Overview of the NTCIR-5 WEB Navigational Retrieval Subtask 2 (Navi-2)

Keizo Oyama[1] Masao Takaku[2] Haruko Ishikawa[3] Akiko Aizawa[4] Hayato Yamana[5]

[1,2,3,4,5] National Institute of Informatics (NII)
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
{[1]oyama, [2]masao, [3]haruko, [4]akiko}@nii.ac.jp

[1,4] Department of Informatics, School of Multidisciplinary Science
The Graduate University for Advanced Studies (SOKENDAI)

[5] Department of Computer Science, School of Science and Engineering
Waseda University
3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan
[5]yamana@yama.info.waseda.ac.jp

Abstract

This paper describes an overview of the Navigational Retrieval Subtask 2 that was conducted from 2004 to 2005 as a subtask of the WEB Task at the Fifth NTCIR Workshop. In the Subtask, 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. 1.36TB web document data and 400 topics were distributed to the participants and, in turn, 35 run results were submitted by 4 participants and 28 by the organizers. Relevance judgments were performed on the documents pooled from the run results, mainly in terms of representativeness of search target items given by the topics. Several kinds of evaluation measures were applied to the run results submitted by each participant. Simple analyses on system evaluation results and on test collection characteristics are given.

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

1 Introduction

This paper describes an overview of the Navigational Retrieval Subtask 2 that was conducted from 2004 to 2005 as a subtask of the WEB Task at the Fifth NTCIR Workshop (NTCIR-5 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 in the NTCIR-4 WEB naming it "Navigational Retrieval Subtask 1 (Navi-1)" [1] and continued it in the NTCIR-5 WEB with a much larger document data set and with more topics. Thus we call this subtask as "Navigational Retrieval Subtask 2 (Navi-2)."

In the Subtask, we attempted to evaluate the retrieval effectiveness of web search systems aiming at "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 to a certain degree.

We used 1.36TB (1.5×10^{12} byte) web document data (NW1000G-04)[1] compiled for Navi-2 and 400 topics created by 17 people in this subtask as a common data set[2]. This data set was distributed to 9 participants, and, in turn, 38 search run results were submitted by 5 groups and 28 by the organizers. However, three out of 38 run results submitted by one out of the 5 groups did not include any relevant or partially relevant documents for some unknown reason. Therefore, system evaluation is done on remaining 35 run

[1] For some circumstances, it is called "NW1000G-04" but not "NW1360G-04".

[2] Approximately 800 topics were also created and distributed as optional ones that will be used for further analysis.

results from 4 groups.

Relevance judgments were performed on the documents pooled from the run 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” [2] 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 search target 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 about: 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 and analysis in Section 7, and conclusion in Section 8.

2 Definition of Navigational Retrieval Subtask 2: Known Item Search

In Navi-2, we tackled with system evaluation for “Known Item Search” just as in Navi-1, but with a much larger document data set and with more topics.

“Known Item Search” 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 already knows about the item but does not necessarily know about its web page.

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 subtask.

For some search target items that exist in the real world, such as products, organizations, stores, persons, facilities, natural things, events, only their information exists on the web; whereas for other search target items, their entity exists on the web, such as information services, blogs, data files, documents and online shops. 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
- Has forgotten the name though remembers the outline

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:

- The searcher has viewed the page and remembers its outline.
- The searcher has viewed the page but does not remember clearly what the page was like.
- The searcher has never viewed 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 to be as follows, although the final relevance judgment depends on subjective views of assessors:

(1) Provider of the representative 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 representative web page

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

3 Document Data Set

The document data set NW1000G-04 consists of web page text files (documents) of approximately 1.36TB (1.5×10^{12} bytes) in total and several kinds of lists.

Four types of document data were prepared by the organizers:

- raw data: web page data just as were crawled from the web; size is 1.36TB.
- euc data: web page data processed from "raw data" converting Japanese character codes to EUC code.
- cooked data: text data processed from "euc data" removing unnecessary HTML tags and elements.
- segmented data: segmented word data generated from "cook data" using a morphological analyzer "mecab".

Three kinds of lists were prepared by the organizers:

- sitelist: a list of crawled web sites consisting of site identifiers and site names.[3]
- doclist: a list of documents in the data set consisting of document identifiers and documents' URL's. Search results can include only documents which are listed in the "doclist".
- linklist[4]: a list of link data consisting of document identifiers and documents' URL's of link source documents and link target documents respectively, both contained in the "doclist".

The web pages were crawled mainly from web servers of Japan from January 2004 through January 2005 with the following steps:

(1) About 450,000 start-up hosts in '*.jp' domain were gathered from previous crawls and were crawled starting with the top page up to 15 hyperlink hops.
(2) URL's in '*.jp' domain found at step (1) in new hosts not included in the above mentioned start-up hosts were collected. Then the hosts were crawled starting with the top page up to 8 hyperlink hops.
(3) URL's out of '*.jp' domain found at steps (1) and (2) in new hosts were collected. Then the hosts were crawled starting with the top page up to 10 pages.
(4) Language identification was performed on the pages fetched at step (3). Then those hosts that included at least one page judged as Japanese were selected.
(5) The hosts selected at step (4) were crawled starting with the top page up to 8 hyperlink hops.

Web pages that were judged to be written in other languages than Japanese or English by a language identifier produced by Basis Technology Corp. were removed from the crawled pages to make the data set.[5,6]

4 Search topics

4.1 Creation and selection

We selected 400 topics out of 891 topics that were created by 17 people for delivery as the result of discarding similar ones and inappropriate ones from several view points. Most of the topic creators are undergraduate and graduate students of various disciplines from 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 for 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.

[3] A site name is composed of a protocol name, a host name and an optional port number.

[4] A significant proportion of link data were lacking in the "linklist" delivered to the participants with the data set for the first time. We found the problem after we started relevance judgment. We soon delivered a fixed "linklist" and requested the participants to redo runs that are using "linklist." In response, four groups and the organizer resubmitted 44 run results in total. We first conducted relevance judgment on the run results using the incomplete "linklist" and delivered system evaluation results to the participants. After we received the run results using the fixed "linklist", we conducted additional relevance judgment and delivered revised system evaluation results. This paper is based on the revised version of the run results and their relevance judgment results.

[5] In the process of removing web pages written in other languages, all or part of web pages from a certain amount of web sites were deleted for some operational error and potentially relevant pages of several topics have been lost. Negative effects on link-based methods are anticipated; however its effects to the system evaluation results have not been assessed by now.

[6] Although pages judged as in languages other than Japanese and English by a language identifier, there still remain a large number of pages in various languages.

However, since the document data set is collected from January 2004 through January 2005, items whose representative web pages were considered not to have existed at that time were excluded from the search topics.[7]

4.2 Format and elements of search topics

A search topic is described in tagged format shown below. The language is Japanese.

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</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**: Contains one search topic.

(2) **NUM**: Topic number used as topic ID.

(3) **TYPE**: Topic type.
One of the codes 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 of the codes defined as follows:
A: Products / services (not including services provided on the web).
B: Companies / organizations (including shops and administrative organs, but not including online shops).
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: Online shops and online services (not including those in F).
H: 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.
Single Japanese sentence briefly describing the information need. Although it should be conceptually consistent with TITLE, the search terms as they are in TITLE may not appear in DESC.

(7) **NARR**: Narrative of the information needs.
Explanation of the information needs which are not fully represented with TITLE and DESC.

(8) **NARR/TERM**: Explanation of terms (Optional).
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 (Optional).
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.
Searcher's position, sex, and experience years of web search.

(12) **USER/@SPECIALITY**: Knowledge level.
Searcher's knowledge level on the search target item; one of the codes 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 little about it.

5 Run conditions

5.1 Search run execution

[7] 842 optional topics were also delivered and are now assessed. They are used, in combination with the above-mentioned topics, for analyzing relationship among search techniques, topic types, search item categories and relevant page styles.

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 corresponding run results using (1) or (2), excluding TYPE and CATEGORY from them.

There is no limitation for the number of run results that can be submitted by a participant.

Both automatic and interactive processing modes are permitted. The run is regarded as interactive when a human has a hand in any way affecting the search results 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
How link information is used for searching and ranking.

URL Information
How URL is used for searching and ranking.

TAG Information
How HTML tags are used for searching and ranking.

Anchor
How anchor text is used for searching and ranking.

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.

Some of these items are used for analyzing system evaluation results and others will be used in the further analyses.

6 Relevance assessment

Four participants submitted 35 run results.[8] Each run result includes up to 100 documents for each topic. Organizers added 28 run results in order to find relevant documents comprehensively so that the test collection can be re-usable.

