcia J. Where should the person stop and the information search interface start? *Information Processing & Management*, 26(5):575–591, 1990. doi: 10.1016/0306-4573(90)90103-9.

Bazire, Mary and Brézillon, Patrick. Understanding context before using it. In *International and Interdisciplinary Conference on Modeling and Using Context*, pages 29–40. Springer, 2005. doi: 10.1007/11508373_3.

Beel, Joeran, Aizawa, Akiko, Breitinger, Corinna, and Gipp, Bela. Mr. dlib: recommendations-as-a-service (raas) for academia. In *Digital Libraries (JCDL), 2017 ACM/IEEE Joint Conference on*, pages 1–2. IEEE, 2017a. doi: 10.1109/jcdl.2017.7991606.

Beel, Joeran, Carevic, Zeljko, Schaible, Johann, and Neusch, Gábor. Rard: The related-article recommendation dataset. *D-Lib Magazine*, 23(7/8), 2017b. doi: 10.1045/july2017-beel.

Beel, Joeran, Dinesh, Siddharth, Mayr, Philipp, Carevic, Zeljko, and Raghvendra, Jain. Stereotype and most-popular recommendationsin the digital library sowiport. In *Proceedings of the 15th International Symposium of Information Science (ISI)*. Humboldt-Universität zu Berlin, 2017c.

Belkin, Nicholas J. Anomalous states of knowledge as a basis for information retrieval. *Canadian journal of information science*, 5(1):133–143, 1980.

Belkin, Nicholas J. Interaction with texts: Information retrieval as information-seeking behavior. In *Information Retrieval '93: Von der Modellierung zur Anwendung*, volume 12, pages 55–66. Universitätsverlag Konstanz, 1993.

Bethlehem, Jelke. Selection bias in web surveys. *International Statistical Review*, 78(2):161–188, 2010. doi: 10.1111/j.1751-5823.2010.00112.x.

Borgman, Christine L. What are digital libraries? Competing visions. *Information Processing and Management*, 35(3):227–243, 1999.

Borlund, Pia. Experimental components for the evaluation of interactive information retrieval systems. *Journal of documentation*, 56(1):71–90, 2000. doi: 10.1108/eum0000000007110.

Borlund, Pia. The iir evaluation model: a framework for evaluation of interactive information retrieval systems. *Information Research*, 8(3), 2003. `http://informationr.net/ir/8-3/paper152.html`, last accessed April 15th, 2020.

Borlund, Pia and Ingwersen, Peter. The development of a method for the evaluation of interactive information retrieval systems. *Journal of documentation*, 53(3): 225–250, 1997. doi: 10.1108/eum0000000007198.

Buckley, Chris, Salton, Gerard, Allan, James, and Singhal, Amit. Automatic query expansion using smart: Trec 3. In *Proceedings of the 3rd Text Retrieval Conference (TREC-3)*, pages 69–80. National Institute of Standards & Technology, 1995.

Candela, Leonardo, Castelli, Donatella, Ferro, Nicola, Ioannidis, Yannis, Koutrika, Georgia, Meghini, Carlo, Pagano, Pasquale, Ross, Seamus, Soergel, Dagobert, Agosti, Maristella, Dobreva-McPherson, Milena, Katifori, Vivi, and Schuldt, Heiko. *The DELOS Digital Library Reference Model - Foundations for Digital Libraries*. 02 2007. DELOS Network of Excellence on Digital Libraries Project no. 507618.

Carevic, Zeljko and Mayr, Philipp. Extending search facilities via bibliometric-enhanced stratagems. In *Proc. of the 2nd Workshop on Bibliometric-enhanced Information Retrieval*, pages 40–46, Vienna, Austria, 2015. CEUR-WS.org.

Carevic, Zeljko and Mayr, Philipp. Survey on high-level search activities based on the stratagem level in digital libraries. In *International Conference on Theory and Practice of Digital Libraries*, pages 54–66. Springer, 2016. doi: 10.1007/978-3-319-43997-6_5.

Carevic, Zeljko and Schaer, Philipp. On the connection between citation-based and topical relevance ranking: Results of a pretest using isearch. In *Proc. of the 1nd Workshop on Bibliometric-enhanced Information Retrieval*, pages 37–44, Amsterdam, Netherlands, 2014. CEUR-WS.org.

Carevic, Zeljko, Lusky, Maria, van Hoek, Wilko, and Mayr, Philipp. Investigating exploratory search activities based on the stratagem level in digital libraries. *International Journal on Digital Libraries*, 19(2-3):231–251, 2018a.

Carevic, Zeljko, Schüller, Sascha, Mayr, Philipp, and Fuhr, Norbert. Contextualised browsing in a digital library's living lab. In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*, JCDL '18, pages 89–98, New York, NY, USA, 2018b. ACM. ISBN 978-1-4503-5178-2. doi: 10.1145/3197026.3197054.

