

Overview of the NTCIR-4 WEB Navigational Retrieval Task 1

Keizo Oyama Koji Eguchi Haruko Ishikawa Akiko Aizawa
National Institute of Informatics (NII)
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
{oyama, eguchi, haruko, akiko}@nii.ac.jp

Abstract

This paper describes an overview of the Navigational Retrieval Task 1 that was conducted from 2002 to 2004 as a subtask of the WEB Task at the Fourth NTCIR Workshop. In the Task, we attempted to assess the retrieval effectiveness of Web search systems from a viewpoint of “Known Item Search” using a common data set, and built a re-usable test collection. 100-gigabyte Web document data constructed at NTCIR-3 and 300 topics prepared in this task were distributed to the participants and, in turn, 16 run results were submitted by 5 groups and some by organizers. Relevance judgments were performed on the documents pooled from the results, mainly from the viewpoint of representativeness of searched items given by the topics. Several kinds of evaluation measures were applied to the run results submitted by each participant. Brief analyses on system evaluation results and on topic difficulties are given.

Keywords: Web Information Retrieval, Evaluation Methods, Test Collections.

1 Introduction

This paper describes an overview of the Navigational Retrieval Task 1 (Subtask B) that was conducted from 2002 to 2004 as a subtask of the WEB Task at the Fourth NTCIR Workshop (NTCIR-4 WEB).

Several kinds of tasks can be associated with the term “Navigational Retrieval”. We selected “Known Item Search” as the first task to tackle with. Thus we call this subtask as “Navigational Retrieval Task 1.”

In the Task, we attempted to evaluate the retrieval effectiveness of Web search systems from a viewpoint of “Known Item Search.” It assumes such a circumstance that a searcher searches for one or a few “representative Web pages” of an item about which the searcher already knows.

We used 100-gigabyte Web document data (NW100G-01) constructed at NTCIR-3[1,2] and 300 topics created by 11 people in this task as a common data set. This data set was distributed to 9 participants, and, in turn, 16 search run results were submitted by 5 groups and some by organizers.

Relevance judgments were performed on the documents pooled from the results. Each run result submitted by the participants was evaluated using the relevance judgments with several kinds of measures. Consequently a re-usable test collection was built.

Similar tasks have been conducted in TREC. One of them is the “Home/Named Page Finding Task”[3] in the TREC-2003 Web Track. It was to evaluate system effectiveness to search for mixture of a home page and a named page by its name.

The “Known Item Search” is different in that one or a few search terms (not necessarily a name) are provided to specify a searched item, rather than a name. Therefore, there may be a few different relevant pages. Moreover, a relevant page may be a single page or a top page of a closely interlinked page group. It is considered to reflect the real search scene more appropriately.

In the following, we describe: the task definition in Section 2, the document set in Section 3, the search topics in Section 4, run conditions in Section 5, relevance assessment in Section 6, system evaluation in Section 7, summary of topic characteristics in Section 8 and conclusion in Section 9.

2 Task Definition

In the Navigational Retrieval Task 1 of the NTCIR-4 WEB Task, we tackled with system evaluation for “**Known Item Search**”. It assumes such a circumstance that a searcher searches for one or a few “representative Web pages” of a given known item. It is supposed that the searcher does not necessarily know about the Web page

itself while the searcher already knows about the item.

2.1 Search target items

An Item which can be a search target is a “known item” which represents a specific thing or a matter, or a collection of specific things or matters. Searches on unspecific things or matters or on unspecific information for information gathering purposes are not handled in this task.

Some search target items do not exist on the Web such as products, organizations, stores, persons, facilities, natural things, events, whereas other search target items do exist on the Web such as information providing sites, search services, data files and documents. Although general information cannot be a search target, information which has a specific content and is assumed to be provided in a “representative Web page” can be a search target.

2.2 Known items

An item is regarded as “known” when a searcher knows beforehand by some means that the search target item exists and can identify the item if search result pages are presented.

However, as in the following examples, the searcher may not be able to describe about the item exactly enough to specify it.

- Knows only an acronym
- Cannot express with a few words or phrases
- Remembers its features but has forgotten its name

On the other hand, the item’s “representative Web page” itself need not necessarily be “known” and may be any of the following three cases:

- Has seen the page and remembers its outline.
- Has actually seen the page but does not remember clearly what the page was like.
- Has never seen the page but take it for granted that such a page exists.

2.3 Representative Web pages

We suppose a “representative Web page” of “a known item” should satisfies the following conditions:

- (1) Provider of the Web page
It is necessary to be an organization or a person that is responsible for the “known item” or an organization or a

person that is generally appreciated as authoritative about the “known item”.

- (2) Content of the Web page

It is necessary to cover information that is strongly related in all aspects with the “known item” comprehensively as far as it is provided by the Web page provider. It is also necessary to include as little information as possible not directly related to the item. The “strongly related information” may either be described in the Web page or be linked from the Web page as it can be recognized explicitly.

3 Document Data Set

The document data set which is searched in is the same as the large document data set (NW100G-01) used in the NTCIR-3 Web retrieval task. It consists of text files of approximately 100-gigabytes in total and their metadata, which were crawled from http servers in “*.jp” domain from August to November in 2001.

Search results can include two types of documents: (1) Web pages whose page data are included in the NW100G-01; and (2) Web pages which have at least one link from Web pages in (1) but are not included in (1), and are actually fetched and stored in referral storage. Document identifiers of all these Web pages are given in a file provided with the distribution data.

Refer to the references [1,2] for more details.

4 Search topics

4.1 Creation and selection

We selected 300 topics for delivery as the result of discarding similar ones and inappropriate ones from several view points out of 456 topics that were created by 11 topic creators. Most of them are undergraduate / graduate students of several universities.

The topics were created and selected with the following procedures:

- (1) Each topic creator recollects a natural search target item in relation with hobby, study, work, daily life, and so on,
- (2) Imagines corresponding “representative Web page”, and
- (3) Writes them down in a free format.
- (4) Organizers select ones

appropriate as the known item search.

- (5) Each topic creator describes it in a given format as a search topic.

When making a search topic, it was not checked if its relevant documents exist in the document data set.

However, since the document data set is collected between August and November in 2001, items whose representative Web pages were considered not to have existed at that time were excluded from the search topics.

4.2 Format and elements of search topics

A search topic is described in the tagged format shown below. The language is Japanese but English translation is also available mainly for publication purpose.

Tag structure

```
<TOPIC>
<NUM>Topic number</NUM>
<TYPE>Type code</TYPE>
<CATEGORY>Category code
</CATEGORY>
<TITLE>Search terms</TITLE>
<DESC>Search description sentence</DESC>
<NARR>
  <TERM>Explanation of terms (optional)
  </TERM>
  <BACK>Explanation of back ground
  (optional) </BACK>
  <RELE>Relevance criteria (optional)
  </RELE>
</NARR>
<USER SPECIALTY="Knowledge level
code">Attributes of searcher</USER>
</TOPIC>
```

The elements corresponding to the tag names are as follows:

- (1) TOPIC: One search topic
- (2) NUM: Topic number used as topic id.
- (3) TYPE: Topic type
 - A code defined as follows:
 - 1: Single search term specifies the known item
 - 2: Combination of search terms specifies the known item
 - 3: Single search term or combination of

search terms represents the known item but cannot specify it

- (4) CATEGORY: Category of the known item
 - One or more codes defined as follows:
 - A: Products / services
 - B: Companies / organizations (including shops and administrative organs)
 - C: Persons
 - D: Facilities (including public and private)
 - E: Sights and historic spots, and natural things (including parks, etc.)
 - F: Information resources (including information sites, data files, etc.)
 - G: Events
 - Z: Others
- (5) TITLE: Search terms
 - Search terms supposed to be entered to a search engine regarding the information needs; up to three terms in the order of importance
- (6) DESC: Search description
 - One Japanese sentence briefly describing the information need; should be conceptually consistent with TITLE
- (7) NARR: Narrative of the information needs (All sub-elements are optional)
- (8) NARR/TERM: Explanation of terms
 - Japanese sentences describing definition of meanings and explaining related terms regarding terms in TITLE and DESC when they have ambiguity or they are not popular
- (9) NARR/BACK: Explanation of back ground
 - Japanese sentences explaining back ground of the information needs and the motivation
- (10) NARR/RELE: Relevance criteria
 - Japanese sentences explaining relevance criteria on the item and the pages when they are not clear just with TITLE and DESC
- (11) USER: Searcher's attributes
 - Position, sex, and experience years of Web search
- (12) USER/@SPECIALITY: Searcher's knowledge level on the searched item
 - A code defined as follows:
 - A: Knows the item in detail.
 - B: Knows the outline of the item.
 - C: Knows the item to the extent the item can be identified from others.
 - D: Knows existence of the item but very little about it.

5 Run conditions

5.1 Search run execution

Participants can use the following combinations of topic elements for the search run execution. The other topic elements must not be used.

- (1) TITLE only (mandatory)
- (2) Any combination of TITLE, DESC, and NARR/BACK
- (3) Any combination of TYPE and CATEGORY added to (1)
- (4) Any combination of TYPE and CATEGORY added to (2)

When submitting run results using (3) and (4), it is strongly recommended to also submit run results without using TYPE or CATEGORY.

A participant may submit up to four run results for each of four topic element combinations, namely, up to 16 run results in all.

Both automatic and interactive processing modes are permitted. The run is regarded as interactive when a human has a hand in any way during search topic processing and/or search execution; otherwise it is regarded as automatic.

5.2 Submission of retrieval results

A participant was required to submit run results and a system description form.

The run results should be in a given format including a query number, a document id, a score, and a run id on each line. The number of retrieved documents should be no more than 100 for each topic for each run.

The system description form includes a concise description of each run including items among others as follows:

Topic Part

The part of the topics used

Query Method

Automatic or interactive

Query Unit

Unit of query, e.g., character, word, phrase

Query Expansion

Techniques used to expand queries

Link Information

Use of link information

Anchor

Use of anchor text, etc.

IR Model

IR model

Ranking

Ranking factor for calculating scores

Index Unit

Unit of index, use of tag/link structure in indexing, etc.