Relevance assessment for 324 out of 400 topics was actually performed by each topic creator and that for remaining 76 topics was performed by someone who can understand the topic well.

Pooling was applied to the run results for the relevance assessment in this subtask.

We used a newly developed assessment system which was designed to enable assessors to view documents and to follow hyperlinks just like browsing the real web pages. A list of documents to be assessed is presented to the assessor; however when he/she finds a document to be (partially) relevant not included in the list, he/she can judge the document through the system and it will be added to the judgment list automatically.

6.1 Pooling

Each pool was made by extracting top N (=20) highly ranked documents from every run result of each topic.

For Navi-1, we selected 10 for N because many search engines return top 10 search results as the first response, and because it was estimated to

[8] Other three run results submitted by a participant included significant number of inappropriate document identifiers and were excluded from the further processing.

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.

For Navi-2, because the number of the documents is more than 10 times larger than for Navi-1, we selected 20 for *N*.

The sequence of assessment is decided as follows: (1) sort the documents with rank as the primary key and top-down-ordered domain name plus path name of URL as the secondary key; and (2) 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.[9]

Although pooling was applied, since web pages are connected with hyperlinks and assessors are required to follow them when necessary for accurate judgment and to find relevant documents aggressively, any of the documents in the document set, not pooled or even not submitted, potentially becomes the object of relevance assessment. We expect most of the relevant documents not in the pool have eventually been assessed by following probable hyperlinks.

6.2 Judgment bases

An assessor was required to use as the judgment bases not only text in the document but also clues that the assessor usually uses in web browsing and that usual searchers are considered to use, e.g. host name, URL pattern, and HTML tags.

Since the document data set holds only text files, assessment is often very difficult without embedded multimedia data such as images and flashes. Hence the assessment system shows to the assessor the web pages embedded with the current data on the web, which may have already been deleted or changed.

Concerning judgment of a frame set page and a page that automatically redirects to another URL, the assessor refers to its link target pages and take them into the judgment bases. The assessor also refers to link source pages of a frame component page and an automatically redirected page if it is potentially relevant or partially relevant. When they are not included in the document data set, the current web pages of the same URL's are referred.

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 such as site top 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 of the search target item satisfying the retrieval needs. More than one independent pages can be "relevant".

B: Partially relevant
A page almost satisfying the retrieval needs; pages as follows, but not limited to them, fit to this relevance level:

- ♦ A representative page of an item having an upper or lower concept of the search target 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 search target item, in terms of contents and reliability.

D: Non-relevant
Otherwise.

6.4 Additional judgment

Aspects which should be taken into account on system evaluation besides relevance as follows 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 the search target item (hereinafter, different item).
- The different item cannot be excluded semantically by only TITLE and DESC.

For instance, when only an informative term in TITLE and DESC is "TOTO" and when NARR/TERM defines it as a name of a totocalcio service in Japan, a representative page of a company named "TOTO" or of a music group "TOTO" is judged as "undistinguishable".

Representative pages of a search target item

[9] In order that the assessors can view a document list in an order appropriate for assessment steps, the assessment system can sort the list with several keys, i.e. URL, page title, document ID, rank, judgment result, additional judgment result and domain-based cluster.

exclusively for mobile phones and so on are also treated as "undistinguishable".

An assessor also gives short description what the page is and how undistinguishable it is.

We define undistinguishability level as below according to how generally well-known the different item is when compared to the search target item.

3: The different item is more well-known than the search target item.
2: The different item is as well-known as the search target item.
1: The different item is less well-known than the search target item.

However, for efficiency reasons, only undistinguishability without level has been assessed by now. We will further investigate if we really need to consider the level.

(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 their contents are completely the same, pages which are considered to have different link target pages or to have different roles in the real web space are not deemed to be duplicate pages.

(3) Other languages

Navi-2 is mainly targeting at search systems for Japanese web pages. However, for certain kinds of search items or searchers, English web pages or others can suffice the information needs. Thus, when there are relevant or partially relevant pages written in another language than Japanese, the language name is judged by the assessor.

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.

- Kansai Lab, NEC Corporation (anonymous[10]) [3]
- Research and Development Strategy Department; Justsystem Corporation (JSWEB) [4]
- Sato Laboratory; Osaka Kyoiku University (OKSAT) [5]
- Software Engineering Center; University of Aizu (OASIS) [6]
- University of Tsukuba, Nagoya University, Toyohashi University of Technology (TNT) [7]

The individual participating groups pursued various objectives. We summarize them as follows[11] (listed in alphabetical order of group ID's):

JSWEB: They experimented with a distributed search system based on Vector Space Model with a term-partitioned inverted file. An agent-based mechanism is used to merge the search results from each index component. They used TF-IDF on full text of respective single documents without document length normalization and attempted two additional methods: "tf limitation" with which too large TF values are replaced with a rather small value; and link structure analysis using in-link page counts and in-link domain counts.[12]

K3100: They experimented with a retrieval method using anchor text for retrieving web documents that are pointed to by them. They attempted to use information obtained by analyzing anchor text count, in-link count and domain structure, etc. For ranking documents, they used several combinations of one or two from six measures: anchor frequency, reference consistency, query term weight, page representativeness, site relevancy and inverse anchor document frequency.[13]

OASIS: They experimented with a distributed search system based on Vector Space Model with a document-partitioned inverted file. Two

[10] The group ID is not shown because they participated as anonymous type.

[11] Summaries of the run result submissions are available in the supplemental CD-ROM attached to the NTCIR-5 Proceedings.

[12] They reported they executed runs with a slightly different scoring equation after they submitted their run results. Please note their run results presented in this paper are hence different from those in their paper.

[13] They reported there were errors in extraction of anchor texts and in counting anchors. These errors may have caused degradation to the run results. Please note that some run results presented in this paper may be different from those in their paper.

methods for merging partial search results from each index component were experimented with. They used TF-IDF on full text of respective single documents and attempted to use term weighting methods taking into account document headings and distance from URL's in the text.[14]

OKSAT: They experimented with three gram-based indices for full text, title part and anchor text respectively. Several methods for merging page-content-based scoring using full text and title part, anchor-text-based scoring and link-analysis-based scoring using in-link count and PageRank are attempted. The page-content-based and anchor-text-based scoring used probabilistic models.[15]

ORGREF: The organizers executed runs to expand the pools and to attempt several link-based and anchor-text-based techniques using four groups of search systems: (1) two content-based baseline systems, one using Boolean model and TF-IDF-based ranking and the other using Okapi BM-25 on page content; (2) two anchor-text-based baseline systems, one using Boolean model and TF-IDF-based ranking and the other using Okapi BM-25 on virtual document comprising source anchor texts; (3) link-analysis-based experimental systems attempting methods to expand the content-based retrieval sets using one-hop forward or backward link analysis; (4) anchor-text-based experimental systems attempting methods using expanded anchor text combined with link structure analysis.[16]

TNT: They experimented with a system combining probabilistic model for full text content and language model for virtual document comprising source anchor texts. Weighted harmonic mean of ranks obtained by the two components is used for the final ranks. PageRank was also applied to modify the scores of anchor-text-based search results.[17]

In total, 44 runs used anchor text, 39 runs used link information and 4 runs used URL information with various combinations.

Many participants suffered from large document data set, which caused various system faults and data processing errors. Four out of nine participants could not submit run results for some unknown reasons, and furthermore, all of the five participants that submitted run results but the organizers executed runs with bug fixes, tuning of scoring functions, or even new approaches after they submitted the formal run results. The size of the document data may have been challenging to average research groups

7.2 System evaluation methods

We delivered 400 topics to the participants and assessed documents in the submitted runs based on the relevance judgment method described in Section 6. Then, we selected such topics that at least one "relevant" document was found in the document set. Consequently, we used 269 topics for the system evaluation.[18]

As the evaluation measures, we calculated DCG (Discounted Cumulative Gain) and WRR (Weighted Reciprocal Rank) [8][9] with cut-off at 10.

Gains used in the DCG calculations are:

$$(G_a, G_b) = (3, 0), (3, 2) \text{ and } (3, 3),$$

and gains used in the WRR calculations are:

$$(\delta_a, \delta_b) = (1, 0), (\beta_a, \beta_b) = (\infty, \infty),$$
$$(\delta_a, \delta_b) = (1, 1), (\beta_a, \beta_b) = (\infty, \infty).$$

Two sets of parameters used in the WRR calculations are such that the measure becomes identical with MRR (Mean Reciprocal Rank), one for "rigid" relevance level (i.e., documents with assessment 'A' are regarded as relevant documents) and the other for "relaxed" relevance level (i.e., documents with assessment 'A' and 'B' are regarded as relevant documents).