Case, Donald Owen. *Looking for information: A survey of research on information seeking, needs, and behavior*. Academic Press, 2002.

Chirita, Paul-Alexandru, Firan, Claudiu S, and Nejdl, Wolfgang. Personalized query expansion for the web. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 7–14. ACM, 2007. doi: 10.1145/1277741.1277746.

Claypool, Mark, Le, Phong, Wased, Makoto, and Brown, David. Implicit interest indicators. In *Proceedings of the 6th international conference on Intelligent user interfaces*, pages 33–40. ACM, 2001. doi: 10.1145/359784.359836.

Cleveland, Gary. Digital libraries: definitions, issues and challenges. IFLA, Universal dataflow and telecommunications core programme, 1998. `https://archive.ifla.org/VI/5/op/udtop8/udt-op8.pdf`, last accessed April 29th, 2020.

Cool, Colleen and Spink, Amanda. Issues of context in information retrieval (ir): an introduction to the special issue. *Information Processing & Management*, 38 (5):605–611, 2002. doi: 10.1016/s0306-4573(01)00054-1.

Croft, W Bruce, Metzler, Donald, and Strohman, Trevor. *Search engines: Information retrieval in practice*. Addison-Wesley Reading, 2009.

di Sciascio, Cecilia, Brusilovsky, Peter, and Veas, Eduardo. A study on user-controllable social exploratory search. In *23rd International Conference on Intelligent User Interfaces*, pages 353–364. ACM, 2018. doi: 10.1145/3172944.3172986.

Doslu, Metin and Bingol, Haluk O. Context sensitive article ranking with citation context analysis. *Scientometrics*, 108(2):653–671, 2016. doi: 10.1007/s11192-016-1982-6.

Ellis, David. A behavioural approach to information retrieval system design. *Journal of documentation*, 45(3):171–212, 1989. doi: 10.1108/eb026843.

Fagan, Jody Condit. Usability studies of faceted browsing: A literature review. *Information Technology and Libraries*, 29(2):58–66, 2010. doi: 10.6017/ital.v29i2.3144.

Fox, Edward A, Goncalves, Marcos Andre, and Shen, Rao. *Theoretical foundations for digital libraries: The 5S (societies, scenarios, spaces, structures, streams) approach*, volume 4. Morgan & Claypool Publishers, 2012. doi: 10.2200/s00434ed1v01y201207icr022.

Fuhr, Norbert. Salton award lecture information retrieval as engineering science. In *ACM SIGIR Forum*, volume 46, pages 19–28. ACM, 2012. doi: 10.1145/2422256.2422259.

Fuhr, Norbert. Some common mistakes in ir evaluation, and how they can be avoided. volume 51, page 32–41, New York, NY, USA, February 2018. Association for Computing Machinery. doi: 10.1145/3190580.3190586.

Fuhr, Norbert, Hansen, Preben, Mabe, Michael, Micsik, András, and Sølvberg, Ingeborg. Digital libraries: A generic classification and evaluation scheme. In *International Conference on Theory and Practice of Digital Libraries*, pages 187–199. Springer, 2001. doi: 10.1007/3-540-44796-2_17.

Fuhr, Norbert, Klas, Claus-Peter, Schaefer, André, and Mutschke, Peter. Daffodil: An integrated desktop for supporting high-level search activities in federated digital libraries. In *International Conference on Theory and Practice of Digital Libraries*, pages 597–612. Springer, 2002. doi: 10.1007/3-540-45747-x_45.

Fuhr, Norbert, Tsakonas, Giannis, Aalberg, Trond, Agosti, Maristella, Hansen, Preben, Kapidakis, Sarantos, Klas, Claus-Peter, Kovács, László, Landoni, Monica, Micsik, András, et al. Evaluation of digital libraries. *International Journal on Digital Libraries*, 8(1):21–38, 2007. doi: 10.1007/s00799-007-0011-z.

Ganguly, Debasis, Roy, Dwaipayan, Mitra, Mandar, and Jones, Gareth JF. Word embedding based generalized language model for information retrieval. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, pages 795–798, 2015. doi: 10.1145/2766462.2767780.

Goncalves, Marcos Andre. *Streams, Structures, Spaces, Scenarios, and Societies (5S): A Formal Digital Library Framework and Its Applications*. PhD thesis, Virginia Tech, 2004.

Gonçalves, Marcos André, Fox, Edward A, Watson, Layne T, and Kipp, Neill A. Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries. *ACM transactions on information systems (TOIS)*, 22(2):270–312, 2004. ISSN 1046-8188. doi: 10.1145/984321.984325.

