WIDIT in TREC-2003 Web Track

Kiduk Yang
School of Library and Information Science
Indiana University
Bloomington, Indiana 47405, U.S.A.
kiyang@indiana.edu

1. Introduction

The Web IR experiment of TREC, otherwise known as the Web track, investigated in its initial stages the strategies for the same ad-hoc retrieval task as was done previously with plain text documents. Although many TREC participants explored methods of leveraging non-textual sources of information such as hyperlinks and document structure, the general consensus among the early Web track participants was that link analysis and other non-textual methods did not perform as well as the content-based retrieval methods fine-tuned over the years (Hawking et al., 1999; Hawking et al., 2000; Gurrin & Smeaton, 2001; Savoy & Rasolofo, 2001).

There have been many speculations as to why link analysis, which showed much promise in previous research and has been so readily embraced by commercial Web search engines, did not prove useful in Web track experiments. Most such speculations point to potential problems with Web track's earlier test collections, from the inadequate link structure of truncated Web data (Savoy & Picard, 1998; Singhal & Kazkiel, 2001), and relevance judgments that penalize the link analysis by not counting the hub pages as relevant (Voorhees & Harman, 2000) and boost the content analysis by counting multiple relevant pages from the same site as relevant (Singhal & Kazkiel, 2001), to unrealistic queries that are too detailed and specific to be representative of real world Web searches (Singhal & Kaszkiel, 2001).

In an effort to address the criticism and problems associated with the early Web track experiments, TREC abandoned the ad-hoc Web retrieval task in 2002 in favor of topic distillation and named page finding task and replaced its earlier Web test collection of randomly selected Web pages with a larger and potentially higher quality domain-specific collection[1]. The topic distillation task in TREC-2002 is described as finding a short, comprehensive list of pages that are good information resources, and the named page finding tasks is described as finding a specific page whose name is described by the query (Hawking & Craswell, 2002; Craswell & Hawking, 2003). Adjustment of the Web track environment brought forth renewed interest in retrieval approaches that leverage Web-specific sources of evidences such as link structure and document structure.

For the home page finding task, where the objective is to find the entry page of a specific site described by the query, Web page's URL characteristics, such as its type and length, as well as the anchor text of Web page's inlinks proved to be useful sources of information to be leveraged (Hawking & Craswell, 2002). In the named page finding task, which is similar to home page finding task except that the target page described by the query is not necessarily the entry point of a Web site but any specific page on the Web, the use of anchor text still proved to be an effective strategy but the use of URL characteristics did not work well as it did in the home page finding task (Craswell & Hawking, 2003). In the topic distillation task, anchor text still seemed to be a useful resource, especially as a mean to boost the performance of content-based methods via fusion (i.e. result merging), although the level of its usefulness fell much below that achieved in named page finding tasks (Hawking & Craswell, 2002; Craswell & Hawking, 2003). Various site compression strategies, which attempt to select the "best" pages of a given site, was another common theme in the topic distillation task, once again demonstrating the importance of fine-tuning the retrieval system according to the task at hand (Amitay et al., 2003; Zhang et al., 2003). It is interesting to note that link analysis (e.g. PageRank, HITS variations) has not yet proven itself to be an effective strategy and the content-based method seems to be still the most dominant factor in the Web track. In fact, the two best results in TREC-2002 topic distillation task were achieved by the baseline systems that used only the content-based methods (MacFarlane, 2003; Zhang et al., 2003).

---

[1] Current test collection of the Web track (i.e. .GOV) consists of 1.25 million Web pages (19 gigabytes) from .gov domain, which is larger, less diverse and likely to be of higher quality than the previous collection (i.e. WT10g), which was a 10 gigabyte subset of the Web crawl from Internet Archive.