Although many topics have multiple relevant documents respectively, 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 fully considered in the current evaluation yet, appropriateness of DCG values as the system effectiveness is left to be investigated. However, because only the first retrieved relevant documents are used for WRR with the above-

[14] Their system evaluation result is not shown in this paper because their run results included no relevant or partially relevant documents for some unknown reason.

[15] In their paper, they present run results using a method for anchor-text-based scoring that is different from one used in the submitted runs. Please note their run results presented in this paper are hence different from those in their paper.

[16] Brief description of the runs is given in Appendix.

[17] They reported there were some bugs in their system and re-executed run results are presented in their paper. Please note their run results presented in this paper are hence slightly different from those in their paper.

[18] There were actually 270 topics having relevant documents. However, two of them were on the same search target item, although TITLE and DESC were different. We therefore excluded one out of the two for the system evaluation.

mentioned parameter settings, the appropriateness is not affected by duplication or link relation. Therefore, we will use WRR as the main evaluation measure in this paper.

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.

As is mentioned above, all participants executed runs with modified systems after they submitted the formal run results. Inversely speaking, they did not have sufficient time to refine their techniques or to tune the scoring functions. Hence the evaluation results presented hereinafter should be considered tentative and we will not get into the details of individual systems or techniques in this paper.

The evaluation results are shown in **Table 1.**[19] The Run ID's are classified into four groups: runs using anchor information and link information, runs using anchor information but link information, runs using link information but anchor information, and runs using neither link information nor anchor information; and arranged in the descending order of the WRR with parameters (δ_a, δ_b) = (1, 0) in each group.

Figures 1 and **2** are graphs plotting WRR values, and **Figures 3, 4** and **5** are graphs plotting DCG values. The orders of the Run ID's are the same as Table 2.

As is seen in the Figure 1 through 5, runs showing high performance, e.g. TNT-1,3 and K3100-7,12, utilize anchor text information in certain ways, although their IR models and ranking methods differ. Besides, several runs utilizing link information, e.g. ORGREF-GC1-LF1, performed fairly well. Runs using content information only stayed at rather low performance. These tendencies are basically the same as Navi-1 results, except that several content-based systems performed much better at Navi-2.

It should be noted that link-based and URL-based methods could make no contribution to improve performance of anchor text based systems except those that use link information among sites rather than among pages like K3100-7 and OKSAT-WEB-F-09.

Graphs of cumulative numbers of topics whose relevant documents were retrieved by several selected runs are shown in **Figures 6 and 7** on rigid and relaxed relevance levels respectively.

At Navi-1, 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.

At Navi-2, the tendencies of curves based on link information and curves based on content information only are the same; however, some of the curves based on anchor information rise rather gradually although well performing ones rise rapidly.

The reason for this difference between Navi-1 and Navi-2 is not clear, but it may be partly because of some faults of the participants' systems.

7.4 Analyses on the effects of some web-specific document characteristics

For web information retrieval, especially for navigational retrieval, web-specific document characteristics affects on the system performance. Here we attempt simple analyses on frameset structure and duplication.

(1) Effect of frameset structure

Anchor text information and most of link information exploited by the participants are considered to be global information, i.e. information provided by general public, and their effectiveness is undoubted.

However, for quite a new page or an unpopular page, such information is not available. Hence, content-based techniques, possibly combined with techniques to exploit local link information, cannot be ignored.

Based on a consideration that one of the reasons why content-based systems poorly performed is because of frameset structure, we attempted a simple compensation on the relevance judgment results such that, if a retrieved page is used as a component of a frameset page with higher relevance than itself, then the page should be regarded as with the same relevance as the frameset page. When the page is used by multiple frameset pages, the above is not applied.

As the result, relevance judgments of 172 documents for 62 topics are modified. **Figure 8** shows the change of WRR (1, 0) values. Note that the scale of "difference" is 50 times smaller than that of the original WRR value.

Although the changes are small compared to

[19] Descriptions of Anchor Info, Link Info and Cont Info are based on the author's interpretation of participants' system description forms and their proceedings papers. Therefore, they are not necessarily consistent with the system descriptions presented in the supplementary CD-ROM.

the absolute value, noticeable increases are observed for several content-based systems. A few link-based systems exploiting local link information also gained noticeable improvement (e.g. ORGREF-GC1-LB2). Contrarily, it is quite reasonable that well-performing systems exploiting anchor text only (without content information) gained no improvement.

The compensation technique used in this analysis is effective to a certain extent in spite of its simplicity. More sophisticated technique is therefore promising for systems not using global information and for topics searching for very new or unpopular pages.

(2) Effect of duplicate pages

Duplicated pages are frequent in the web space. Although duplicates are usually useless or even obstructive from the point of users' view, they have an effect to shift some evaluation measures to higher side.

Thus, in order to grasp the degree of such an effect, we compensate the DCG values using the judgment of duplication. Note that redundancy caused by links is not considered in this attempt. According to the intuition obtained by monitoring the relevance assessment process, redundancy by links is much larger than that of duplicates.

Figures 9 to 11 show the change of DCG values for parameters (3, 0), (3, 2) and (3, 3) respectively. A short bar below each plot is the compensated value. In these graphs, duplicate pages appearing secondly or later are regarded as non-relevant. If such duplication is just ignored, the plot will be in-between.

The effects vary largely among runs and no consistent tendency is observed among the four groups. However, it is feasible that the system ranking with compensated DCG measures is more similar to that with WRR measures.

The results suggest it is necessary to compensate redundancy caused by links in order to evaluate systems accurately with topics having multiple relevant pages. Thus, we will further investigate methods to detect and to assess the redundancy and methods to compensate evaluation measures.

8 Conclusion

In this paper, an overview of the Navigational Retrieval Subtask 2 of the WEB Task at the Fifth NTCIR Workshop was described. It aimed at evaluating web search engine systems for retrieving representative web pages of known items.

We used a 1.36TB web document data set, NW1000G-04, constructed for this subtask.

400 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 20 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. We expect most of the relevant documents not in the pool have eventually been assessed.

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 four groups, it seems that a group utilizing anchor text performed best, another group utilizing anchor text and link information performed almost the same or a little worse, another group utilizing link information but anchor text performed fairly well, and the other group using content information only performed rather poorly.

Link information used in combination with anchor text is ineffective or even harmful. However, link information among sites may be effective.

By analyzing the effect of frameset pages, we pointed out a possibility to improve performance of systems that do not use global information.

We are currently conducting analysis of the effects on evaluation results by duplication and redundancy of interlinked pages. By analyzing duplication alone, necessity for proper handling of these redundancies in system evaluation was shown. In addition, we will investigate effects of undistinguishable documents and, if necessary, a method for their compensation. Topic by topic and page by page analyses are also the issues to be conducted.

We have labeled the topics with types and categories. Using them together with other data obtained through Navi-2, we are going to analyze relationship among types and categories of topics, styles and link statistics of relevant pages, and search techniques.

Moreover, using the submitted run result data and the relevance judgment, we will test various measures suitable for navigational retrieval in terms of their stability.

Finally, new evaluation measures suitable for navigational retrieval such as a measure taking into account costs of inputting search terms and browsing retrieved documents will be investigated.

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.

Web crawling was performed in cooperation with e-Society Foundation Software project of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

We greatly appreciate the efforts of all the participants in the Navigational Retrieval Subtask 2 of the WEB Task at the Fifth NTCIR Workshop. We also appreciate the helpful advice of Professors Jun Adachi and Noriko Kando, National Institute of Informatics, and the intensive work on document data preparation by Mr. Shin Kato and Mr. Kengo Minamide, Waseda University.

References

[1] Keizo Oyama, Koji Eguchi, Haruko Ishikawa and Akiko Aizawa: "Overview of the NTCIR-4 WEB Navigational Retrieval Task 1", *in Proc. NTCIR-4*, Tokyo, 2004 (available from <http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings4/WEB/NTCIR4-OV-WEB-B-OyamaK.pdf>).

[2] Craswell, N. 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 >).

[3] Tateishi, K., Kawai, H., Kusui, D. and Fukushima, T.: "Verification of Effective Retrieval Method for Anchor Text on Navigational Retrieval", *in Proc. NTCIR-5*, Tokyo, 2004.

[4] Tanioka, H., Yamamoto, K. and Nakagawa, T.: "A Distributed Retrieval System for NTCIR-5 WEB Task", *in Proc. NTCIR-5*, Tokyo, 2004.