Gonçalves, Marcos André, Moreira, Bárbara L, Fox, Edward A, and Watson, Layne T. "What is a good digital library?"–a quality model for digital libraries. *Information processing & Management*, 43(5):1416–1437, 2007. doi: 10.1016/j.ipm.2006.11.010.

Greenwald, Glen. The Gurdian, 2013. `https://www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order`, last accessed May 1st, 2020.

Hearst, Marti. *Search user interfaces*. Cambridge University Press, 2009. doi: 10.1017/cbo9781139644082. Accessed: 2019-08-28.

Hearst, Marti A. Clustering versus faceted categories for information exploration. *Communications of the ACM*, 49(4):59–61, 2006. doi: 10.1145/1121949.1121983.

Hemminger, Bradley M, Lu, Dihui, Vaughan, KTL, and Adams, Stephanie J. Information seeking behavior of academic scientists. *Journal of the American society for information science and technology*, 58(14):2205–2225, 2007. doi: 10.1002/asi.20686.

Hienert, Daniel and Lusky, Maria. Where do all these search terms come from?–two experiments in domain-specific search. In *European Conference on Information Retrieval*, pages 15–26. Springer, 2017. doi: 10.1007/978-3-319-56608-5_2.

Hienert, Daniel and Mutschke, Peter. A usefulness-based approach for measuring the local and global effect of iir services. In *Proceedings of the 2016 ACM Conference on Human Information Interaction and Retrieval*, pages 153–162. ACM, 2016. doi: 10.1145/2854946.2854962.

Hienert, Daniel, Schaer, Philipp, Schaible, Johann, and Mayr, Philipp. A novel combined term suggestion service for domain-specific digital libraries. In *International Conference on Theory and Practice of Digital Libraries*, pages 192–203. Springer, 2011. doi: https://doi.org/10.1007/978-3-642-24469-8_21.

Hienert, Daniel, Sawitzki, Frank, and Mayr, Philipp. Digital library research in action: supporting information retrieval in sowiport. *D-Lib Magazine*, 21(3):8, 2015a. doi: 10.1045/march2015-hienert.

Hienert, Daniel, van Hoek, Wilko, Weber, Alina, and Kern, Dagmar. Whose–a tool for whole-session analysis in iir. In *European Conference on Information Retrieval*, pages 172–183. Springer, 2015b. doi: 10.1007/978-3-319-16354-3_18.

Ingwersen, Peter. Context in information interaction–revisited 2006. In *Proceedings of the Fourth Biennial DISSAnet Conference. Pretoria: University of Pretoria*, pages 13–23, 2006.

Ingwersen, Peter and Järvelin, Kalervo. Information retrieval in context: Irix. In *Acm sigir forum*, volume 39, pages 31–39. ACM, 2005a.

Ingwersen, Peter and Järvelin, Kalervo. *The turn: Integration of information seeking and retrieval in context*. Springer-Verlag, Dordrecht, 2005b. ISBN 978-1-4020-3851-8.

Ingwersen, Peter and Järvelin, Kalervo. On the holistic cognitive theory for information retrieval. In *Proceedings of the First International Conference on the Theory of Information Retrieval (ICTIR) (Budapest, Hungary, 2007). Foundation for Information Society*, pages 135–147, 2007.

Isah, Abdulmumin, Serema, Batlang Comma, Mutshewa, Athulang, and Kenosi, Lekoko. Digital libraries: analysis of delos reference model and 5s theory. *Journal of Information Science Theory and Practice*, 1(4):38–47, 2013. doi: 10.1633/jistap.2013.1.4.3.

Jamali, Hamid R. and Nicholas, David. Interdisciplinarity and the information-seeking behavior of scientists. *Information Processing & Management*, 46(2): 233 – 243, 2010. doi: 10.1016/j.ipm.2009.12.010.

Joachims, Thorsten, Granka, Laura, Pan, Bing, Hembrooke, Helene, and Gay, Geri. Accurately interpreting clickthrough data as implicit feedback. *SIGIR Forum*, 51 (1):4–11, August 2017. doi: 10.1145/3130332.3130334.

Jones, Steve, Cunningham, Sally Jo, McNab, Rodger, and Boddie, Stefan. A transaction log analysis of a digital library. *International Journal on Digital Libraries*, 3(2):152–169, 2000. doi: 10.1007/s007999900022.

Kacem, Ameni and Mayr, Philipp. Analysis of search stratagem utilisation. *Scientometrics*, 116(2):1383–1400, August 2018. doi: 10.1007/s11192-018-2821-8.

Kang, Ruogu and Fu, Wai-Tat. Exploratory information search by domain experts and novices. In *Proceedings of the 15th international conference on Intelligent user interfaces*, pages 329–332. ACM, 2010. doi: 10.1145/1719970.1720023.