In our earlier studies (Yang, 2002a; Yang, 2002b), where we investigated various fusion approaches for ad-hoc retrieval using the WT10g collection, we found that simplistic approach that combine the results of content- and link-based retrieval results did not enhance retrieval performance in general. TREC participants in recent Web track environment, however, found that use of non-textual information such as hyperlinks, document structure, and URL could be beneficial for specific tasks such as topic distillation and named/home page finding tasks. We believe that this is not only due to the change in the retrieval environment (i.e. test collection, retrieval task) but also the result of more dynamic approach to combining multiple sources of evidence than straightforward result merging. Thus, our focus in TREC-2003 Web track was in exploring fusion strategies that utilize various information sources in a dynamic manner to optimize retrieval for specific search environment. For our experiment, we used the experimental fusion retrieval system called WIDIT[2] to combine content and link information, and then reranked the combined result based on heuristics arrived at from dynamic system tuning process.

2. WIDIT

Basic approach of WIDIT in the Web track consisted of four main phases: indexing, searching, result merging, and reranking. Indexing phase involved indexing various sources of evidence to generate multiple indexes, which was followed by the searching phase that produced multiple result sets from using different query formulations against multiple indexes. The result sets were combined using weighted sum formula, after which a reranking heuristics were applied to optimize the ranking of the merged results. The overview of WIDIT system architecture is displayed in Figure 1.

2.1 Indexing Module

WIDIT preprocessed documents by removing HTML tags and stopwords and applying the simple plural remover (Frakes & Baeza-Yates, 1992)[3]. The stopwords consisted of non-meaningful words such as words in a standard stopword list, non-alphabetical words, words consisting of more than 25 or less than 3 characters, and words that contain 3 or more repeated characters. Hyphenated words were split into parts before applying the stopword exclusion, and acronyms and abbreviations were kept as index terms[4].

In addition to extracting body text terms (i.e. terms between <body> and </body> tags), WIDIT extracted terms from document title, meta keywords and descriptions, and "emphasized" text (e.g. text with <b>, , <font>, <u>, <h1> tags) as well as extracting terms from the anchor texts of incoming links. Thus, WIDIT created three sets of term indexes: first based on document content (i.e. body index), second based on document structure (header index), and third based on link structure (anchor index).

In order to enable incremental indexing as well as to scale up to larger collections, each of the indexes consisted of smaller subcolllections, which were created and searched in parallel. The whole collection term statistics were derived after the creation of the subcollections.

2.2 Retrieval Module

The retrieval component of WIDIT was based on a Vector Space Model (VSM) using the SMART length-normalized term weights as was implemented in IRIS (Yang & Maglaughlin, 2000). Documents were ranked in decreasing order of the inner product of document and query vectors,

$$\mathbf{q}^{\mathrm{T}}\mathbf{d}_i = \sum_{k=1}^{t} q_k d_{ik}, \qquad (1)$$

---

[2] WIDIT (Web Information Discovery Integrated Tool; http://widit.slis.indiana.edu/), which extends IRIS research (http://ils.unc.edu/iris/) at the University of North Carolina, is an experimental IR system with a suite of modular retrieval tools designed for easy integration of multiple Web IR approaches. WIDIT is currently being developed in the School of Library and Information Science at Indiana University.

[3] The simple plural remover was chosen to speed up indexing time and to minimize the overstemming effect of more aggressive stemmers.

[4] Acronym and abbreviation identification was based on simple pattern matching of punctuations and capitalizations.

where q_k is the weight of term k in the query, d_{ik} is the weight of term k in document i, and t is the number of terms in the index. SMART *Lnu* weights with the slope of 0.3 were used for document terms (Buckley et al., 1997), and SMART *ltc* weights (Buckley et al., 1995) were used for query terms. *Lnu* weights attempt to match the probability of retrieval given a document length with the probability of relevance given that length (Singhal et al., 1996).

Two sets of queries, one resulting from simple stop and stemming, and another with phrases, acronyms and abbreviations extracted, were applied against three sets of document indexes[5] to produce six sets of retrieval results[6].