Index Technique

Techniques used to process index terms

Index Structure

Index structure

Filtering

Filtering method for extracting useful pages or for discarding unnecessary pages

Resource

External resources used for indexing, filtering, or searching, other than the data provided by the organizers

These items will be used in the further analysis.

6 Relevance assessment

Five participants submitted 16 run results. Each run result include up to 100 documents for each topic. Organizers added 68 run results in order to find relevant documents comprehensively so that the test collection becomes re-usable.

Pooling was applied to the run results for the relevance assessment in this task. However, we requested assessors to find relevant documents as far as possible by following possible hyperlinks and by searching probable URLs. Consequently, any document included in the document data set potentially becomes the object of relevance assessment.

Although we tried that relevance assessment of each topic is done by the topic creator, about half of the topics were actually done so.

On the completion of the assessment, all of the relevance judgment results were inspected by the organizers, and a few topics that did not meet the judgment criteria set by the organizers were carefully re-judged.

6.1 Pooling

Each pool was made by extracting top N (=10) highly ranked documents from every run result of each topic.

The sequence of assessment is decided as follows: (1) add random number to each of all documents in the pool, (2) sort the documents with the rank as the primary key and the random number as the secondary key, and (3) remove duplicates leaving one which appears first. With this method, highly ranked documents are assessed first without losing fairness among runs, resulting good assessment efficiency.

The value 10 for N was selected because many search engines return top 10 search results as the first response, and because it will include more than 80 percent of relevant documents included in all the submitted documents (up to 100 for each run) according to rough sample assessment.

We expect most of the relevant documents not included in the pool have eventually been assessed by following possible hyperlinks. We will verify this afterwards by doing additional assessment.

6.2 Judgment bases

The assessor should use not only text in the document but also clues which the assessor usually uses in Web browsing and which usual searchers are considered to use, e.g. page titles, host names, URL patterns, and various kinds of HTML tags, as the judgment bases.

Concerning judgment of frame set pages and pages that automatically jump to other URLs, the assessor refer to their link target pages as far as they are included in the document data set and take them into the judgment bases.

Moreover, in order to identify the provider of the page and for other purposes, the assessor may refer to the current real Web page of the same URL or the related Web pages.

6.3 Relevance judgment

Relevance of each document to the search topic was judged into one of the following levels by absolute evaluation:

A: Relevant

A representative page appropriate for the searched item satisfying the retrieval needs

B: Partially relevant

A page partially satisfying the retrieval needs; pages as follows fit to this:

- ◆ A representative page of an item having an upper or lower concept of the searched item; an easy-to-find hyperlink to the relevant document should be provided in the page
- ◆ A page that can be regarded as a substitute for the representative page of the searched item

D: Non-relevant
Otherwise.

6.4 Additional judgment

Aspects which should be taken into account on system evaluation other than relevance are judged.

(1) Undistinguishability

A non-relevant page that satisfies all the following conditions is judged as undistinguishable.

- The page is a representative page of an item different from and not directly related to the searched item (hereinafter, different item).
- The different item cannot be excluded semantically only by TITLE and DESC.

We define undistinguishability as follows according to how generally well-known the different item is when compared to the searched item:

3: The different item is more well-known than the searched item.

2: The different item is as well-known as the searched item.

1: The different item is less well-known than the searched item.

0: Not undistinguishable.

(2) Duplicate pages

When there are relevant or partially relevant pages which have identical entity or which are corresponding pages within mirror sites judging from their contents, URL's, etc., these pages are judged as duplicate pages. Even if content text are completely same, pages which are considered to have different link target pages or images are not deemed to be duplicate pages.

7 System evaluation

7.1 Summary of participation

Five groups, listed below in alphabetical order of affiliations, submitted their completed run results, with the organizers also submitting the results from their own search systems along with those of the participants in an attempt to improve the comprehensiveness of the pool. Their group ID's are shown in parentheses.

- Beijing Center for Japanese Studies and Tokai University Junior College (W3SJP2003)
- National Institute of Informatics, University of Tokyo, and KYA group Corporation (R2D2)

- NEC Corporation (anonymous¹)
- Osaka Kyoiku University (OKSAT)
- Toyohashi University of Technology, University of Tsukuba, and University of Nagoya (TKB)

The individual participating groups pursued various objectives. We summarize them as follows (listed in alphabetical order of group IDs):

K3100: Experimented with a retrieval method using site anchor text for indexing Web documents that are pointed to by them, and ranking with two relevancy measures: reference consistency and specificity of word combination.

OKSAT: Experimented with gram-based indices using textual contents of Web documents. The retrieval module was based on a probabilistic model.

ORGREF: Performed by the organizers to expand the pool using four types of search systems: (1) baseline systems using Boolean-type search systems and ranking by tf-idf; (2) a method to expand the baseline retrieval sets using one-hop forward link analysis; (3) same as the above except using one-hop backward link analysis; (4) a method using expanded anchor text for indexing Web documents that are pointed to by them.

R2D2: Experimented with a system based on the Relevance-based Superimposition Model in combination with two ranking methods using URL heuristics and link analysis.

TKB: Experimented with a search system on four conditions, i.e. combination of with/without PageRank and with/without pseudo-relevance feedback.

W3SJP2003²: Experiment with Microsoft IIS Index Server.

Summaries of the run result submissions are shown in **Table 1** (shown at the end of this paper).

¹ The group ID is not shown because they participated as anonymous type.

² Although the group submitted their run result, they could not submit their Working Note paper because of unavoidable circumstances.

7.2 System evaluation methods

We evaluated each run result on two different document sets and on two different relevance levels described below, resulting four combinations.

Document data sets

Although we defined the document data set as in Section 3, many run results submitted by the participants did not include the part (2) because their systems cannot search the part. Therefore, we decided to evaluate each run on two different document data sets as follows:

(DS-1) Document data set with and without delivered page data

The document data set which is defined in Section 3. It includes not only documents whose page data are delivered to the participants, but also documents which have only in-links from one or more of them and are fetched and stored by the organizers.

(DS-2) Document set with delivered page data

An additional document data set for comparison. It includes only the part (1) described in Section 3, i.e., only documents whose page data are delivered to the participants.

Relevance levels

Based on the relevance judgment described in Section 6.3, we evaluated each run on two relevance levels as follows:

(RL-1) Rigid:

Documents with assessment A ('relevant') are regarded as relevant documents.

(RL-2) Relaxed:

Documents with assessment A ('relevant') and B ('partially relevant') are regarded as relevant documents.

We delivered 300 topics to the participants and assessed them all³. For each document set, we selected such topics satisfying both of the

³ At the time of NTIR-4 Working Notes, assessment results of only 144 topics were available. Furthermore, those of a few topics did not meet the judgment criteria set by the organizers and were re-assessed thereafter. Therefore, system evaluation results in this paper may slightly differ from that of the Working Notes paper.

following conditions¹:

- condition 1:** at least one relevant document at the 'rigid' relevance level was found or was considered to exist in the Web space when the documents were crawled; and
- condition 2:** at least one relevant document at the 'relaxed' relevance level was found in the document set.

Consequently, we used the following two topic sets for system evaluation:

- (TS-1):** 218 topics for (DS-1); and
- (TS-2):** 197 topics for (DS-2).

As the evaluation measures, we calculated DCG (Discounted Cumulative Gain) and WRR (Weighted Reciprocal Rank)[1,2] at top ranked 10 document level for each of above mentioned four combinations, averaging over (TS-1) and (TS-2) respectively.

Gains used in the DCG calculation are: $(G_A, G_B) = (3, 0)$ for (RL-1), and $(G_A, G_B) = (3, 2)$ for (RL-2).

Parameters used in the WRR calculation are such that the measure becomes identical with MRR (Mean Reciprocal Rank).

Although many topics respectively have multiple relevant documents, most of them are redundant, i.e., either duplicated web pages or closely linked web pages. Therefore, for such a group of pages, the top ranked relevant document has importance and the others have little.

Because duplication and link relation are not considered in this evaluation, appropriateness of DCG values as the system effectiveness is left to be investigated. However, because only the first retrieved relevant document is used in MRR, the appropriateness is the same regardless of considering duplication or link relation.

System evaluation considering the redundancy and using the full topics will be done later.

An evaluation measure taking costs of inputting search terms and browsing retrieval result documents into account will be applied in the future study.

Moreover, in case an undistinguishable document is included in the run results, some compensation may be necessary. We will further investigate the evaluation methods regarding their effects.

¹ For Working Notes, we selected such topics that at least one relevant document at the 'rigid' relevance level was found in the submitted run results.

7.3 Summary of evaluation results

We computed the effectiveness of individual run results shown in Section 7.1 based on the evaluation method described in Section 7.2, on four combinations of document data sets and relevance levels.

The evaluation results are shown in **Table 2** (shown at the end of this paper). The Run IDs are classified into three groups: runs using anchor information, runs using link information but anchor information, and runs using no link information nor anchor information; and arranged in the descending order of the WRR at the combination of (DS-1) and (RL-1) in each group.

Figure 1, 2, 3 and **4** (shown at the end of this paper) are graphs plotting WRR values, and **Figure 5, 6, 7** and **8** (ibid.) are graphs plotting DCG values. The orders of the Run IDs are the same as Table 2.

As seen in the Figure 1 through 8, in general, runs showing high performance, e.g. K3100-tt-02 and ORGREF-AT0-P1, utilize anchor text information in certain ways, although their retrieval methods differ. Besides these runs, several runs utilizing link information performed well. Runs using content information only stayed at relatively low performance.

Graphs of WRR (Figures 1 through 4) and of DCG (Figures 5 through 8) show roughly same tendencies, though each system shows slight differences. If duplicated documents and closely linked documents are given penalties in DCG calculation, the differences are expected to be even less.

Differences between graphs on (DS-1) and (DS-2) (e.g., Figures 1 and 3, Figures 2 and 4) are mainly caused by differences of system concepts in handling hyperlinks to web pages yet to be fetched.

It may be understood that graphs on (RL-1) (i.e., Figure 1, 3, 5 and 7) show performance of finding exact web pages while graphs on (RL-2) (i.e., Figures 2, 4, 6 and 8) show performance of finding allowable web pages.