[5] Sato, T. and Nakakubo, H.: "NTCIR-5 WEB Navi-2 Experiments at Osaka Kyoiku University", *in Proc. NTCIR-5*, Tokyo, 2004.

[6] Klyuev, V.: "OASIS at NTCIR-5: WEB Navigational Retrieval Subtask", *in Proc. NTCIR-5*, Tokyo, 2004.

[7] Fujii, A., Itou, K., Akiba, T. and Ishikawa, I.: "Exploiting Anchor Text for the Navigational Web Retrieval at NTCIR-5", *in Proc. NTCIR-5*, Tokyo, 2004.

[8] 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>).

[9] 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>).

Table 1. Selected evaluation results

Run ID: Indicates the indication code of the system run result. The first part denotes the group ID.
Anchor Info: Indicates how anchor text information is used for searching and ranking.
Link Info: Indicates how link information is used for searching and ranking.
Cont Info: Indicates how content information is used for searching and ranking.
WRR: Indicates weighted reciprocal rank with cut-off at 10.
DCG: Indicates discounted cumulative gain with cut-off at 10.

Run ID	Anchor Info	Link Info	Cont. Info	WRR		DCG		
				(1, 0)	(1, 1)	(3, 0)	(3, 2)	(3, 3)
TNT-1	indexing document	PageRank	full text	0.563402	0.635335	2.294993	2.71621	2.926819
TNT-2	indexing document	PageRank	(no)	0.560248	0.633276	2.310088	2.709968	2.909908
K3100-7	indexing link	site in-link analysis	(no)	0.396163	0.481281	1.606749	1.945853	2.115404
K3100-9	indexing link	in-link analysis	(no)	0.391254	0.483519	1.637614	1.992713	2.170262
K3100-11	indexing link	site in-link analysis	(no)	0.38815	0.474401	1.616461	1.965038	2.139327
K3100-14	indexing link	in-link analysis	(no)	0.370791	0.451512	1.497412	1.811604	1.968700
K3100-6	indexing link	in-link analysis	(no)	0.369794	0.449984	1.489859	1.801403	1.957174
K3100-8	indexing link	in-link analysis	(no)	0.366614	0.453200	1.52139	1.846183	2.008579
K3100-13	indexing link	in-link analysis	(no)	0.365376	0.443401	1.491671	1.806268	1.963566
K3100-10	indexing link	in-link analysis	(no)	0.362611	0.450319	1.512988	1.844902	2.010859
ORGREF-C20-P2	indexing link	in-link analysis	(no)	0.340245	0.392794	1.500649	1.834971	2.002132
ORGREF-C20-P1	indexing link	in-link analysis	(no)	0.329762	0.382683	1.469356	1.809488	1.979555
ORGREF-DR-LF2	indexing document	out-link analysis	full text	0.216763	0.256607	0.907171	1.164596	1.293308
ORGREF-DR-LF1	indexing document	out-link analysis	full text	0.212611	0.253163	0.880983	1.123083	1.244133
ORGREF-AR-LF1	indexing document	out-link analysis	anchor element	0.207770	0.235975	0.823229	0.980588	1.059268
ORGREF-AR-LF2	indexing document	out-link analysis	anchor element	0.205589	0.236713	0.838122	1.011278	1.097856
ORGREF-R-LF1	indexing document	out-link analysis	(no)	0.186502	0.209586	0.788666	0.942205	1.018975
ORGREF-R-LF2	indexing document	out-link analysis	(no)	0.183997	0.208820	0.776957	0.938099	1.018671
OKSAT-WEB-F-07	indexing link	in-link analysis	full text	0.157633	0.200119	0.839731	1.170654	1.336116
OKSAT-WEB-F-09	indexing link	site PageRank	full text	0.155719	0.196909	0.794031	1.113906	1.273843
OKSAT-WEB-F-11	indexing link	PageRank	full text	0.142564	0.192375	0.767063	1.080034	1.236520
OKSAT-WEB-F-06	indexing link	in-link analysis	full text	0.132411	0.176910	0.72798	1.091934	1.273910
OKSAT-WEB-F-08	indexing link	site PageRank	full text	0.122029	0.170772	0.669778	1.025856	1.203895
OKSAT-WEB-F-10	indexing link	PageRank	full text	0.114206	0.164634	0.644963	0.984571	1.154375
ORGREF-DR-LB2	indexing document	in-link analysis	full text	0.078157	0.112558	0.396406	0.707989	0.863781
ORGREF-R-LB2	indexing document	in-link analysis	(no)	0.067187	0.111139	0.383187	0.737005	0.913913
ORGREF-C20-S2	indexing link	in-link analysis	(no)	0.064858	0.077636	0.298495	0.407049	0.461325
ORGREF-AR-LB2	indexing document	in-link analysis	anchor element	0.043149	0.067390	0.272105	0.516234	0.638299
TNT-3	indexing document	(no)	full text	0.585791	0.645240	2.401664	2.814922	3.021552
TNT-4	indexing document	(no)	(no)	0.581489	0.642297	2.429589	2.819776	3.014870
K3100-12	indexing link	(no)	(no)	0.393260	0.472778	1.582700	1.918424	2.086285
K3100-1	indexing link	(no)	(no)	0.390765	0.461257	1.583006	1.921605	2.090904
K3100-4	indexing link	(no)	(no)	0.351624	0.427316	1.464084	1.797185	1.963735
K3100-5	indexing link	(no)	(no)	0.349265	0.426480	1.402175	1.697344	1.844929
K3100-3	indexing link	(no)	(no)	0.348672	0.425887	1.401370	1.696539	1.844124
K3100-2	indexing link	(no)	(no)	0.348278	0.421775	1.410705	1.716097	1.868792
ORGREF-R	indexing document	(no)	(no)	0.255089	0.308829	1.206347	1.552115	1.725000
ORGREF-GA1	indexing document	(no)	(no)	0.232751	0.275956	1.077592	1.415049	1.583777
ORGREF-DR	indexing document	(no)	full text	0.208910	0.256489	1.009247	1.339523	1.504662
OKSAT-WEB-F-05	indexing link	(no)	full text	0.148553	0.192800	0.775163	1.127437	1.303574
OKSAT-WEB-F-03	indexing link	(no)	full text	0.143524	0.188055	0.759515	1.090453	1.255922

Run ID	Anchor Info	Link Info	Cont. Info	WRR		DCG		
				(1, 0)	(1, 1)	(3, 0)	(3, 2)	(3, 3)
ORGREF-AR	indexing document	(no)	anchor element	0.142845	0.196072	0.766389	1.114841	1.289066
OKSAT-WEB-F-04	indexing link	(no)	full text	0.118319	0.165361	0.652176	1.054150	1.255138
OKSAT-WEB-F-02	indexing link	(no)	full text	0.116652	0.162582	0.650996	1.014602	1.196405
ORGREF-GC1-LF1	(no)	out-link analysis	full text	0.254303	0.317984	0.99885	1.431530	1.647870
ORGREF-GC1-LF2	(no)	out-link analysis	full text	0.251133	0.317572	0.995999	1.454977	1.684466
ORGREF-A-LF2	(no)	out-link analysis	anchor element	0.200236	0.235800	0.819132	0.995759	1.084073
ORGREF-A-LF1	(no)	out-link analysis	anchor element	0.198006	0.233198	0.806133	0.997064	1.092530
ORGREF-D-LF2	(no)	out-link analysis	full text	0.174130	0.205635	0.773949	0.998296	1.110469
ORGREF-D-LF1	(no)	out-link analysis	full text	0.166628	0.194526	0.725894	0.945375	1.055115
ORGREF-GC1-LB2	(no)	in-link analysis	full text	0.068964	0.157391	0.411909	0.989661	1.278537
ORGREF-D-LB2	(no)	in-link analysis	full text	0.064053	0.096963	0.312760	0.578341	0.711132
JSWEB-2	(no)	in-link analysis	full text	0.020818	0.037138	0.070272	0.161821	0.207596
JSWEB-4	(no)	in-link analysis	full text	0.020818	0.040395	0.070272	0.188712	0.247932
ORGREF-A-LB2	(no)	in-link analysis	anchor element	0.013573	0.042766	0.105639	0.305277	0.405096
TNT-5	(no)	(no)	full text	0.080069	0.116153	0.381351	0.664939	0.806732
ORGREF-GC1	(no)	(no)	full text	0.058186	0.124882	0.312234	0.770008	0.998895
OKSAT-WEB-F-00	(no)	(no)	full text	0.043947	0.092764	0.247505	0.658076	0.863361
OKSAT-WEB-F-01	(no)	(no)	full text	0.043204	0.088458	0.242958	0.624698	0.815568
ORGREF-D	(no)	(no)	full text	0.041454	0.066975	0.194908	0.400849	0.503820
ORGREF-A	(no)	(no)	anchor element	0.014013	0.031413	0.085743	0.215275	0.280042
JSWEB-1	(no)	(no)	full text	0.001157	0.010264	0.010233	0.062159	0.088122
JSWEB-3	(no)	(no)	full text	0.001157	0.010176	0.010233	0.083852	0.120662