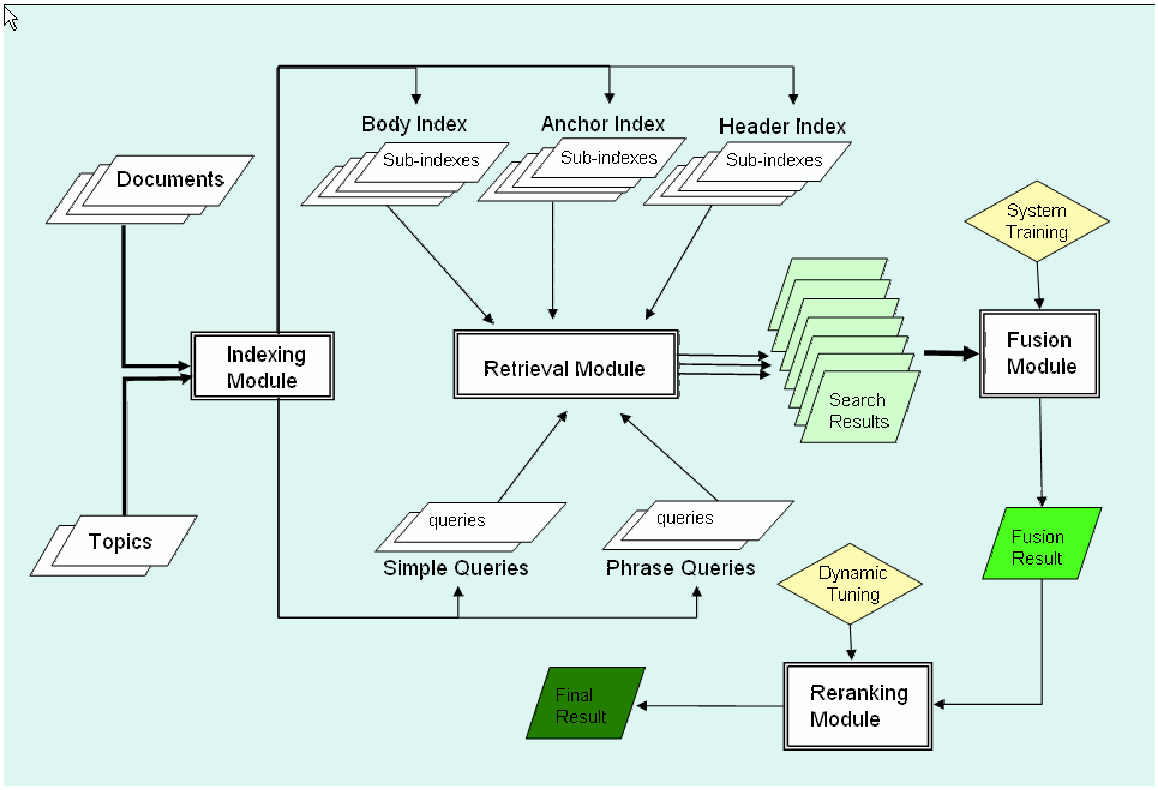

Figure 1. WIDIT System Architecture

2.3 Fusion Module

In post-retrieval fusion, where multiple sets of search results are combined after retrieval time, two of the most common fusion formulas are *Similarity Merge* (Fox & Shaw, 1995; Lee, 1997) and *Weighted Sum* (Bartell et al., 1994; Thompson, 1990). The similarity merge formula multiplies the sum of fusion component scores for a document by the number of fusion components that retrieved the document (i.e. overlap), based on the assumption that documents with higher overlap are more likely to be relevant. Instead of relying on overlap, the weighted sum formula sums fusion component scores weighted with the relative contributions of the fusion components that retrieved them, which is typically estimated based on training data. Both formulas compute the fusion score of a document by a linear combination of fusion component scores.

---

[5] Body text index consisted of title and body text terms. Anchor text index consisted of title and inlink anchor text terms. Header text index consisted of title, meta keywords and descriptions, and emphasized text terms.

[6] In practice, retrieval for each document index consisted of parallel searches of 46 subcollections using the whole collection term weights, whose results were merged and sorted by document score.