Ratios of WRR on (RL-2) to WRR on (RL-1) for well performed systems based on anchor text information distribute between 1.2 and 1.3, and those based on link information distribute between 1.2 and 1.7. On the other hand, ratios of DCG on (RL-2) to DCG on (RL-1) for well performed systems based on anchor text information distribute between 1.2 and 1.3 (i.e., about the same as that of WRR), and those based on link information distribute between 1.4 and 1.9 (i.e., slightly higher than that of WRR). This may

suggests that link-based systems tend to find more allowable web pages at higher ranks than anchor-based systems do.

For systems based on content information only, the ratios of both WRR and DCG distribute between 1.3 and 2.5 varying from system to system, and no specific tendency over systems can be seen.

Figures 9, 10, 11 and 12 (shown at the end of this paper) show, on the four combinations respectively, graphs of cumulative numbers of topics whose relevant documents were retrieved by several selected runs. Selected runs were those recommended by the participants. (These also roughly represent the well performed runs from different search methods.)

We can see a tendency that curves of runs based on anchor information rise rapidly within rank 10 and nearly level thereafter, while those based on content information only rise gradually over all rank range, and those based on link

information perform intermediately.

Comparing these graphs with those of Figures 1 through 4, how the form of the curves affects WRR values can be seen. For instance, comparing K3100-tt-02 with ORGREF-AT40-P1 on Figures 10 and 12, differences in retrieved topics below rank 5 (i.e., > 5) make little effects on WRR values in Figure 2, while differences above rank 5 (i.e., ≤ 5) make relatively large effects in Figure 4.

8 Summary of topic characteristics

As described in section 4.2, topics that were used for the task are characterized by three types (TYPE 1 to 3) and 8 categories (CATEGORY A to Z). The number of topics and their proportion in percentage (in bracket) for each type and category in all 218 topics are listed in the second column of Table 3 and 4 respectively.

We investigated the distribution of types and categories of topics in ‘easy topics’ and in

Table 3 Proportion of ‘TYPE’ in topics

TYPE	ALL topics	Easy topics	Difficult topics
1	150 (68.8%)	8 (80.0%)	10 (37.0%)
2	61 (28.0%)	2 (20.0%)	15 (55.6%)
3	7 (3.2%)	0 (0%)	2 (7.4%)
Total	218 (100%)	10 (100%)	27 (100%)

Table 4 Proportion of ‘CATEGORY’ in topics

CATEGORY	ALL topics			Easy topics	Difficult topics
	ALL	TYPE 1	TYPE 2		
A	34 (15.6%)	11 (32.4*%)	23 (67.6*%)	0 (0%)	5 (18.5%)
B	59 (27.1%)	50 (84.7*%)	9 (15.3*%)	2 (20.0%)	2 (7.4%)
C	26 (11.9%)	21 (80.8*%)	4 (15.4*%)	2 (20.0%)	4 (14.8%)
D	22 (10.1%)	18 (81.8*%)	4 (18.2*%)	2 (20.0%)	1 (3.7%)
E	19 (8.7%)	16 (84.2*%)	2 (10.5*%)	1 (10.0%)	5 (18.5%)
F	23 (10.6%)	10 (43.5*%)	10 (43.5*%)	1 (10.0%)	6 (22.2%)
G	9 (4.1%)	6 (66.7*%)	3 (33.3*%)	0 (0%)	1 (3.7%)
Z	3 (1.4%)	2 (66.7*%)	1 (33.3*%)	0 (0%)	0 (0%)
Combo	23 (10.5%)	16 (69.6*%)	6 (26.1*%)	2 (20.0%)	3 (11.1%)
Total	218 (100%)	--	--	10 (100%)	27(99.9%)

* The values are calculated in proportion to the number of topics in each CATEGORY.

‘difficult topics’. The easy topics are determined by the magnitude of averaged reciprocal rank calculated over all runs (RS-1) or only over the selected runs (RS-2) for each topic, and we used only the first 10 of them. The difficult topics are those for which no runs could find the relevant web pages.

The investigations were conducted for the eight cases, i.e. the combination of (RS-1,2), (DS-1,2) and (RL-1,2). The values listed in the tables are for the case of (DS-1), (RL-1) and (RS-2). However similar trends were observed in other seven cases, both in the distribution of types and categories.

The third and the fourth columns of the **Table 3** indicate the number of topics and its proportion in the total number of the ‘easy’ or the ‘difficult’ topics in percentage (in bracket). The first 10 easy topics were dominated by TYPE 1 while more than 50% of the difficult topics were TYPE 2. It is to be noted that the quarter of all TYPE 2 and 3 topics (the first column) belong to ‘difficult topics’.

On the last two columns of the **Table 4**, the number and its proportion in the total number of ‘easy’ or ‘difficult’ topics in percentage (in bracket) for each CATEGORY are listed. The number and proportion in percentage of TYPE 1 and 2 in each CATEGORY are also listed in the third and the fourth columns of the same table.

The distributions of categories A to Z are distorted from those of 218 topics in the ‘easy’ and the ‘difficult’ topics. Higher concentration of TYPE 1 and 2 in the easy and the difficult topics respectively only partly explains the distortion and requires further investigation.

Some of other features that were observed in the relevant documents of the ‘easy’ and ‘difficult’ topics in the results of the case (DS-1), (RL-1) and (RS-2) are as follows. The relevant documents of the easy topics are far more frequently linked from other web-pages than those of the difficult topics. They are more likely to be higher in the hierarchy of the web site and contain exact query term(s) in their text. Frameset pages did not appear as relevant documents of the easy topics while approximately 20% relevant documents in the difficult topics were frameset pages. The further investigation on the relationship between topic types, types of relevant documents and the search methods are currently conducted by the authors.

9 Conclusion

In this paper, an overview of the Navigational Retrieval Task 1, a subtask of the WEB Task at the Fourth NTCIR Workshop was described. It aimed at evaluating web search engine systems for retrieving representative web pages of known items.

We used a 100-gigabyte Web document data set, NW100G-01, constructed at the Third NTCIR Workshop.

300 topics were delivered to the participants. Each run result submitted by participants included up to 100 documents per topic. Pooling was done with top-ranked 10 documents of every run results for every topic. Relevance was assessed not only on the documents in the pool but also on the documents hyperlinked from ones in the pool. All topics were assessed and topics including at least one relevant document were used for the evaluation.

The run results submitted by the participants were evaluated with DCG and WRR. Classifying systems to three groups, it seems that a group utilizing anchor text and link information performed best, another group utilizing link information but anchor text performed second, and the other group only using content information performed worst.

We will also attempt to use other evaluation measures than WRR or DCG, which can reflect users’ intuition more appropriately. Topic by topic analysis is also the future study.

Since the number of the participating systems is not sufficient, and since the document data size is not large enough compared to the actual operating web search engines, we cannot conclude merits and demerits of each method at this point. Therefore, we are planning to conduct the Navigational Retrieval Task 2 continuously at the Fifth NTCIR Workshop with larger document data and, hopefully, with participation of more systems.

Acknowledgements

This work was partially supported by the Grants-in-Aid for Scientific Research on Priority Areas of “Informatics” (#13224087) from the Ministry of Education, Culture, Sports, Science and Technology, Japan. We greatly appreciate the efforts of all the participants of the Navigational Retrieval Task 1 of the WEB Task at the Fourth NTCIR Workshop. We also appreciate the helpful advice of Professors Jun Adachi and Noriko Kando, National Institute of Informatics.

References

- [1] Koji Eguchi, Keizo Oyama, Emi Ishida, Noriko Kando, and Kazuko Kuriyama: "Overview of the Web Retrieval Task at the Third NTCIR Workshop", *NII Technical Report*, No.NII-2003-002E (Jan. 2003) (available from: <<http://research.nii.ac.jp/TechReports/03-002E.html>>)
- [2] K. Eguchi, K. Oyama, E. Ishida, N. Kando, and K. Kuriyama, "Evaluation Methods for Web Retrieval Tasks Considering Hyperlink Structure", *IEICE Transactions on Information and Systems*, Vol.E86-D, No.9, pp.1804-1813 (2003) (available from: <<http://search.ieice.org/2003/files/e000d09.htm#e86-d,9,1804>>)
- [3] N. Craswell, and D. Hawking, "Overview of the TREC 2003 Web Track", in *Proc. 2003 Text Retrieval Conference: TREC 2003*, Gaithersburg, Maryland, November 18-21, 2003. (available from: <<http://trec.nist.gov/pubs/trec12/papers/WEB.OVERVIEW.pdf>>)

Table 1. Summaries of run result submissions

RunID: Indicates the indication codes of the system run results. Each one starts with the group ID.

TopicPart: Indicates the part of the topic used. The characters 'T', 'D', and 'B' respectively indicate TITLE, DESC, and BACK.

QMethod: Indicates 'automatic' or 'interactive'. 'Automatic' indicates a run without any human intervention during query processing and search; 'interactive' indicates a run other than 'automatic'.

QExpan: Indicates if query expansion is used.

ContInfo: Indicates if fulltext content information is used for indexing documents.

LinkInfo: Indicates if link information is used for indexing documents.

AnchorInfo: Indicates if anchor text information is used for indexing documents.