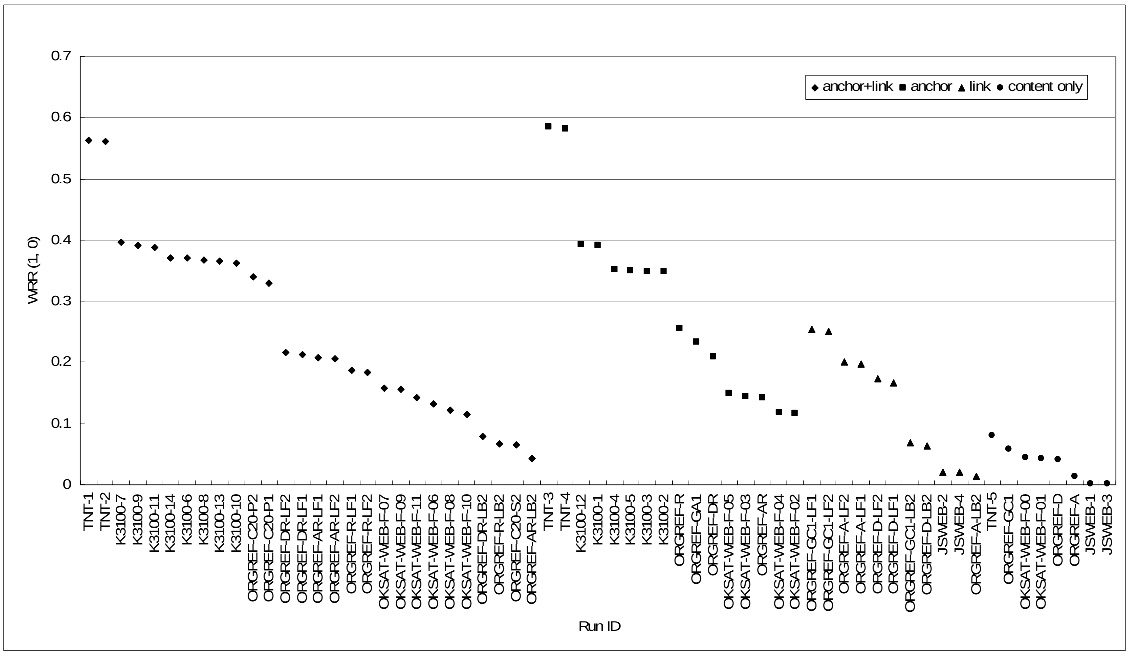

Figure 1. WRR values of run results for (δ_a, δ_b) = (1, 0).

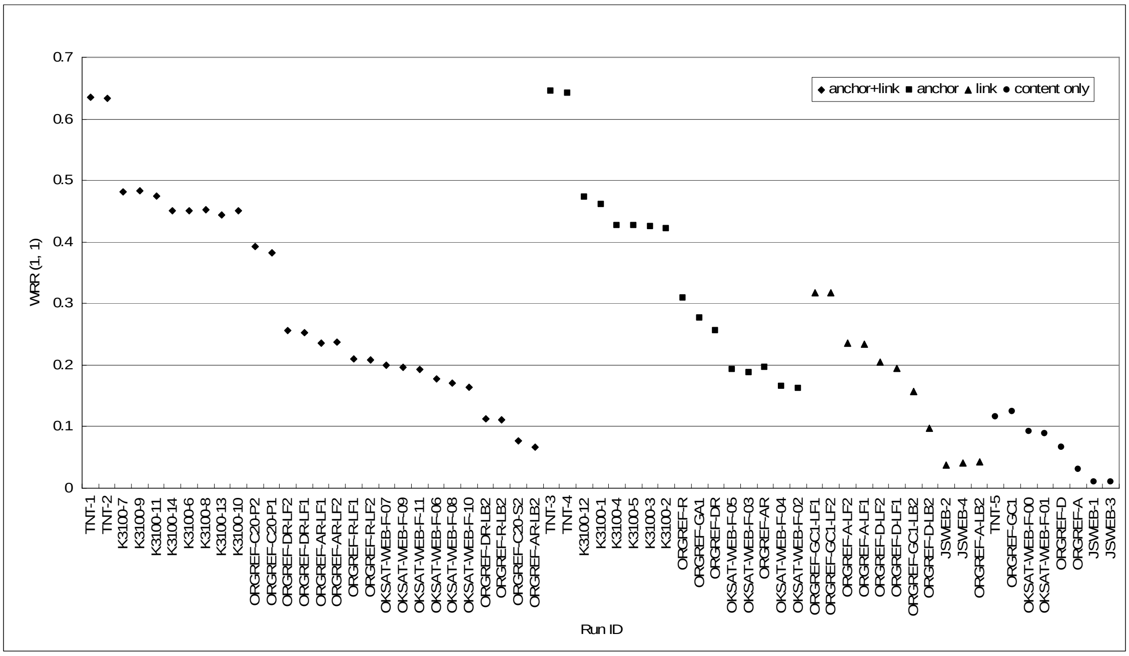

Figure 2. WRR values of run results for (δ_a, δ_b) = (1, 1).

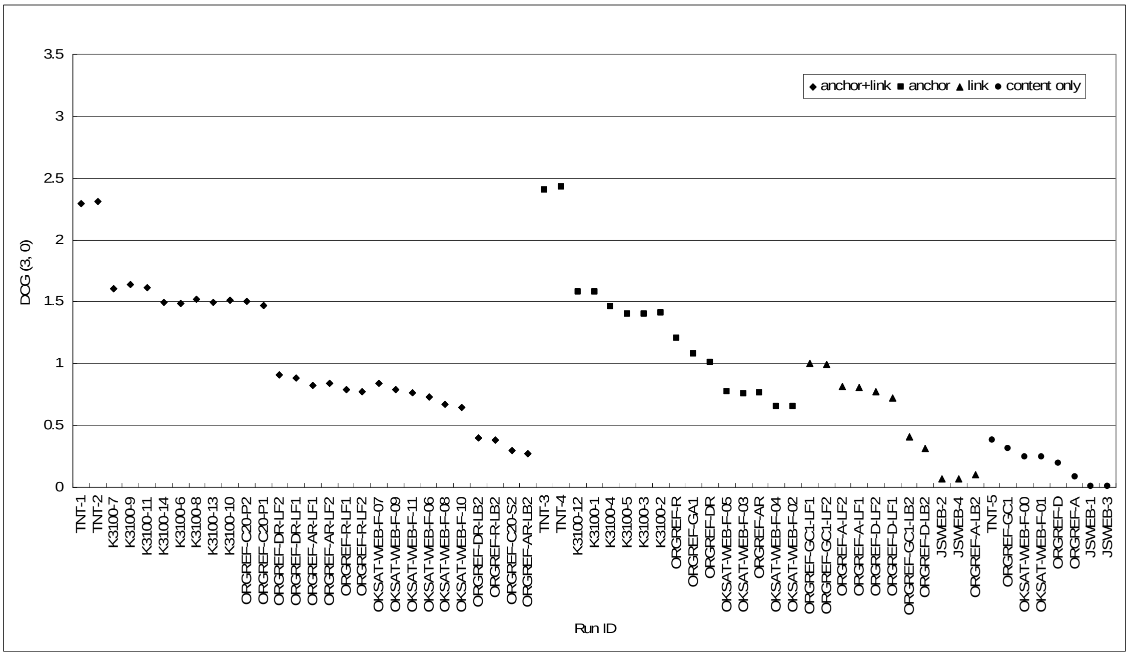

Figure 3. DCG values of run results for (G_a, G_b) = (3, 0).