In our earlier study (Yang, 2002b), similarity merge approach proved ineffective when combining content- and link-based results, so this time we tried three variations of the weighted sum fusion formula, which were shown to be more effective in combining fusion components that are dissimilar (Yang, 2002a). Equation (2) describes the simple *Weight Sum* (WS) formula, which sums the normalized system scores multiplied by system contribution weights. Equation (3) describes the *Overlap Weight Sum* (OWS) formula, which multiplies the WS score by overlap. Equation (4) describes the *Weighted Overlap Weighted Sum* (WOWS) formula, which multiplies the WS score by overlap weighted by system contributions:

$$FS_{WS} = \sum(w_i * NS_i), \quad (2)$$

$$FS_{OWS} = \sum(w_i * NS_i * olp), \quad (3)$$

$$FS_{WOWS} = \sum(w_i * NS_i * w_i * olp), \quad (4)$$

where: FS = fusion score of a document,
w_i = weight of system i,
NS_i = normalized score of a document by system i,
$= (S_i - S_{min}) / (S_{max} - S_{min})$
olp = number of systems that retrieved a given document.

The normalized document score, NS_i, is computed by Lee's min-max formula (1996, 1997), where S_i is the retrieval score of a given document and S_{max} and S_{min} are the maximum and minimum document scores by method i.

One of the main challenges in using the weighted fusion formula lies in determination of the optimum weights for each system (w_i). We assessed various weight combinations[7] (e.g. 0.9 for body text, 0.08 for header text, 0.01 for anchor text) with the training data of past Web track results to tune our fusion module.

2.4 Reranking Module

In order to optimize retrieval performance in top ranks, fusion results were reranked based on combinations of site compression technique and content-link evidence ranking heuristic. The site compression involved clustering results by sites, sorting sites by the highest document score of each site, and floating the top n documents from top m sites to the top ranks. The content-link evidence ranking heuristic consisted of a set of ranking and document score boosting rules arrived at by dynamic tuning process involving interactive retrieval and manual system tuning in real time. The dynamic tuning process was applied to the best single and best fusion systems to "tune" the ranking heuristic.

The dynamic tuning component of WIDIT produces retrieval results that display individual scores for each source of evidence such as inter/intrasite in/outdegree, phrase/proximity match counts in body/header/anchor texts, and query term matches in URL as well as ranking and retrieval scores before/after fusion and site compression. We performed a series of dynamic tuning sessions using training data, which involved repeated cycles of retrieval and tuning the reranking heuristic based on real time evaluation of retrieval results. In contrast to the static tuning of fusion formulas, dynamic tuning process, though ad-hoc, allows tuning of systems with numerous parameters by leveraging human intelligence. The main components of content-link evidence ranking heuristic we used were inter/intrasite in/outdegree (e.g. boost score if large outdegree for topic distillation, boost score if large indegree for home/named page finding), phrase/proximity match (e.g. boost ranking if phrase match in title or anchor text), and query term match in URL (e.g. boost to top 10 rank if acronym match in URL). The reranking heuristics for home/named page finding task also involved the query classification component, which assigned different emphasis on evidence sources according to the query type. The queries were classified into either named page or home page based on term occurrence patterns observed in training queries.

[7] Eight weight combinations for each fusion formula (24 systems per task) were examined.

2.5 Topic Distillation vs. Home/Named Page Finding Tasks

WIDT systems for topic distillation and home/named page finding runs shared the indexing and searching modules, but used different fusion formulas arrived at from different training data. The reranking module of topic distillation systems employed both site compression and content-link evidence reranking heurisitic, but home/named page finding systems used only the reranking heuristics that included the query classification component and emphasized the acronym matching.

The training data for topic distillation systems were 2002 topic distillation topics and relevance judgments, and home/named page finding system used 2002 named page finding topics and relevance judgments as training data. Unfortunately, both training data were problematic. The topic distillation task in TREC-2003 was about finding relevant "home pages" given a broad query, which introduced bias towards home page finding approaches not present in the training data. As the name indicates, home/named page finding task in TREC-2003 combined home page and named page finding tasks, only half of which were present in the training data used. In addition to suboptimal training data, the topic distillation systems were trained based on precision at rank 10 (P@10), whereas the main evaluation metric for topic distillation in TREC-2003 was changed to R-precision.