RunID	TopicPart	QMethod	QExpan	ContInfo	LinkInfo	AnchorInfo
K3100-ds-01	D	automatic	no	no	yes	yes
K3100-ds-02	D	automatic	no	yes	yes	yes
K3100-ds-03	D	automatic	no	no	yes	yes
K3100-ds-04	D	automatic	no	no	yes	yes
K3100-tt-01	T	automatic	no	no	yes	yes
K3100-tt-02	T	automatic	no	yes	yes	yes
K3100-tt-03	T	automatic	no	no	yes	yes
K3100-tt-04	T	automatic	no	no	yes	yes
oksat-bf01	T	interactive	yes	yes	no	no
r2d2-a1	T	automatic	no	yes	yes	no
r2d2-a2	T	automatic	no	yes	no	no
TKB-01	T	automatic	yes	yes	yes	no
TKB-02	T	automatic	yes	yes	no	no
TKB-03	T	automatic	no	yes	yes	no
TKB-04	T	automatic	no	yes	no	no
W3SJP2003-001	TDB	interactive	yes	yes	yes	-
ORGREF-NMZ-AND	T	automatic	no	yes	no	no
ORGREF-NMZ-OR	T	automatic	no	yes	no	no
ORGREF-OT-D	T	automatic	no	yes	no	no
ORGREF-OT-DT	T	automatic	no	yes	no	no
ORGREF-AT0-P1	T	automatic	no	no	yes	yes
ORGREF-AT0-P2	T	automatic	no	no	yes	yes
ORGREF-AT0-P3	T	automatic	no	no	yes	yes
ORGREF-AT0-P4	T	automatic	no	no	yes	yes
ORGREF-AT0-P5	T	automatic	no	no	yes	yes
ORGREF-AT0-P6	T	automatic	no	no	yes	yes
ORGREF-AT0-S1	T	automatic	no	no	yes	yes
ORGREF-AT0-S2	T	automatic	no	no	yes	yes
ORGREF-AT200-P1	T	automatic	no	no	yes	yes
ORGREF-AT200-P2	T	automatic	no	no	yes	yes
ORGREF-AT200-P3	T	automatic	no	no	yes	yes
ORGREF-AT200-P4	T	automatic	no	no	yes	yes
ORGREF-AT200-P5	T	automatic	no	no	yes	yes
ORGREF-AT200-P6	T	automatic	no	no	yes	yes
ORGREF-AT200-S1	T	automatic	no	no	yes	yes
ORGREF-AT200-S2	T	automatic	no	no	yes	yes
ORGREF-AT40-P1	T	automatic	no	no	yes	yes
ORGREF-AT40-P2	T	automatic	no	no	yes	yes
ORGREF-AT40-P3	T	automatic	no	no	yes	yes

RunID	TopicPart	QMethod	QExpan	ContInfo	LinkInfo	AnchorInfo
ORGREF-AT40-P4	T	automatic	no	no	yes	yes
ORGREF-AT40-P5	T	automatic	no	no	yes	yes
ORGREF-AT40-P6	T	automatic	no	no	yes	yes
ORGREF-AT40-S1	T	automatic	no	no	yes	yes
ORGREF-AT40-S2	T	automatic	no	no	yes	yes
ORGREF-NMZ-AND-LB1	T	automatic	no	yes	yes	no
ORGREF-NMZ-AND-LB2	T	automatic	no	yes	yes	no
ORGREF-NMZ-AND-LB3	T	automatic	no	yes	yes	no
ORGREF-NMZ-AND-LB4	T	automatic	no	yes	yes	no
ORGREF-NMZ-AND-LB5	T	automatic	no	yes	yes	no
ORGREF-NMZ-AND-LF1	T	automatic	no	yes	yes	no
ORGREF-NMZ-AND-LF2	T	automatic	no	yes	yes	no
ORGREF-NMZ-AND-LF3	T	automatic	no	yes	yes	no
ORGREF-NMZ-AND-LF4	T	automatic	no	yes	yes	no
ORGREF-NMZ-AND-LF5	T	automatic	no	yes	yes	no
ORGREF-NMZ-OR-LB1	T	automatic	no	yes	yes	no
ORGREF-NMZ-OR-LB2	T	automatic	no	yes	yes	no
ORGREF-NMZ-OR-LB3	T	automatic	no	yes	yes	no
ORGREF-NMZ-OR-LB4	T	automatic	no	yes	yes	no
ORGREF-NMZ-OR-LB5	T	automatic	no	yes	yes	no
ORGREF-NMZ-OR-LF1	T	automatic	no	yes	yes	no
ORGREF-NMZ-OR-LF2	T	automatic	no	yes	yes	no
ORGREF-NMZ-OR-LF3	T	automatic	no	yes	yes	no
ORGREF-NMZ-OR-LF4	T	automatic	no	yes	yes	no
ORGREF-NMZ-OR-LF5	T	automatic	no	yes	yes	no
ORGREF-OT-D-LB1	T	automatic	no	yes	yes	no
ORGREF-OT-D-LB2	T	automatic	no	yes	yes	no
ORGREF-OT-D-LB3	T	automatic	no	yes	yes	no
ORGREF-OT-D-LB4	T	automatic	no	yes	yes	no
ORGREF-OT-D-LB5	T	automatic	no	yes	yes	no
ORGREF-OT-D-LF1	T	automatic	no	yes	yes	no
ORGREF-OT-D-LF2	T	automatic	no	yes	yes	no
ORGREF-OT-D-LF3	T	automatic	no	yes	yes	no
ORGREF-OT-D-LF4	T	automatic	no	yes	yes	no
ORGREF-OT-D-LF5	T	automatic	no	yes	yes	no
ORGREF-OT-DT-LB1	T	automatic	no	yes	yes	no
ORGREF-OT-DT-LB2	T	automatic	no	yes	yes	no
ORGREF-OT-DT-LB3	T	automatic	no	yes	yes	no
ORGREF-OT-DT-LB4	T	automatic	no	yes	yes	no
ORGREF-OT-DT-LB5	T	automatic	no	yes	yes	no
ORGREF-OT-DT-LF1	T	automatic	no	yes	yes	no
ORGREF-OT-DT-LF2	T	automatic	no	yes	yes	no
ORGREF-OT-DT-LF3	T	automatic	no	yes	yes	no
ORGREF-OT-DT-LF4	T	automatic	no	yes	yes	no
ORGREF-OT-DT-LF5	T	automatic	no	yes	yes	no

Table 2. Summary of evaluation results

RunID: Indicates the indication codes of the system run results. Each one starts with the group ID.

ContInfo: Indicates if fulltext content information is used for indexing documents.

LinkInfo: Indicates if link information is used for indexing documents.

AnchorInfo: Indicates if anchor text information is used for indexing documents.

WRR: Indicates weighted reciprocal rank at top-10 document level.

DCG: Indicates discounted cumulative gain at top-10 document level.

(DS-1): Indicates that the search target document set includes documents with & without delivered page data.

(DS-2): Indicates that the search target document set includes only documents with delivered page data.

Rigid: Indicates rigid relevance level.

Relaxed: Indicates relaxed relevance level.

RunID	Cont Info	Link Info	Anchor Info	WRR				DCG			
				(DS-1)		(DS-2)		(DS-1)		(DS-2)	
				Rigid	Relaxed	Rigid	Relaxed	Rigid	Relaxed	Rigid	Relaxed
ORGREF-AT0-P1	no	yes	yes	0.4521	0.5123	0.4272	0.4911	2.0396	2.4362	1.7648	2.1857
ORGREF-AT0-P2	no	yes	yes	0.4521	0.5120	0.4272	0.4907	2.0387	2.4329	1.7648	2.1831
ORGREF-AT40-P2	no	yes	yes	0.4436	0.5254	0.4111	0.4967	1.9027	2.3254	1.6572	2.1215
ORGREF-AT40-P1	no	yes	yes	0.4426	0.5246	0.4126	0.4985	1.8996	2.3188	1.6680	2.1329
K3100-tt-02	yes	yes	yes	0.4181	0.5124	0.4626	0.5670	1.8026	2.3811	1.9948	2.6350
K3100-tt-01	no	yes	yes	0.3806	0.4494	0.3372	0.4038	1.6401	1.9681	1.3943	1.7395
K3100-tt-04	no	yes	yes	0.3472	0.4156	0.3199	0.3794	1.6948	2.1063	1.4446	1.8297
ORGREF-AT200-P2	no	yes	yes	0.3423	0.4301	0.3205	0.4106	1.4497	1.8692	1.3108	1.7436
ORGREF-AT200-P1	no	yes	yes	0.3412	0.4272	0.3209	0.4120	1.4393	1.8374	1.3148	1.7313
K3100-ds-02	yes	yes	yes	0.3279	0.4097	0.3629	0.4534	1.4056	1.8566	1.5554	2.0545
K3100-tt-03	no	yes	yes	0.3220	0.3784	0.2994	0.3539	1.5925	1.9240	1.3379	1.6675
K3100-ds-04	no	yes	yes	0.2946	0.3545	0.2755	0.3301	1.4148	1.7718	1.2079	1.5342
K3100-ds-03	no	yes	yes	0.2480	0.2966	0.2335	0.2844	1.2073	1.4835	1.0325	1.3077
K3100-ds-01	no	yes	yes	0.2025	0.2548	0.1921	0.2454	0.9399	1.1606	0.8497	1.0860
ORGREF-AT0-P4	no	yes	yes	0.1963	0.2559	0.2000	0.2639	0.9946	1.3128	0.9375	1.2670
ORGREF-AT0-P6	no	yes	yes	0.1963	0.2559	0.2000	0.2639	0.9946	1.3128	0.9375	1.2670
ORGREF-AT0-S2	no	yes	yes	0.1852	0.2463	0.1921	0.2595	0.9380	1.2419	0.9060	1.2157
ORGREF-AT40-P4	no	yes	yes	0.1379	0.1969	0.1339	0.2039	0.6848	0.9898	0.6454	1.0148
ORGREF-AT40-P6	no	yes	yes	0.1379	0.1969	0.1339	0.2039	0.6848	0.9898	0.6454	1.0148
ORGREF-AT40-S2	no	yes	yes	0.1198	0.1678	0.1177	0.1747	0.5783	0.8251	0.5845	0.8692
ORGREF-AT0-P3	no	yes	yes	0.1126	0.1541	0.1131	0.1580	0.5652	0.7777	0.5743	0.7858
ORGREF-AT0-P5	no	yes	yes	0.1126	0.1541	0.1131	0.1580	0.5652	0.7719	0.5743	0.7794
ORGREF-AT0-S1	no	yes	yes	0.0914	0.1226	0.0912	0.1316	0.4641	0.6240	0.4691	0.6409
ORGREF-AT40-P3	no	yes	yes	0.0692	0.0979	0.0663	0.1031	0.3769	0.5378	0.3527	0.5515
ORGREF-AT40-P5	no	yes	yes	0.0692	0.0918	0.0663	0.0955	0.3769	0.5228	0.3527	0.5281
ORGREF-AT200-P4	no	yes	yes	0.0518	0.0768	0.0276	0.0607	0.2687	0.3862	0.1697	0.3190
ORGREF-AT200-P6	no	yes	yes	0.0518	0.0768	0.0276	0.0607	0.2687	0.3862	0.1697	0.3190
ORGREF-AT40-S1	no	yes	yes	0.0379	0.0519	0.0338	0.0565	0.2206	0.3015	0.2074	0.3150
ORGREF-AT200-P3	no	yes	yes	0.0339	0.0494	0.0105	0.0336	0.1837	0.2696	0.0681	0.1776
ORGREF-AT200-P5	no	yes	yes	0.0339	0.0494	0.0105	0.0336	0.1837	0.2663	0.0681	0.1732
ORGREF-AT200-S2	no	yes	yes	0.0207	0.0304	0.0119	0.0213	0.1168	0.1593	0.0865	0.1330
ORGREF-AT200-S1	no	yes	yes	0.0182	0.0228	0.0041	0.0131	0.0973	0.1242	0.0295	0.0686
ORGREF-OT-D-LF2	yes	yes	no	0.2684	0.3422	0.2340	0.3181	1.1557	1.6230	0.9713	1.4639
ORGREF-OT-D-LF1	yes	yes	no	0.2640	0.3261	0.2265	0.3071	1.1466	1.6060	0.9071	1.4161
ORGREF-NMZ-AND-LF1	yes	yes	no	0.2615	0.3146	0.2269	0.2943	1.0487	1.4667	0.8760	1.3339
ORGREF-OT-DT-LF1	yes	yes	no	0.2563	0.3222	0.2275	0.3145	1.1338	1.6483	0.9124	1.5042
ORGREF-NMZ-AND-LF2	yes	yes	no	0.2533	0.3087	0.2223	0.2922	0.9963	1.3894	0.8741	1.3006
ORGREF-OT-D-LF3	yes	yes	no	0.2308	0.2989	0.2131	0.3000	1.1772	1.6039	1.0249	1.4999