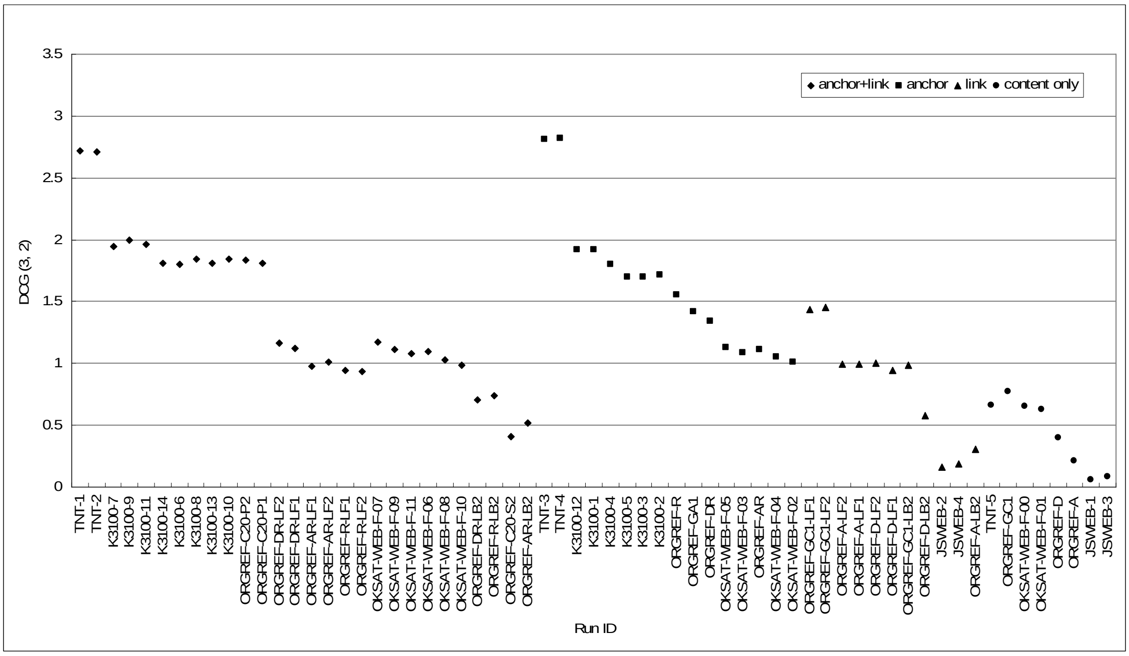

Figure 4. DCG values of run results for (G_a, G_b) = (3, 2).

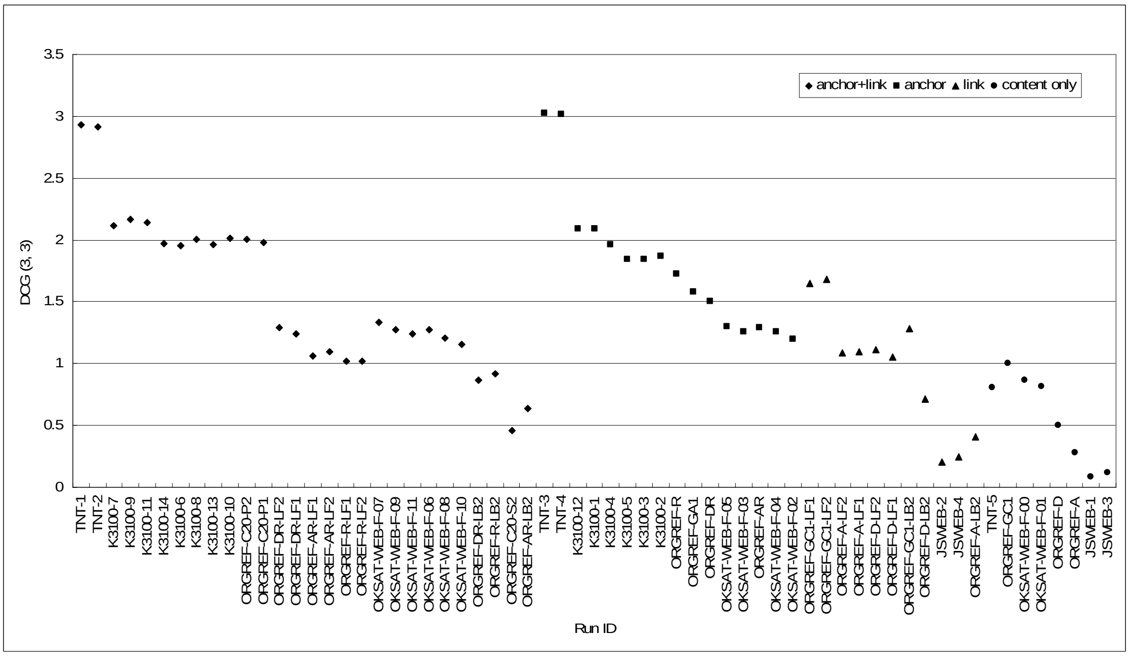

Figure 5. DCG values of run results for (G_a, G_b) = (3, 3).

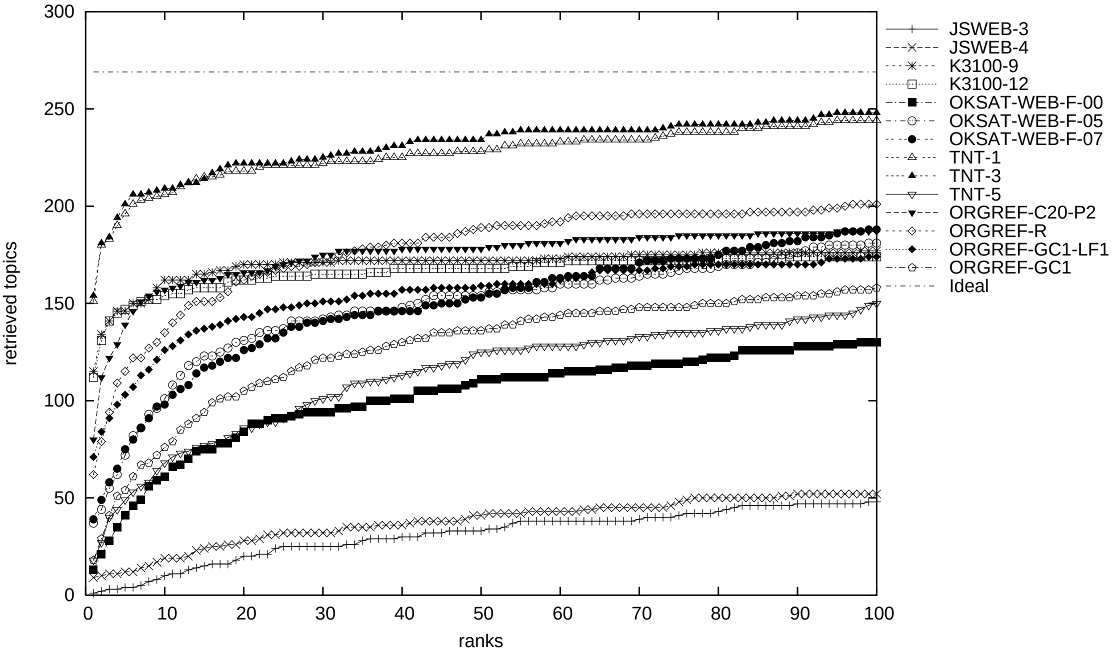

Figure 6. Cumulative number of topics whose relevant documents were retrieved (rigid relevance level).

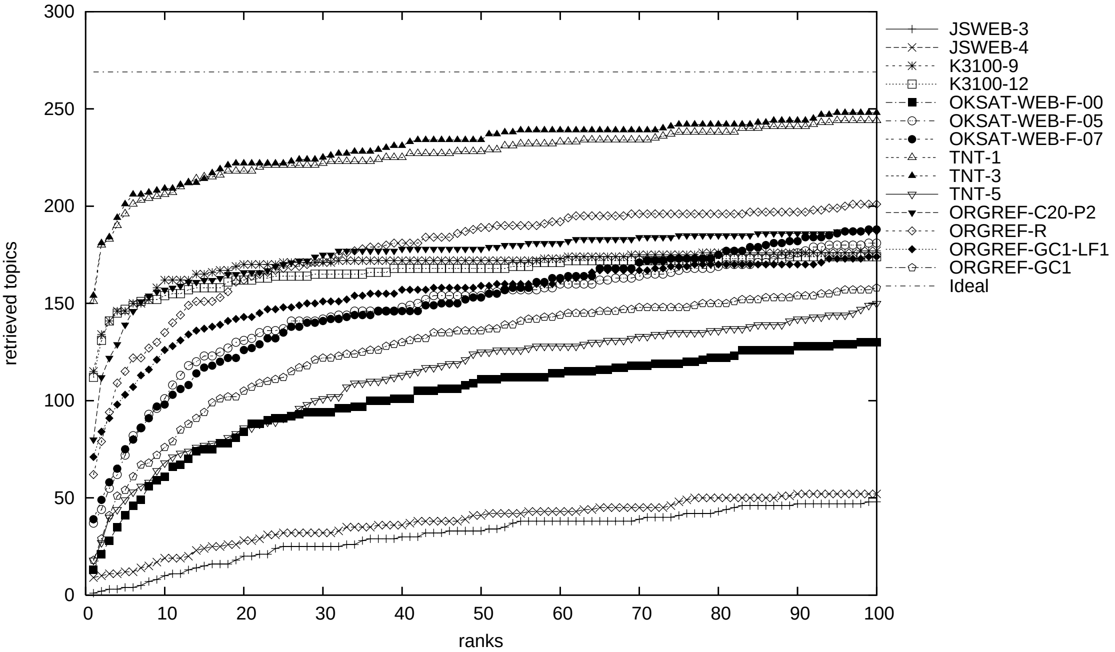

Figure 7. Cumulative number of topics whose relevant documents were retrieved (relaxed relevance level).