Kantor, Paul B and Voorhees, Ellen M. The trec-5 confusion track: Comparing retrieval methods for scanned text. *Information Retrieval*, 2(2-3):165–176, 2000.

Kelly, Diane. Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval*, 3(1—2):1–224, 2009.

Kelly, Diane and Belkin, Nicholas J. Reading time, scrolling and interaction: exploring implicit sources of user preferences for relevance feedback. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 408–409. ACM, 2001. doi: 10.1145/383952.384045.

Kelly, Diane and Teevan, Jaime. Implicit feedback for inferring user preference: A bibliography. *SIGIR Forum*, 37(2):18–28, September 2003. doi: 10.1145/959258.959260.

Kenter, Tom and de Rijke, Maarten. Short text similarity with word embeddings. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, CIKM '15, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450337946. doi: 10.1145/2806416.2806475.

Kuhlthau, Carol C. Inside the search process: Information seeking from the user's perspective. *Journal of the American society for information science*, 42(5): 361–371, 1991.

Kuhlthau, Carol C. A principle of uncertainty for information seeking. *Journal of documentation*, 49(4):339–355, 1993. doi: 10.1108/eb026918.

Kules, Bill and Capra, Robert. Influence of training and stage of search on gaze behavior in a library catalog faceted search interface. *Journal of the American Society for Information Science and Technology*, 63(1):114–138, 2012.

Kules, Bill, Capra, Robert, Banta, Matthew, and Sierra, Tito. What do exploratory searchers look at in a faceted search interface? In *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries*, JCDL '09, page 313–322, New York, NY, USA, 2009. Association for Computing Machinery. doi: 10.1145/1555400.1555452.

Lesk, Michael. *Practical Digital Libraries: Books, Bytes, and Bucks*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1997. ISBN 1558604596.

Li, Xinyi and de Rijke, Maarten. Characterizing and predicting downloads in academic search. *Information Processing & Management*, 56(3):394–407, 2019. doi: 10.1016/j.ipm.2018.10.019.

Li, Yuelin and Belkin, Nicholas J. A faceted approach to conceptualizing tasks in information seeking. *Information Processing & Management*, 44(6):1822 – 1837, 2008. doi: 10.1016/j.ipm.2008.07.005. Adaptive Information Retrieval.

Lykke, Marianne, Larsen, Birger, Lund, Haakon, and Ingwersen, Peter. Developing a test collection for the evaluation of integrated search. In *European Conference on Information Retrieval*, pages 627–630. Springer, Springer Berlin Heidelberg, 2010. doi: 10.1007/978-3-642-12275-0_63.

Mangan, Rachel. Web of Science Core Collection - Descriptive Document, 2018. `https://clarivate.libguides.com/ld.php?content_id=45175981`, last accessed May 1st, 2020.

Manning, Christopher D, Raghavan, Prabhakar, and Schütze, Hinrich. *Introduction to information retrieval*. Cambridge university press, 2008. ISBN 978-0-521-86571-5. `http://nlp.stanford.edu/IR-book/information-retrieval-book.html`, last accessed May 1st, 2020.

Marchionini, Gary. Exploratory search: From finding to understanding. *Commun. ACM*, 49(4):41–46, April 2006. doi: 10.1145/1121949.1121979.

Mayr, Philipp. How do practitioners, phd students and postdocs in the social sciences assess topic-specific recommendations? In *Proceedings of the Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL)*, pages 84–92, 2016.

Mayr, Philipp and Petras, Vivien. Cross-concordances: terminology mapping and its effectiveness for information retrieval. In *74th IFLA World Library and Information Congress*, Quebec, Canada, 2008. IFLA. `http://www.ifla.org/IV/ifla74/papers/129-Mayr_Petras-en.pdf`, last accessed May 1st, 2019.

Meho, Lokman I and Tibbo, Helen R. Modeling the information-seeking behavior of social scientists: Ellis's study revisited. *Journal of the American society for*

Information Science and Technology, 54(6):570–587, 2003. doi: 10.1002/asi.10244.

Morita, Masahiro and Shinoda, Yoichi. Information filtering based on user behavior analysis and best match text retrieval. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '94, page 272–281, Berlin, Heidelberg, 1994. Springer-Verlag.

Mutschke, Peter and Mayr, Philipp. Science models for search: a study on combining scholarly information retrieval and scientometrics. *Scientometrics*, 102 (3):2323–2345, 2015. doi: 10.1007/s11192-014-1485-2.

Niu, Xi, Hemminger, Bradley M, Lown, Cory, Adams, Stephanie, Brown, Cecelia, Level, Allison, McLure, Merinda, Powers, Audrey, Tennant, Michele R, and Cataldo, Tara. National study of information seeking behavior of academic researchers in the united states. *Journal of the American Society for Information Science and Technology*, 61(5):869–890, 2010. doi: 10.1002/asi.21307.