RunID	Cont Info	Link Info	Anchor Info	WRR				DCG			
				(DS-1)		(DS-2)		(DS-1)		(DS-2)	
				Rigid	Relaxed	Rigid	Relaxed	Rigid	Relaxed	Rigid	Relaxed
ORGREF-OT-DT-LF2	yes	yes	no	0.2196	0.2994	0.1943	0.2943	1.0235	1.5830	0.8546	1.4934
ORGREF-NMZ-OR-LF1	yes	yes	no	0.2171	0.2577	0.1985	0.2469	0.9105	1.2556	0.7980	1.1647
ORGREF-NMZ-AND-LF3	yes	yes	no	0.2166	0.2708	0.2094	0.2742	1.0347	1.4573	0.9625	1.4468
ORGREF-OT-DT-LF3	yes	yes	no	0.2156	0.2912	0.1892	0.2892	1.0358	1.5529	0.8528	1.4440
ORGREF-NMZ-OR-LF2	yes	yes	no	0.2138	0.2630	0.2004	0.2610	0.8569	1.2261	0.8020	1.1973
ORGREF-OT-D-LF4	yes	yes	no	0.1976	0.2618	0.1576	0.2393	1.0035	1.4163	0.7672	1.2382
ORGREF-OT-DT-LF4	yes	yes	no	0.1941	0.2615	0.1716	0.2659	0.9625	1.3801	0.7567	1.2479
ORGREF-NMZ-OR-LF3	yes	yes	no	0.1882	0.2239	0.1911	0.2393	0.9251	1.2512	0.8876	1.2735
ORGREF-NMZ-AND-LF4	yes	yes	no	0.1576	0.2048	0.1401	0.1961	0.7547	1.0524	0.6159	0.9737
TKB-03	yes	yes	no	0.1545	0.2337	0.1710	0.2586	0.7309	1.2162	0.8088	1.3458
TKB-01	yes	yes	no	0.1522	0.2484	0.1684	0.2748	0.6838	1.2524	0.7567	1.3859
ORGREF-OT-D-LF5	yes	yes	no	0.1407	0.2082	0.0898	0.1748	0.6756	1.0286	0.4716	0.8876
ORGREF-NMZ-OR-LF4	yes	yes	no	0.1372	0.1620	0.1272	0.1626	0.6818	0.8954	0.5668	0.8446
ORGREF-OT-DT-LF5	yes	yes	no	0.1237	0.1832	0.0776	0.1548	0.5462	0.8601	0.3738	0.7378
ORGREF-NMZ-AND-LF5	yes	yes	no	0.1139	0.1564	0.0753	0.1161	0.5076	0.7090	0.3123	0.5224
ORGREF-OT-DT-LB2	yes	yes	no	0.0971	0.1653	0.1075	0.1829	0.5084	0.9736	0.5626	1.0774
ORGREF-OT-DT-LB3	yes	yes	no	0.0873	0.1356	0.0966	0.1501	0.4133	0.7975	0.4573	0.8825
ORGREF-NMZ-OR-LF5	yes	yes	no	0.0814	0.1168	0.0595	0.0926	0.3822	0.5387	0.2562	0.4194
ORGREF-OT-DT-LB1	yes	yes	no	0.0743	0.1468	0.0822	0.1625	0.3730	0.8007	0.4128	0.8860
W3SJP2003-001	yes	yes	no	0.0735	0.1201	0.0814	0.1329	0.4029	0.6415	0.4458	0.7099
ORGREF-OT-DT-LB4	yes	yes	no	0.0671	0.1212	0.0742	0.1341	0.3124	0.5803	0.3457	0.6422
ORGREF-OT-D-LB2	yes	yes	no	0.0651	0.1364	0.0720	0.1509	0.3963	0.7588	0.4386	0.8397
ORGREF-OT-D-LB1	yes	yes	no	0.0594	0.1244	0.0657	0.1376	0.3428	0.6644	0.3793	0.7352
ORGREF-OT-D-LB3	yes	yes	no	0.0560	0.1015	0.0619	0.1123	0.3071	0.5771	0.3398	0.6386
ORGREF-OT-DT-LB5	yes	yes	no	0.0514	0.0954	0.0569	0.1055	0.2519	0.4695	0.2787	0.5196
ORGREF-NMZ-AND-LB2	yes	yes	no	0.0499	0.1142	0.0552	0.1264	0.2556	0.5518	0.2829	0.6106
ORGREF-OT-D-LB5	yes	yes	no	0.0464	0.0815	0.0514	0.0902	0.2131	0.3954	0.2358	0.4375
ORGREF-OT-D-LB4	yes	yes	no	0.0457	0.0883	0.0506	0.0977	0.2233	0.4288	0.2471	0.4745
ORGREF-NMZ-OR-LB1	yes	yes	no	0.0452	0.1037	0.0501	0.1148	0.2370	0.5097	0.2623	0.5640
ORGREF-NMZ-AND-LB1	yes	yes	no	0.0451	0.1130	0.0499	0.1251	0.2416	0.5735	0.2673	0.6347
ORGREF-NMZ-AND-LB3	yes	yes	no	0.0447	0.0917	0.0494	0.1015	0.2121	0.4500	0.2347	0.4979
ORGREF-NMZ-OR-LB2	yes	yes	no	0.0438	0.1026	0.0485	0.1136	0.2377	0.5006	0.2630	0.5539
ORGREF-NMZ-OR-LB3	yes	yes	no	0.0404	0.0804	0.0447	0.0890	0.1997	0.3937	0.2209	0.4356
ORGREF-NMZ-OR-LB4	yes	yes	no	0.0336	0.0761	0.0372	0.0842	0.1506	0.3226	0.1666	0.3570
ORGREF-NMZ-OR-LB5	yes	yes	no	0.0314	0.0750	0.0348	0.0830	0.1682	0.3623	0.1862	0.4009
ORGREF-NMZ-AND-LB4	yes	yes	no	0.0290	0.0782	0.0321	0.0866	0.1444	0.3543	0.1598	0.3920
ORGREF-NMZ-AND-LB5	yes	yes	no	0.0287	0.0800	0.0318	0.0885	0.1677	0.3920	0.1856	0.4338
r2d2-a1	yes	yes	no	0.0250	0.0551	0.0276	0.0609	0.1322	0.2502	0.1463	0.2769
r2d2-a2	yes	no	no	0.0802	0.1189	0.0888	0.1316	0.3654	0.4963	0.4044	0.5492
ORGREF-NMZ-AND	yes	no	no	0.0775	0.1077	0.0857	0.1191	0.4634	0.6865	0.5128	0.7597
ORGREF-NMZ-OR	yes	no	no	0.0713	0.0926	0.0789	0.1025	0.4248	0.5777	0.4701	0.6393
ORGREF-OT-DT	yes	no	no	0.0699	0.1578	0.0774	0.1747	0.4107	0.8641	0.4545	0.9563
TKB-04	yes	no	no	0.0667	0.1125	0.0738	0.1245	0.3547	0.6304	0.3925	0.6976
ORGREF-OT-D	yes	no	no	0.0517	0.1119	0.0572	0.1238	0.2824	0.6171	0.3125	0.6829
TKB-02	yes	no	no	0.0435	0.1067	0.0481	0.1181	0.2136	0.5382	0.2364	0.5955
oksat-bf01	yes	no	no	0.0327	0.0615	0.0362	0.0680	0.1808	0.3420	0.2001	0.3785

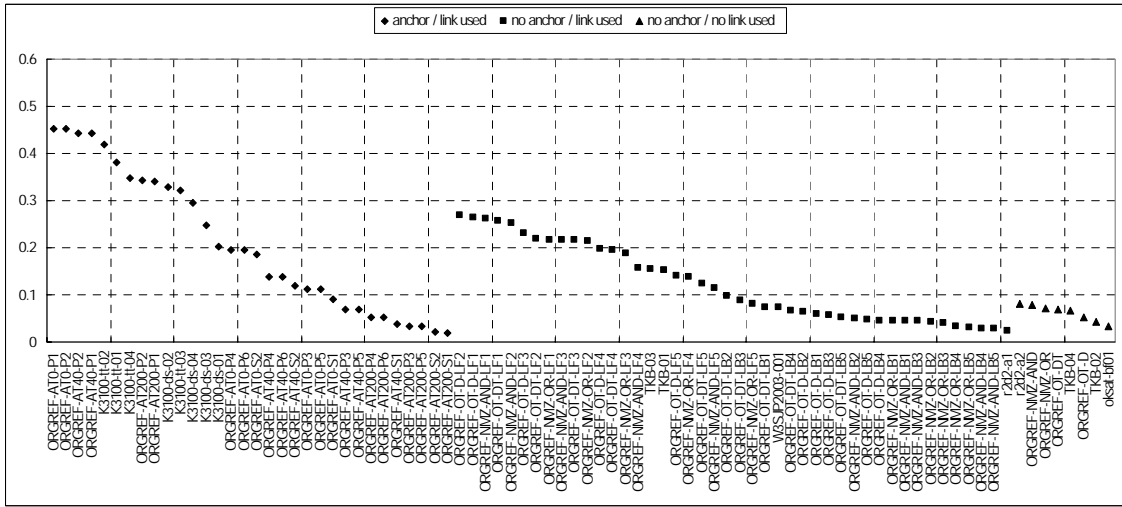


Figure 1. WRR values of run results on (DS-1) and (RL-1).