3. Results

In the topic distillation test runs using training data, the best single system was *widittdb1* (P@10 = 0.168), which used simple query and body text index. The best fusion runs, which used fusion weights of 0.9, 0.01, 0.09 for body, header, and anchor text respectively, improved the baseline (*widittdb1*) results by 8% (P@10 = 0.182). The reranking of fusion results improved the baseline results by 17% (P@10 = 0.196). In the named page finding test runs, same baseline system resulted in mean reciprocal rank (MRR) of 0.471, which fusion with weights 0.8, 0.19, and 0.01 improved by 10% (MRR = 0.520) but reranking failed to improve (MRR = 0.471).

The official run results were comparable to test run results in that fusion improved retrieval performance in both tasks and reranking further enhanced performance in topic distillation runs according to P@10. The official home/named page finding runs, which consisted of the baseline and the best fusion run were comparable to test runs in that the fusion run (MRR = 0.400) improved the baseline result by 10% (MRR = 0.362). The official topic distillation runs showed 21% improvement by fusion (P@10 = 0.092) and 29% improvement by reranking (P@10 = 0.098) over baseline (P@10 = 0.076).

3.1 TREC System Rankings

By official TREC system rankings, which ranked all TREC topic distillation runs by R-precision, WIDIT appeared to perform rather poorly when compared with other TREC systems (Table 1). By R-precision, the best WIDIT run, which was the baseline run, was ranked 71 of 107 systems (18 of 23 groups). When we reranked the systems using P@10, however, WIDIT runs were ranked much higher (Table 2). In fact, the best WIDIT run, which was the reranked fusion run, ranked 28 of 107 systems (12 of 23 groups). When ranked by mean average precision (Table 3), WIDIT ranked 38 of 107 systems (11 of 23 groups). Table 4 and 5, which shows system rankings for home/named page finding runs, reflects the poor performance level of WIDIT[8], which we tried to compensate for in post-submission runs described in the next section.

Table 1. Topic Distillation ranking by Mean R-Precision (MRP)

	MRP	MAP	avgP@10
Best TREC system	**0.1636**	0.1543	0.1240
Median TREC system	**0.0699**	0.0896	0.0700
Best WIDIT system	**0.0736**	0.1016	0.0760
Worst TREC system[9]	**0.0181**	0.0250	0.0160

[8] WIDIT fusion run ranked 48 of 75 systems (13 of 19 groups) by MRR, and 41 of 75 (13 of 19 groups) systems by mean success rate at rank 10.

[9] Outlier was excluded to keep the system performance comparisons in perspective.

Table 2. Topic Distillation ranking by Mean Precision at rank 10 (avgP@10)

	MRP	MAP	**avgP@10**
Best TREC system	0.1485	0.1387	**0.1280**
Best WIDIT system	0.0626	0.0787	**0.0980**
Median TREC system	0.0871	0.1057	**0.0807**
Worst TREC system	0.0181	0.0250	**0.0160**

Table 3. Topic Distillation ranking by Mean Average Precision (MAP)

	MRP	**MAP**	avgP@10
Best TREC system	0.1636	**0.1543**	0.1240
Best WIDIT system	0.0736	**0.1016**	0.0760
Median TREC system	0.0699	**0.0896**	0.0700
Worst TREC system	0.0230	**0.0222**	0.0200

Table 4. Home/Name Page ranking by Mean Reciprocal Rank (MRR)

	MRR	avgS@10
Best TREC system	**0.727**	89.3
Median TREC system	**0.496**	64.3
Best WIDIT system	**0.400**	66.3
Worst TREC system	**0.065**	8.7

Table 5. Home/Name Page ranking by Mean Success Rate at rank 10 (avgS@10)

	MRR	**avgS@10**
Best TREC system	0.727	**89.3**
Median TREC system	0.465	**68.3**
Best WIDIT system	0.400	**66.3**
Worst TREC system	0.065	**8.7**

3.2 Post-submission Runs

After receiving the results of the official TREC runs, we conducted post-submission analysis to discover and address the shortcomings of the submitted systems. In retrospect, it was easy to see the negative effects of improper system tuning that resulted from biased training data as well as system overtuning that penalized system rankings based on the different evaluation measure than one used in training. To compensate for these shortcomings, we conducted another cycle of dynamic tuning iterations that involved the adjustments in reranking and query classification heuristics and the implementation of home page identification method. The WIDIT dynamic tuning component was modified to produce ranking by both P@10 and R-precision, so that system tuning could be optimized for both evaluation measures.