Pirolli, Peter and Card, Stuart. Information foraging in information access environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, page 51–58, USA, 1995. ACM Press/Addison-Wesley Publishing Co. doi: 10.1145/223904.223911.

Pontis, Sheila and Blandford, Ann. Understanding "influence:" an exploratory study of academics' processes of knowledge construction through iterative and interactive information seeking. *Journal of the Association for Information Science and Technology*, 66(8):1576–1593, 2015. doi: 10.1002/asi.23277.

Pontis, Sheila, Blandford, Ann, Greifeneder, Elke, Attalla, Hesham, and Neal, David. Keeping up to date: An academic researcher's information journey. *Journal of the Association for Information Science and Technology*, 68(1):22–35. doi: 10.1002/asi.23623.

Reichle, Erik D, Pollatsek, Alexander, and Rayner, Keith. Using ez reader to simulate eye movements in nonreading tasks: A unified framework for understanding the eye–mind link. *Psychological review*, 119(1):155–185, 2012. doi: 10.1037/a0026473.

Riley, Jenn. Understanding metadata. *National Information Standards Organization (NISO)*, 2017. `https://groups.niso.org/apps/group_public/download.php/17446/UnderstandingMetadata.pdf`, last accessed May 1st, 2019.

Robins, David. Interactive information retrieval: Context and basic notions. *Informing Science Journal*, 3:57–62, 2000. doi: 10.28945/577.

Rocchio, Joseph. Relevance feedback in information retrieval. *The Smart retrieval system-experiments in automatic document processing*, pages 313–323, 1971.

Roy, Dwaipayan, Ganguly, Debasis, Bhatia, Sumit, Bedathur, Srikanta, and Mitra, Mandar. Using word embeddings for information retrieval: How collection and term normalization choices affect performance. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, page 1835–1838, New York, NY, USA, 2018. Association for Computing Machinery. doi: 10.1145/3269206.3269277.

Salton, G. *Automatic information organization and retrieval*. McGraw-Hill computer science series. McGraw-Hill, 1968.

Salton, Gerard and Buckley, Chris. Improving retrieval performance by relevance feedback. *Journal of the American society for information science*, 41(4):288–297, 1990.

Sanderson, Mark. Test collection based evaluation of information retrieval systems. *Foundations and Trends in Information Retrieval*, 4(4):247–375, 2010. ISSN 1554-0669. doi: 10.1561/1500000009.

Sanderson, Mark and Croft, Bruce W. The history of information retrieval research. *Proceedings of the IEEE*, 100(Special Centennial Issue):1444–1451, 2012. doi: 10.1109/jproc.2012.2189916.

Saracevic, Tefko. Relevance reconsidered. In *Proceedings of the second conference on conceptions of library and information science (CoLIS 2)*, pages 201–218. ACM New York, 1996.

Saracevic, Tefko. The stratified model of information retrieval interaction: Extension and applications. In *Proceedings of the Annual Meeting-American Society for Information Science*, volume 34, pages 313–327, 1997.

Saracevic, Tefko. Digital library evaluation: Toward an evolution of concepts. *LIBRARY TRENDS*, 49(2):350–369, 2000.

Shen, Xuehua, Tan, Bin, and Zhai, ChengXiang. Implicit user modeling for personalized search. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, CIKM '05, page 824–831, New York, NY, USA, 2005. Association for Computing Machinery. doi: 10.1145/1099554.1099747.

SSOAR. Social science open access repository, 2020. `https://www.gesis.org/en/ssoar/home/about-ssoar`, last accessed May 1st, 2020.

Tenopir, Carol, King, Donald W, Spencer, Jesse, and Wu, Lei. Variations in article seeking and reading patterns of academics: What makes a difference? *Library & Information Science Research*, 31(3):139–148, 2009. doi: https://doi.org/10.1016/j.lisr.2009.02.002.

van Hoek, Wilko and Carevic, Zeljko. Building user groups based on a structural representation of user search sessions. In *International Conference on Theory and Practice of Digital Libraries*, pages 459–470. Springer, 2017.

White, Ryen W. *Interactions with search systems*. Cambridge University Press, New York, 2016. doi: 10.1017/CBO9781139525305.

White, Ryen W. and Buscher, Georg. Text selections as implicit relevance feedback. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, page 1151–1152, New York, NY, USA, 2012. Association for Computing Machinery. doi: 10.1145/2348283.2348514.

White, Ryen W and Roth, Resa A. Exploratory search: Beyond the query-response paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 1(1):1–98, 2009. doi: 10.2200/s00174ed1v01y200901icr003.