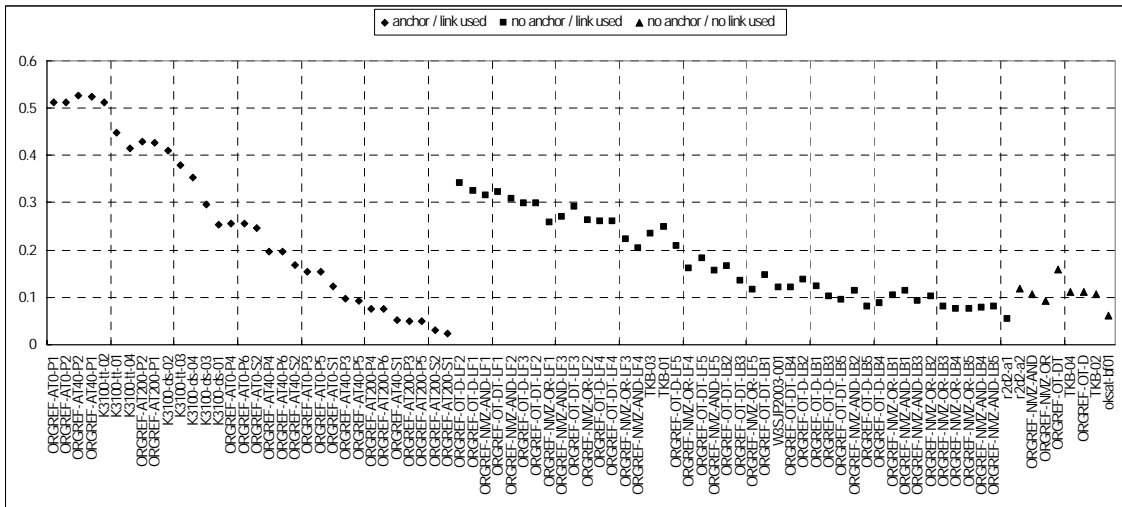


Figure 2. WRR values of run results on (DS-1) and (RL-2).

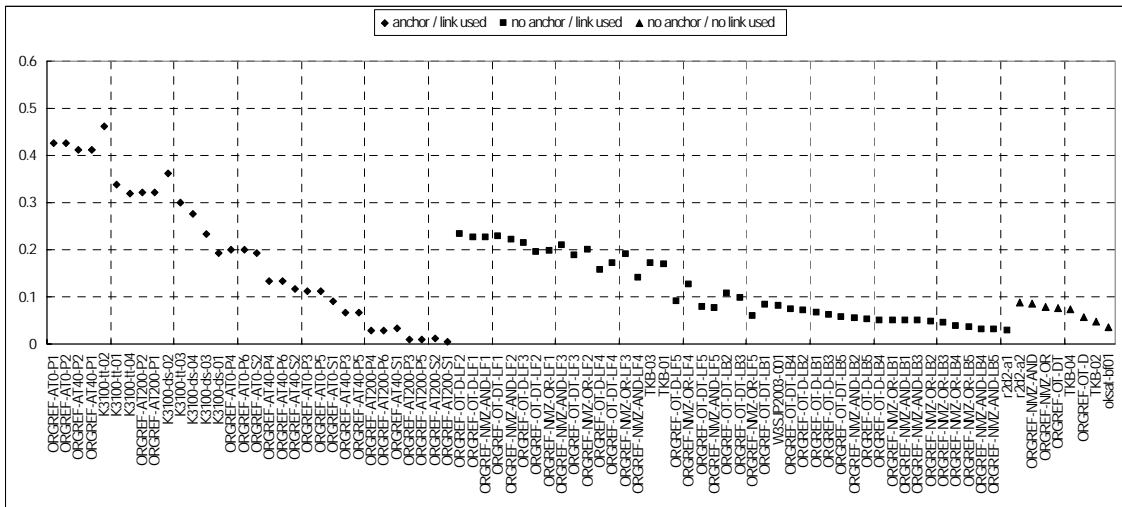


Figure 3. WRR values of run results on (DS-2) and (RL-1).

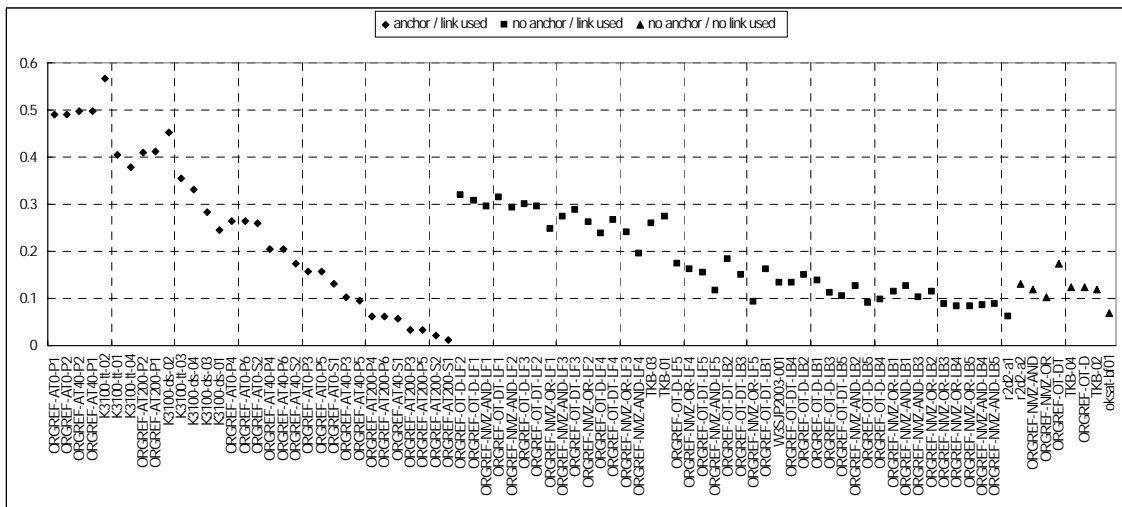


Figure 4. WRR values of run results on (DS-2) and (RL-2).

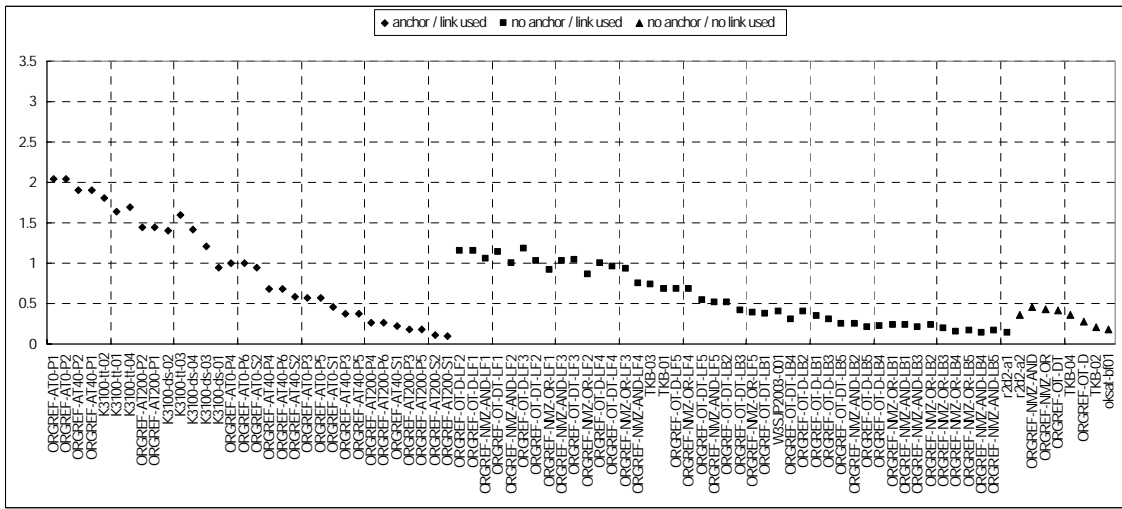


Figure 5. DCG values of run results on (DS-1) and (RL-1).

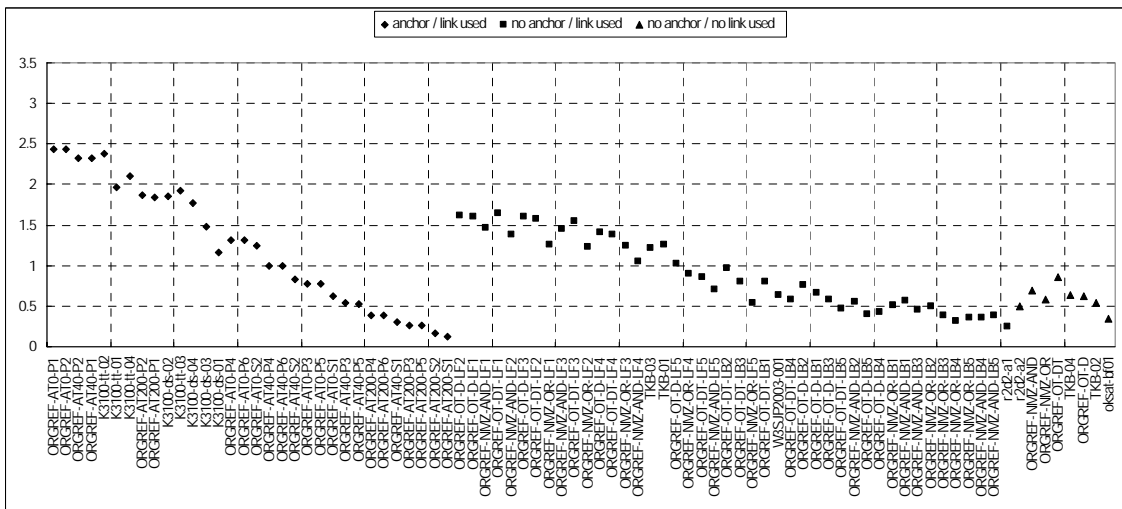


Figure 6. DCG values of run results on (DS-1) and (RL-2).