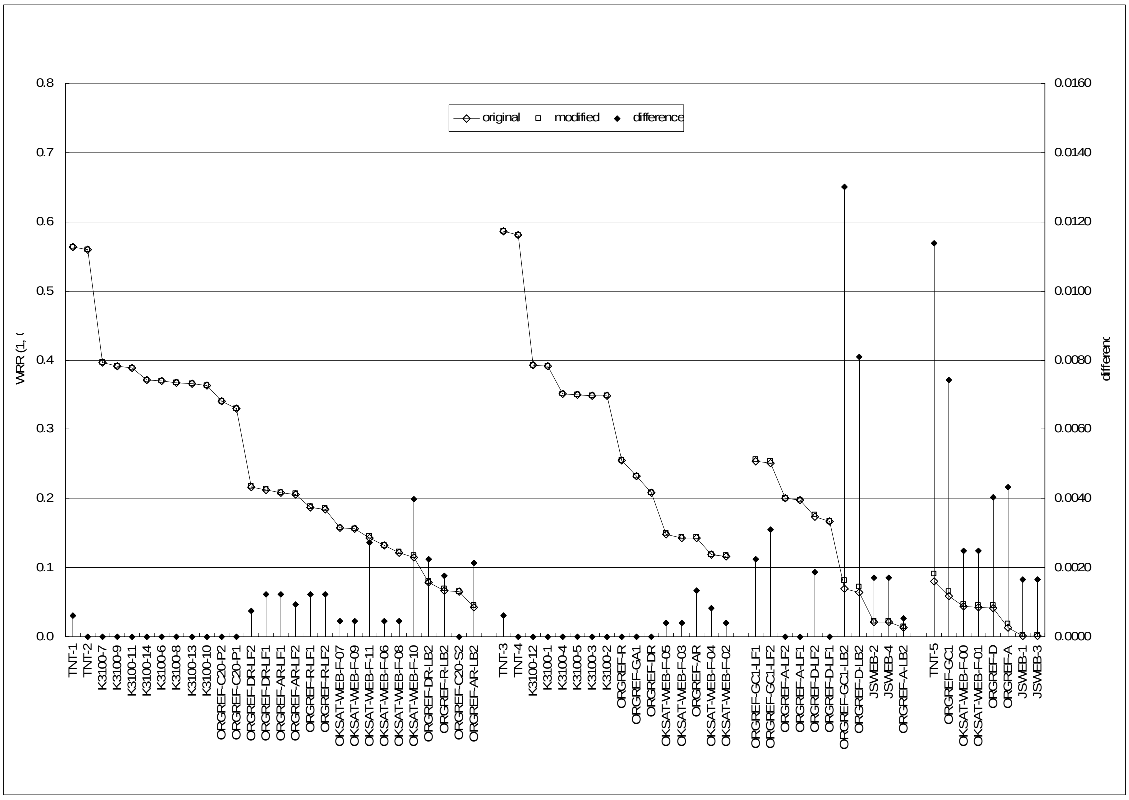

Figure 8. Effect of frameset pages on system evaluation with WRR (1, 0).

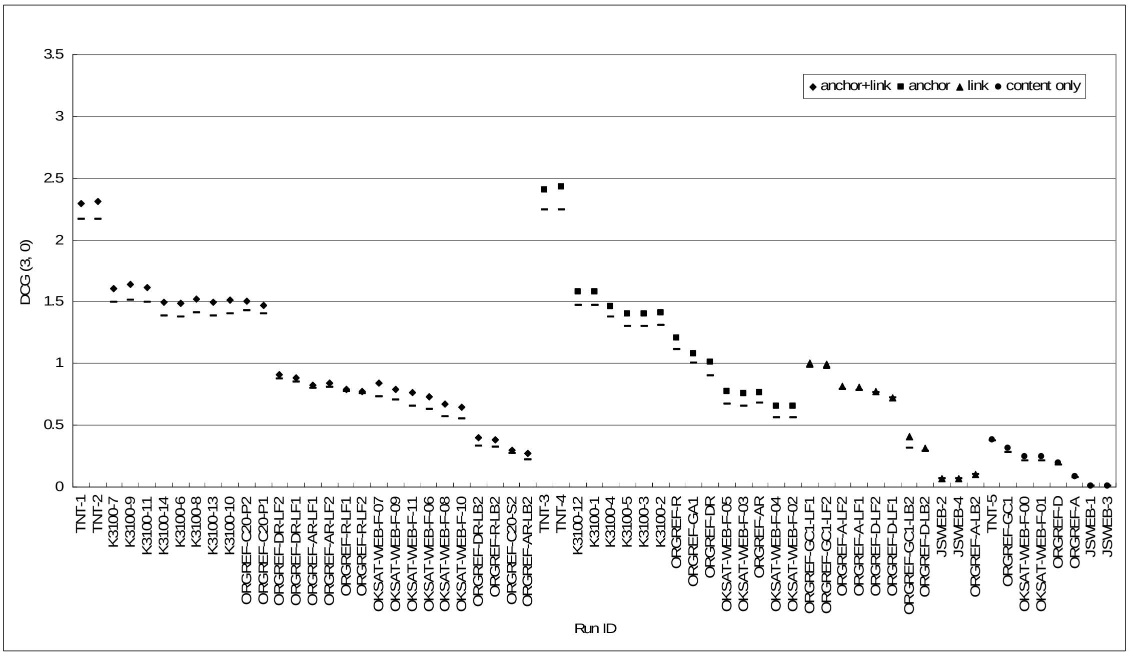

Figure 9. Effect of duplicate pages on system evaluation with DCG (3, 0, 0).

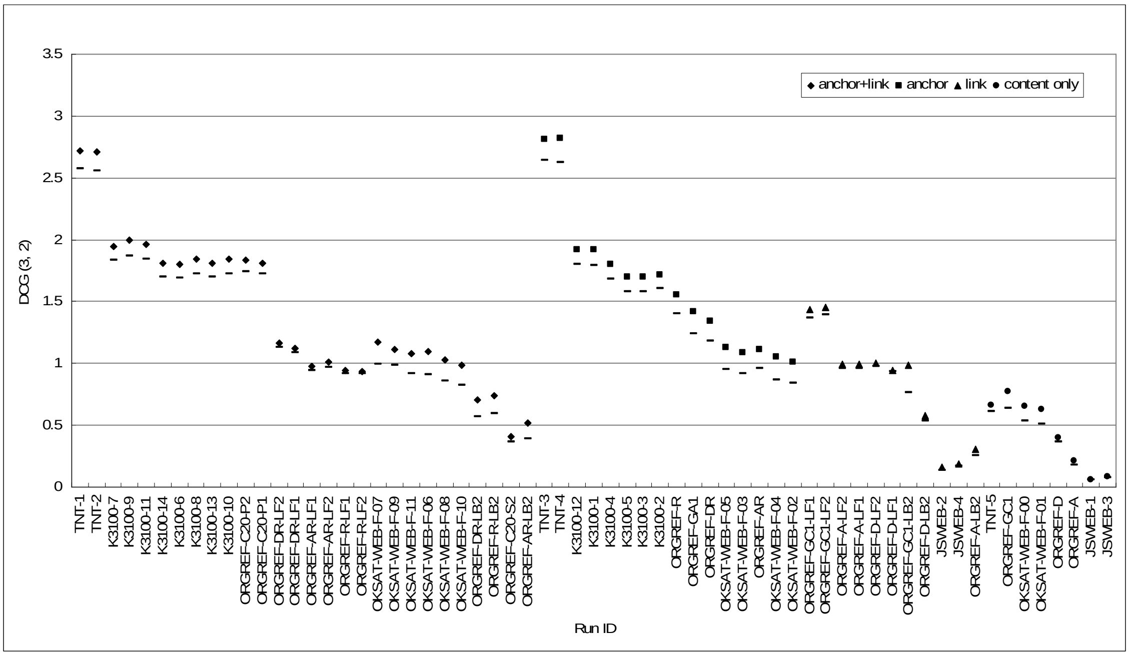

Figure 10. Effect of duplicate pages on system evaluation with DCG (3, 2, 0).