White, Ryen W, Jose, Joemon M, and Ruthven, Ian. Comparing explicit and implicit feedback techniques for web retrieval: Trec-10 interactive track report. In *Proceedings of the Tenth Text Retrieval Conference (TREC-10)*, pages 534–538, 2001.

White, Ryen W, Kules, Bill, and Bederson, Ben. Exploratory search interfaces: categorization, clustering and beyond: report on the xsi 2005 workshop at the human-computer interaction laboratory, university of maryland. In *ACM SIGIR Forum*, volume 39, pages 52–56. ACM, 2005.

White, Ryen W, Kules, Bill, Drucker, Steven M, et al. Supporting exploratory search. *Communications of the ACM*, 49(4):36–39, 2006. doi: 10.1145/1121949.1121978.

White, Ryen W., Bailey, Peter, and Chen, Liwei. Predicting user interests from contextual information. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '09, page 363–370, New York, NY, USA, 2009. Association for Computing Machinery. doi: 10.1145/1571941.1572005.

Wildemuth, Barbara M. The effects of domain knowledge on search tactic formulation. *Journal of the American Society for Information Science and Technology*, 55(3):246–258, 2004. doi: 10.1002/asi.10367.

Wilson, TD. Human information behavior. *Informing Science: the International Journal of an Emerging Transdiscipline*, 3:49–56, 2000. doi: 10.28945/576.

Wilson, Tom D. Models in information behaviour research. *Journal of Documentation*, 55(3):249–270, 1999. doi: 10.1108/EUM0000000007145.

Witten, Ian H, Bainbridge, David, and Nichols, David M. *How to build a digital library*. Morgan Kaufmann, 2009. ISBN 978-0-12-374857-7. doi: 10.1016/C2009-0-19701-5.

Zamani, Hamed and Croft, W. Bruce. Embedding-based query language models. In *Proceedings of the 2016 ACM International Conference on the Theory of Information Retrieval*, ICTIR '16, page 147–156, New York, NY, USA, 2016. Association for Computing Machinery. doi: 10.1145/2970398.2970405.

Zapilko, Benjamin, Schaible, Johann, Mayr, Philipp, and Mathiak, Brigitte. Thesoz: A skos representation of the thesaurus for the social sciences. *Semantic Web*, 4(3):257–263, 2013. doi: 10.3233/sw-2012-0081.

Appendices

A

Tables

A.1 Task on organising journal articles (Online Survey)

Table A.1: Task on organizing journal articles. Mean values range from lowest rank (6) to highest rank (1) (N≥128).

Ranking option	M	SD	Mdn	Mode
By the entered query terms (alcohol consumption germany)	2.08	1.34	2	1
By similarity to the current document based on title (Developments in alcohol consumption..)	2.23	1.32	2	2
By title	3.95	1.49	4	5
By issue and date	3.95	1.66	4	6
By number of citations	4.08	1.42	4	4
By author	4.42	1.31	5	6

A.2 Usage frequency of the six stratagems in comparison to queries and recommendations

Table A.2: Total usage frequency of the six stratagems in comparison to queries and recommendations.

Type	Search Activity	Usage frequency	# Participants
Stratagem	Keywords	50	16
	References	27	18
	Citations	26	18
	Author	25	13
	Classifications	5	5
	Journals	4	2
	total stratagem	**137**	
Other	Queries	69	22
	Recommendations	42	16
	total other	**111**	

A.3 Percentaged usage frequency of search activities within a session limit of ten steps

Table A.3: Percentaged usage frequency of search activities within a session limit of ten steps.

Type	Search activity	Step 1	Steps 2-4	Steps 5-7	Steps 8-10
Stratagem	Keywords	21.8%	16.1%	18.7%	33.3%
	Citations	21.8%	8.6%	10.9%	8.3%
	References	18.7%	11.8%	3.1%	13.8%
	Author	3.1%	10.7%	14.0%	0%
	Classifications	3.1%	3.2%	1.5%	0%
	Journal	0%	3.2%	0%	0%
Other	Queries	15.6%	29.0%	37.5%	33.3%
	Recommendations	15.6%	17.2%	14.0%	11.1%

A.4 Task on organising journal articles (User Study)

Table A.4: Task on organizing journal articles. Mean values range from lowest rank (6) to highest rank (1) (N=32).

Ranking option	M	SD	Mdn	Mode
By the entered query terms (alcohol consumption germany)	1.81	1.07	1.50	1
By similarity to the current document based on title (Developments in alcohol consumption..)	2.47	1.44	2	2
By issue and date	4.00	1.30	4	5
By title	4.09	1.74	4	6
By number of citations	4.31	1.49	4	6
By author	4.31	1.24	4.5	5

A.5 Descriptive statistics for the period of the transaction log study

Table A.5: Descriptive statistics for the period of the study.