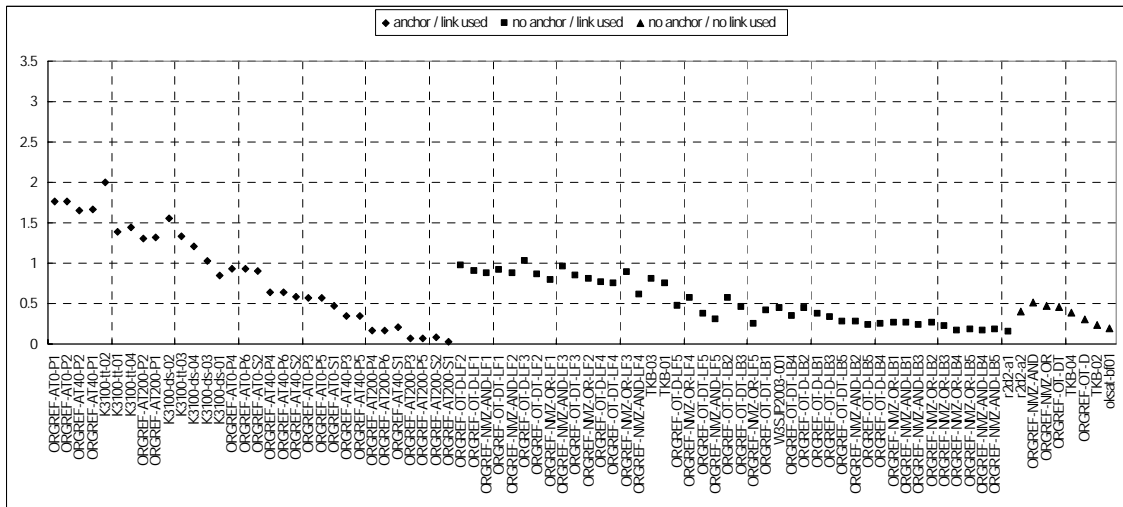


Figure 7. DCG values of run results on (DS-2) and (RL-1).

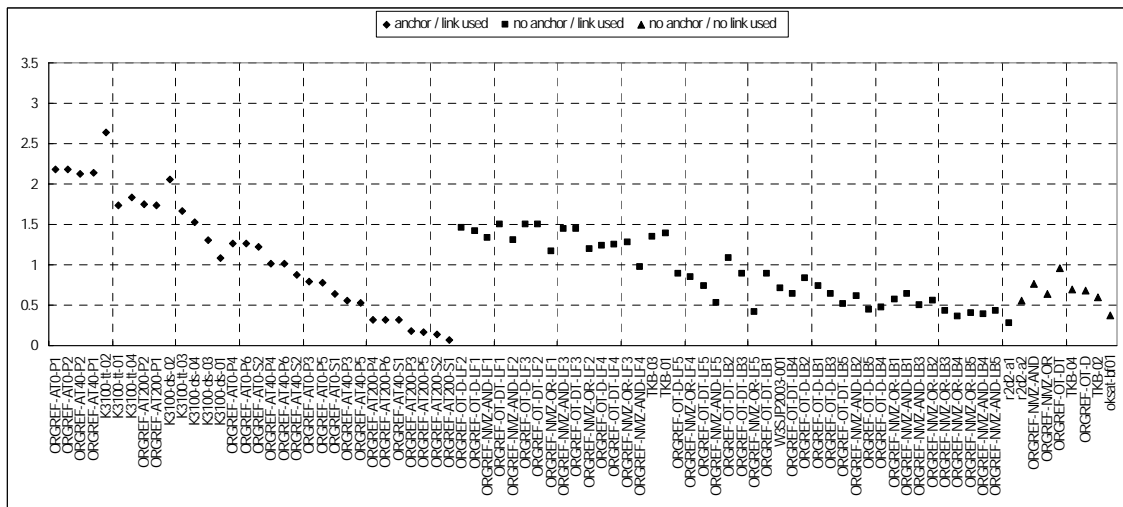


Figure 8. DCG values of run results on (DS-2) and (RL-2).

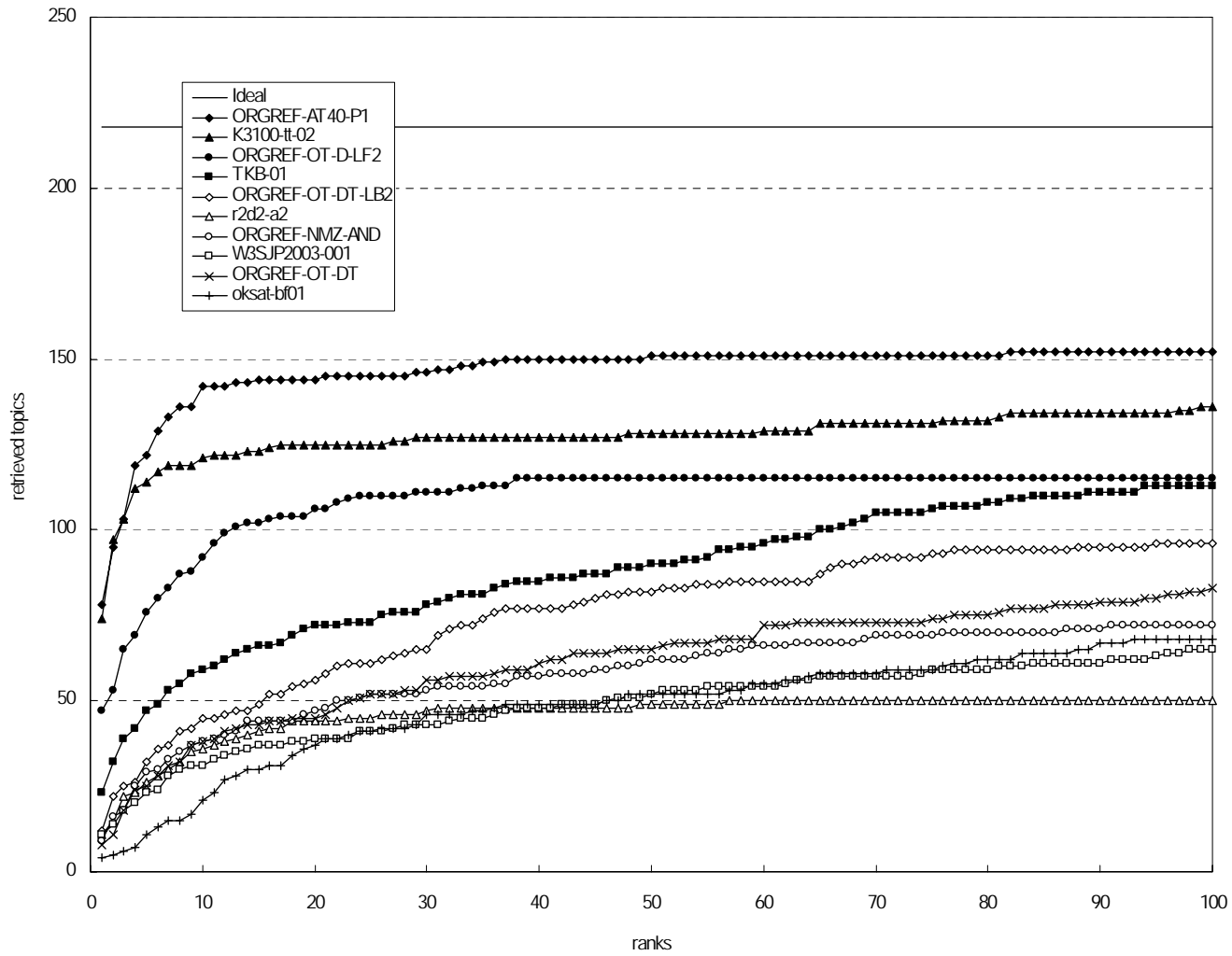


Figure 9. Cumulative number of topics whose relevant documents were retrieved on (DS-1) and (RL-1).