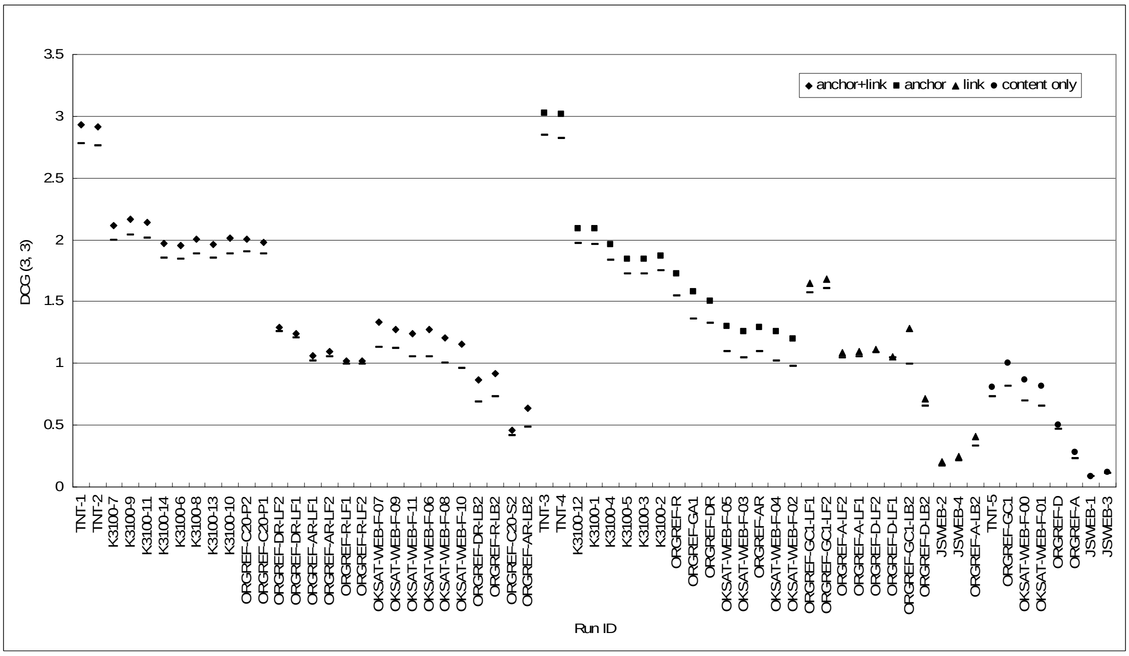

Figure 11. Effect of duplicate pages on system evaluation with DCG (3, 3, 0).

Appendix. Description of organizers' runs

The organizers executed 28 runs to expand the pool and to collect more relevant and partially relevant documents as well as to attempt various combinations of content-based, link-based and anchor-text-based techniques. These runs have run IDs with prefix "ORGREF-".

(1) Content-based systems

ORGREF-GC1:

A base line system for content-based systems using GETA search engine. IR model is probabilistic and ranking method is OKAPI BM-25. Text part indexed is full text. Topic part used is TITLE only. Index unit and query unit are both single word. No query expansion or relevance feedback is used.

ORGREF-D:

A base line system for content-based systems using Opentext Livelink 8 full text search engine. IR model is Boolean and ranking method is abbreviated tf-idf. Text part indexed is full text. Topic part used is TITLE only. Index unit is single character for Japanese characters and single word for alphanumeric. Query unit is phrase as is given in TITLE element. No query expansion or relevance feedback is used.

ORGREF-A:

An experimental system using Opentext Livelink 8 full text search engine. IR model is Boolean and ranking method is abbreviated tf-idf. Text part indexed is anchor element only. Topic part used is TITLE only. Index unit is single character for Japanese characters and single word for alphanumeric. Query unit is phrase as is given in TITLE element. No query expansion or relevance feedback is used.

(2) Anchor text (virtual document)

ORGREF-GA1:

A base line system for anchor-text-based systems using GETA search engine. IR model is probabilistic and ranking method is OKAPI BM-25. Text part indexed is anchor texts of in-links concatenated as a single text. Topic part used is TITLE only. Index unit and query unit are both single word. No query expansion or relevance feedback is used.

ORGREF-R:

A base line system for anchor-text-based systems using Opentext Livelink 8 full text search engine. IR model is Boolean and ranking method is abbreviated tf-idf. Text part indexed is anchor texts of in-links concatenated as a single text. Topic part used is TITLE only. Index unit is single character for Japanese characters and single word for alphanumeric. Query unit is phrase as is given in TITLE element. No query expansion or relevance feedback is used.

(3) Combination of content and anchor text

ORGREF-DR:

An experimental system for combining partial text and anchor-text index using Opentext Livelink 8 full text search engine. IR model is Boolean and ranking method is abbreviated tf-idf. Text part indexed is anchor element in the content and anchor texts of in-links concatenated as a single text. Topic part used is TITLE only. Index unit is single character for Japanese characters and single word for alphanumeric. Query unit is phrase as is given in TITLE element. No query expansion or relevance feedback is used.

ORGREF-AR:

An experimental system for combining full text and anchor-text index using Opentext Livelink 8 full text search engine. IR model is Boolean and ranking method is abbreviated tf-idf. Text part indexed is full text and anchor texts of in-links concatenated as a single text. Topic part used is TITLE only. Index unit is single character for Japanese characters and single word for alphanumeric. Query unit is phrase as is given in TITLE element. No query expansion or relevance feedback is used.

(4) Expansion with forward link

ORGREF-GC1-LF1:
ORGREF-GC1-LF2:
ORGREF-D-LF1:
ORGREF-D-LF2:
ORGREF-A-LF1:
ORGREF-A-LF2:
ORGREF-R-LF1:
ORGREF-R-LF2:
ORGREF-DR-LF1:

ORGREF-DR-LF2:
ORGREF-AR-LF1:
ORGREF-AR-LF2:

Experimental systems for collecting frequently referenced pages from pages retrieved by content-based systems corresponding to the prefix part of run IDs before "-LF1" and "-LF2". Score of a page *d* is calculated with the following equation:

$$Score_k(d) = \sum_{s \in R} wf_k(s,d),$$

$$wf_1(s,d) = \begin{cases} 1 & \text{if link from } s \text{ to } d \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

$$wf_2(s,d) = \begin{cases} \sigma(s) & \text{if link from } s \text{ to } d \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

where *R* is a content-base retrieval document set and $\sigma(s)$ is a score of the retrieved document *s*.

(5) Expansion with backward link

ORGREF-GC1-LB2:
ORGREF-D-LB2:
ORGREF-A-LB2:
ORGREF-R-LB2:
ORGREF-DR-LB2:
ORGREF-AR-LB2:

Experimental systems for collecting frequently referenced pages from pages retrieved by content-based systems corresponding to the prefix part of run IDs before "-LB2". Score of a page *d* is calculated with the following equation:

$$Score_k(d) = \sum_{d \in R} wb_k(s,d),$$

$$wb_2(s,d) = \begin{cases} \sigma(d) & \text{if link from } s \text{ to } d \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

where *S* is a content-base retrieval document set and $\sigma(d)$ is a score of the retrieved document *d*.

(6) Extended anchor text and link analysis

ORGREF-C20-P1:
ORGREF-C20-P2:
ORGREF-C20-S2:

Experimental systems using anchor text and its left and right context for selecting hyperlinks, and link structure for scoring linked pages. Using the same system as ORGREF-D, links are first searched with each query term *t* within 20 index units before and after distinct anchor texts. Then, score of a page *d* is calculated with the following equation:

$$p_1(d) = \prod_t L_{d,t}$$

$$p_2(d) = \prod_t \log(L_{d,t} + 1)$$

$$s_2(d) = \sum_t \log(L_{d,t} + 1) / \log(L_{d,*} + 1) \cdot \log(N / D_t)$$

where $L_{d,t}$ is a number of *d*'s in-links searched with *t*, $L_{d,*}$ is a number of all *d*'s in-links, D_t is a number of documents searched with *t*, and *N* is a number of all documents in the document set.