Approach	Total stratagem usage	Document views from stratagem search	Mean interactions per session	Mean dwell time (s)
(A) Baseline	25,426	1,985	7.61	123.79
(B) Similarity	25,475	3,212	7.91	134.98
(C) Interaction Context	26,135	2,627	7.76	123.27

A.6 Mean first relevant for different history sizes (H)

Table A.6: Mean first relevant for different history sizes (H).

Approach	MFR H $\in [2, 5]$	MFR H $\in [6, 10]$	MFR H $\in [11, \infty]$
(A) Baseline	4.52 (N=802)	5.06 (N=276)	5.49 (N=112)
(B) Similarity	3.04 (N=1491)	3.29 (N=508)	3.59 (N=215)
(C) Interaction context	3.58 (N=1201)	3.77 (N=370)	3.80 (N=160)

A.7 Usefulness of stratagem browsing per session

Table A.7: Usefulness of stratagem browsing per session.

Approach	Local usefulness	Global usefulness
(A) Baseline	232	5,385
(B) Similarity	628	5,684
(C) Session Context	334	5,294

A.8 Descriptive metadata of Sowiport records

Table A.8: Descriptive metadata of records in Sowiport associated with their browsing capabilities.

Metadata Field	Browsable
Title	No
Abstract	No
Keywords	Yes
Classifications	Yes
Publication Information	
Journal	Yes
Published Date	No
DOI	No
Page Numbers	No
Authors	Yes
References	Yes (if available)
Citations	Yes (if available)

B

Survey questionnaire

1 Welcome

Welcome, and thank you for participating in the survey on:

Higher Search Activities in Digital Libraries.

General Information:
This survey aims to understand the usage of higher search activities in Digital Libraries (DLs).

Filling out the survey will take approximately 5 minutes.

What is the survey about?
If you choose to complete this survey, you will help us to identify possible requirements for supporting higher search activities in retrieval sessions. Higher search activities are for example:

- Reading or browsing a journal or conference proceedings central to one's topic of interest
- Using citations to determine which other works have cited a certain document
- Following footnotes or references to other related materials

I appreciate your participation very much, as you thereby support me in my Ph.D. thesis. Participation in this survey is voluntary and you are free to withdraw or discontinue participation at any time.

We will publish the results of this survey after thorough analysis. We do not collect any sensitive personal data and the results will be anonymous.

Click on "Continue" to begin the survey.

Thank you very much for your interest and support. If you have any questions or comments, do not hesitate to email me at:
zeljko.carevic@gesis.org

2 Journal Run

Journal and Conference Proceedings 1 of 2

How useful are conference proceedings or journals as a source for relevant documents during your search task?

Not at all useful	Rather not useful	Neither useful nor not useful	Rather useful	Very useful

How satisfied are you with the support of current Digital Libraries (e.g. ACM DL, Web of Science) in browsing through conference proceedings or journals?

Not at all satisfied	Rather unsatisfied	Neither satisfied nor unsatisfied	Rather satisfied	Very satisfied

How important is the quality of a conference (ranking) or a journal (e.g. the impact factor) for your confidence in the source?

Not at all important	Rather unimportant	Neither unimportant nor important	Rather important	Very important

3 Journal Run 2

Journal and Conference Proceedings 2 of 2

How often do you browse through conference proceedings or journals to find relevant documents?

○ Never ○ Rarely ○ Sometimes ○ Often ○ Very often

If you never or rarely browse through conference proceedings or journals, please explain why.

After finding a document (e.g. ACM DL, Web of Science) that is relevant for your current search task: How often do you browse through the conference proceedings or the journal the document was published in?

○ Never ○ Rarely ○ Sometimes ○ Often ○ Very often

If you never or rarely browse through conference proceedings or journals starting from a relevant document, please explain why.

4 Citations

Citations 1 of 2

How important is the number of citations a document has received to you?

○ Not at all important ○ Rather unimportant ○ Neither important nor unimportant ○ Rather important ○ Very important

How would you rate the usefulness of citation rankings (e.g. h-index) where documents are ranked by the number of received citations?

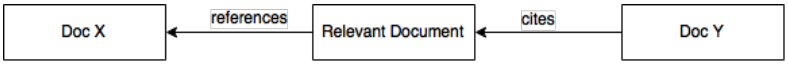
○ Not at all useful ○ Rather not useful ○ Neither useful nor not useful ○ Rather useful ○ Very useful

Assuming there is a key document in a particular field. How important is it to you to find central authors citing that particular document?

○ Not at all important ○ Rather unimportant ○ Neither important nor unimportant ○ Rather important ○ Very important

5 Citations 2

Citations 2 of 2



Starting from a relevant document: How often do you use references (see illustration) to find other relevant documents for your search task?