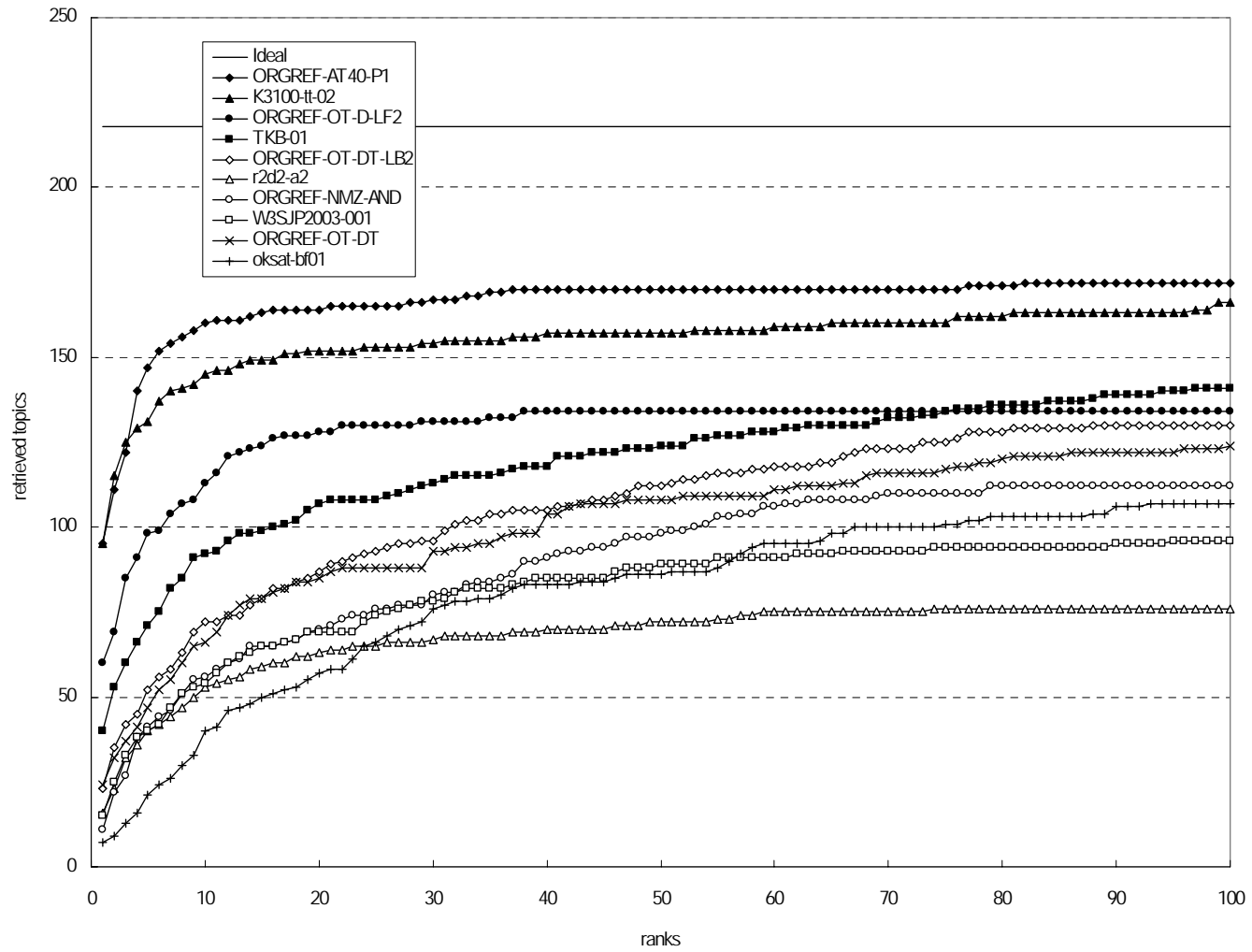


Figure 10. Cumulative number of topics whose relevant documents were retrieved on (DS-1) and (RL-2).

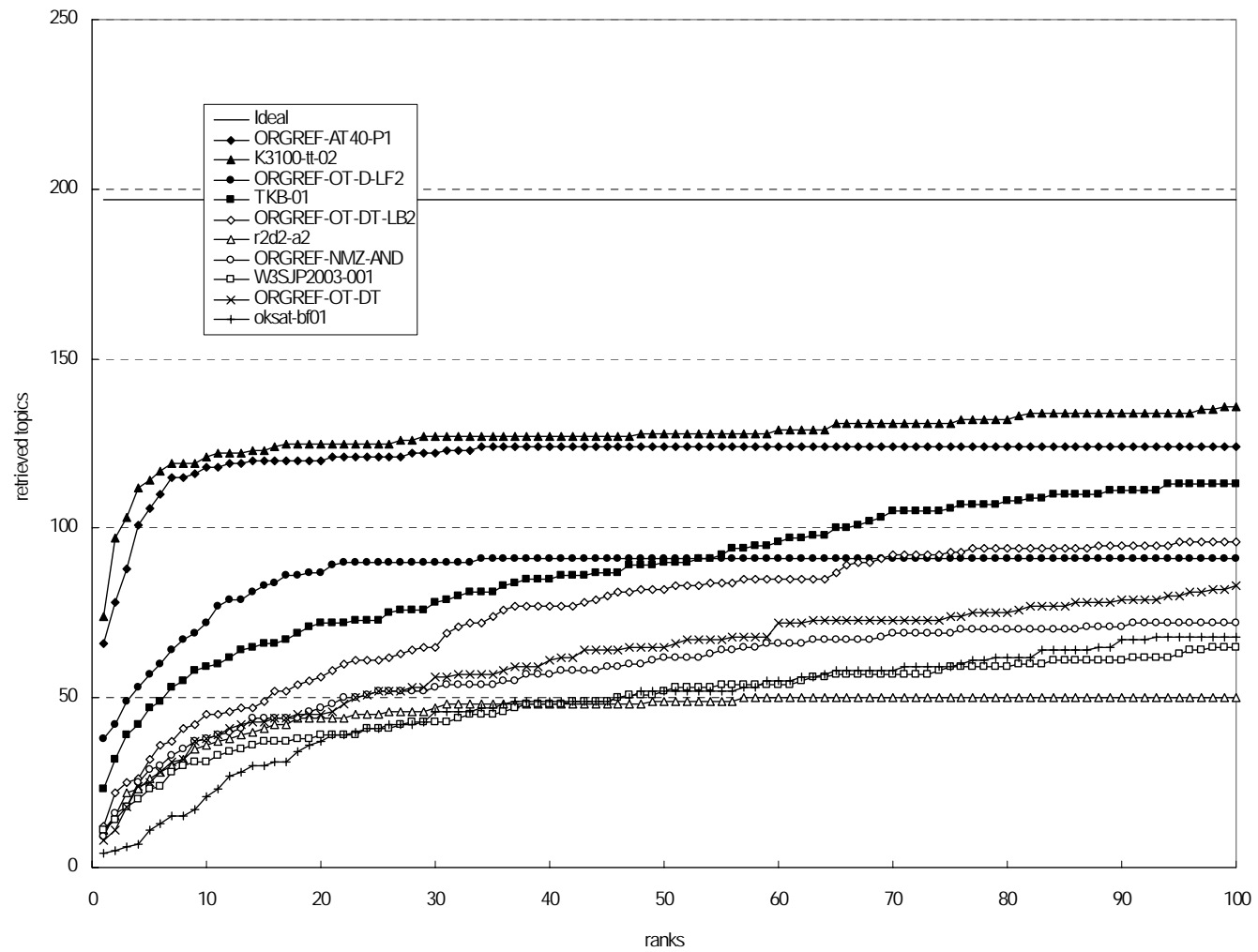


Figure 11. Cumulative number of topics whose relevant documents were retrieved on (DS-2) and (RL-1).

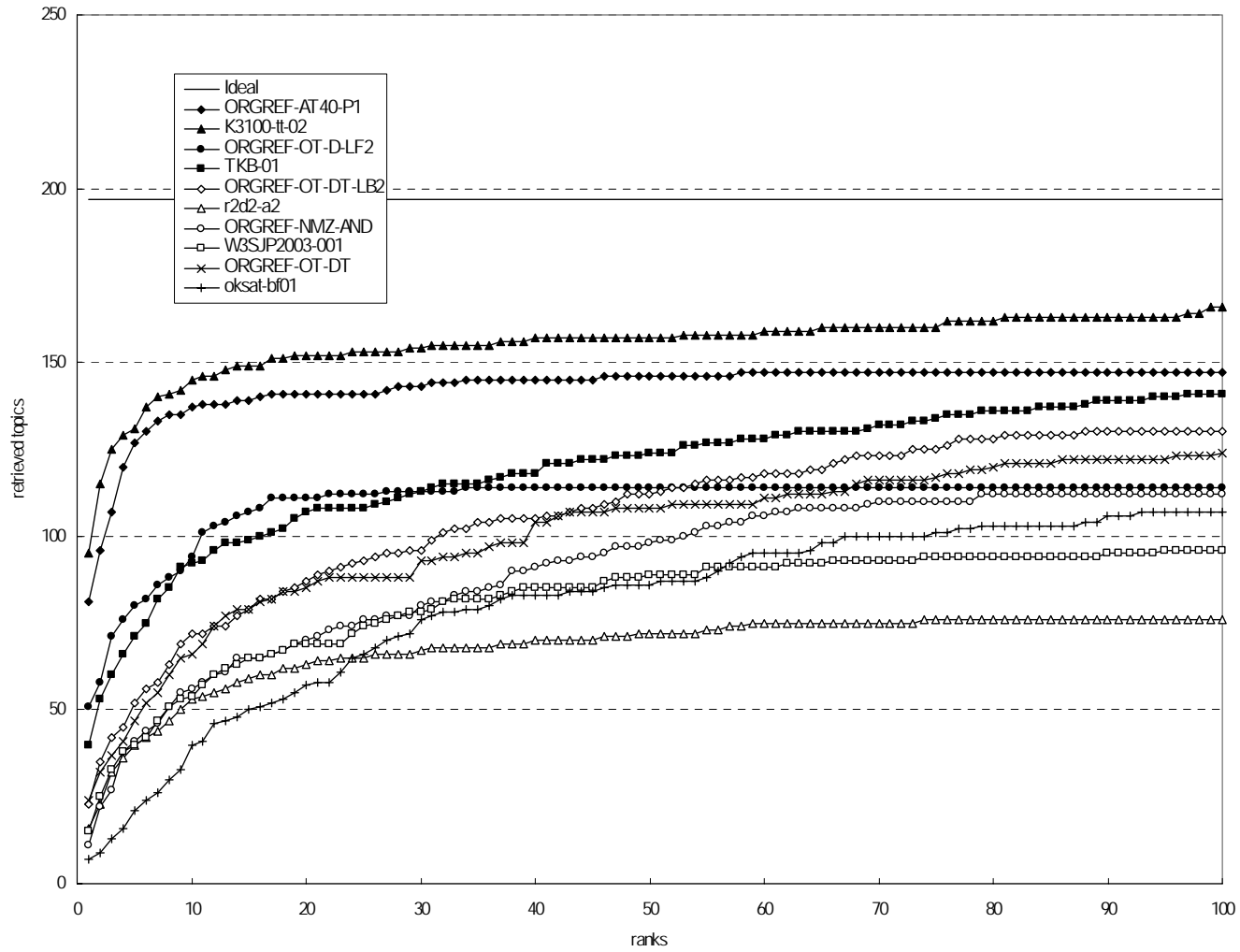


Figure 12. Cumulative number of topics whose relevant documents were retrieved on (DS-2) and (RL-2).