○ Never ○ Rarely ○ Sometimes ○ Often ○ Very often

If you never or rarely use references to find other relevant documents, please explain why.

Starting from a relevant document: How often do you use citations (see illustration) to find other relevant documents for your search task?

Never Rarely Sometimes Often Very often

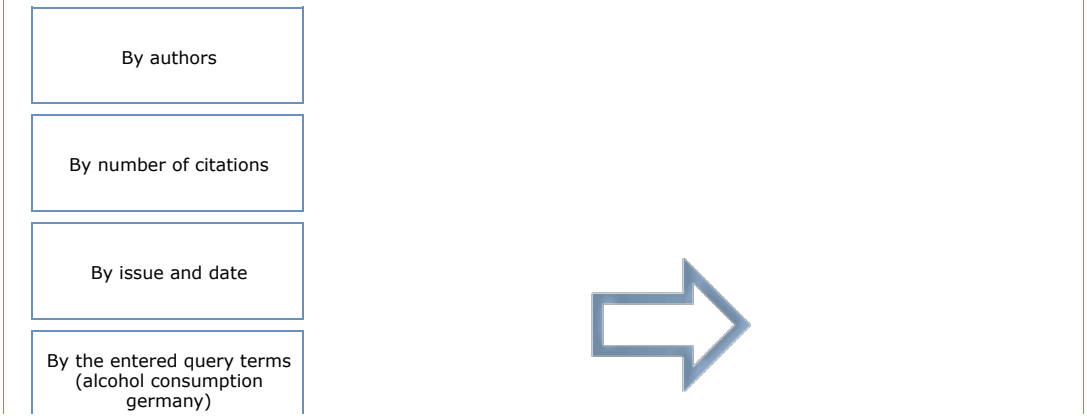
If you never or rarely use citations to find other relevant documents, please explain why.

6 Scenario 2

Please consider the following situation: You are about to write an essay about 'Alcohol Consumption in Germany and its Demographic Distribution'. You start your search by entering the search terms 'alcohol consumption germany'. You find a relevant document (see illustration) that was published in a journal named 'Addiction'. After reading the document you want to see more material from that particular journal.

How should documents from that journal be ranked in the result list? Please order the following options from best to worst.

Please move all options from left to right. Please note that some options refer to the illustration.



By title

By similarity to the current document based on title (Developments in alcohol consumption..)

7 Basic Stratagems

Please consider the following scenario. You want to find out about the current state of the art in a particular field. You have already found one document that is useful to your current work task.

Which of the following search activities do you perform to find other related materials? Please order the following options from best to worst.

Please move all options from left to right.

I use the keywords that describe the current document as search terms .

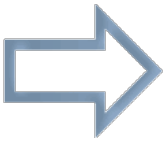
I browse a thesaurus to find other classification terms that are related to the current document.

I inspect the list of documents that cite the current document.

I follow references (footnotes) in the current document.

I look through the journal/ conference proceedings the current document was published in.

I look for other papers the author(s) has/have published.



What other search activities do you perform to find relevant documents?

8 Personal Information

Thank you very much for completing the survey :-)

Last but not least, please provide some (demographic) information (and, if you like, also some feedback on this survey).

Please specify your gender.

○ Male ○ Female

Please specify your age.

Please specify your highest academic degree.

None
Bachelor Degree
Diploma
Master Degree
Ph.D.
Professor

Please specify the discipline you are working in.

Other
Arts and Literature
Astronomy / Astrophysics
Biological Sciences
Business Administration
Chemistry
Computer and Information Science
Earth Sciences
Economics
Education
Electrical and Electronic Engineering
Engineering
Environmental Sciences
Humanities
Law
Linguistics
Management Science
Materials
Mathematics
Medicine
Philosophy
Physics
Psychology
Social Sciences
Sports and Recreation

Do you work in academia or industry?

☐ Academia ☐ Industry

Please specify your job title.

How often do you use Digital Libraries like ACM DL or Web of Science?

○ Very rarely ○ Rarely ○ Sometimes ○ Often ○ Very Often

How often do you use Google Scholar?

○ Very rarely ○ Rarely ○ Sometimes ○ Often ○ Very Often

How would you consider your experience in searching Digital Libraries?

None at all | Little | Moderate | High | Expert

Experience in searching Digital Libraries

Do you have any comments or feedback about the survey?

What did you miss in this survey? What did you like in this survey? What did you not like in this survey?

If you are interested in the outcome of this survey, please leave your email address and we will send you the results after thorough analysis and publication.

9 Endseite

Thank you very much for completing this survey and for helping me with my Ph.D. thesis. It is very much appreciated. You can close the window now.

Best regards,
Zeljko Carevic