Home page identification was based on URL typing (Tomlinson, 2003; Kraaij et al., 2002), where Web pages are classified into categories of root, subroot, path, and file. The root page was defined as URL with zero slash counts or URL that ends with home page file name (e.g. index.htm, default.htm) and 1 slash count. Subroot page was defined as home page ending with 2 slash count, and path page was defined as home page ending with 3 or more slash count. The file page was defined as URL that meets none of these conditions. Based on the observations from post-submission analysis, which suggested the strong performance of fusion results in top ranks as well as the importance of home page finding approaches, following rank boosting rules were added to the reranking heuristic:

- Keep top 5 ranks static
- Boost the rank of potential home pages (root, subroot, path)
- Boost the rank of file type page with two or more query terms in URL
- Stop if top 20 ranks are filled

The query classification heuristics for home/named page finding task was also improved by adding additional rules such as classifying queries that ends in all capitalized words as home page queries.

3.3 Post-submission Results

The post-submission topic distillation run results that introduced home page finding bias and tuned the system for both P@10 and R-precision improved system ranking by MRP from 71 to 23 (Table 6), system ranking by avgP@10 from 28 to 25 (Table 7), and system ranking by MAP from 38 to 10 (Table 8). The post-submission home/named page finding run results, however, were little different from official results.

Table 6. Post-submission Topic Distillation ranking by Mean R-Precision (MRP)

	MRP	MAP	avgP@10
Best TREC system	**0.1636**	0.1543	0.1240
Best WIDIT system	**0.1139**	0.1281	0.0980
Median TREC system	**0.0699**	0.0896	0.0700
Worst TREC system	**0.0181**	0.0250	0.0160

Table 7. Post-submission Topic Distillation ranking by Mean Precision at rank 10 (avgP@10)

	MRP	MAP	**avgP@10**
Best TREC system	0.1485	0.1387	**0.1280**
Best WIDIT system	0.0626	0.1216	**0.0981**
Median TREC system	0.0871	0.1057	**0.0807**
Worst TREC system	0.0181	0.0250	**0.0160**

Table 8. Post-submission Topic Distillation ranking by Mean Average Precision (MAP)

	MRP	**MAP**	avgP@10
Best TREC system	0.1636	**0.1543**	0.1240
Best WIDIT system	0.1139	**0.1281**	0.0980
Median TREC system	0.0699	**0.0896**	0.0700
Worst TREC system	0.0230	**0.0222**	0.0200

3. Discussion

In this year's topic distillation task, there were on the average 10.32 relevant documents per topic (5 topics with 1 relevant document, 3 topics with 2 relevant documents, 20 topics with 5 or fewer relevant documents, 33 topics with 10 or fewer relevant documents). The best WIDIT topic distillation run results had 17 topics with zero P@10, 23 topics with zero R-Precision, and only 4 topics with relevant documents at top 20 ranks. The key question for WIDIT topic distillation runs, therefore, is why they performed so poorly with topics with few relevant documents. Whether this is due to some WIDIT specific factors or it is a TREC system-wide phenomenon remains to be seen.

In the home/named page finding task, WIDIT exhibited suboptimal performance, which we attribute largely to incomplete training data (e.g. omission of home page finding training data) based on the following observation: the best WIDIT home/named page run had 24 home page topics with first correct answer beyond rank 100, compared to 6 named page topics with first correct answer beyond rank 100.

Overall, we believe fusion is a promising area of investigation for Web IR. Our results show that exploiting the richness of Web search environment by combining multiple sources of evidence via result merging and dynamic system tuning can enhance retrieval performance in the topic distillation task. As for the home/named page finding task, we suspect our approach was hampered by incomplete training data, which will be investigated in a follow-up study.

References

Amitay, E., Carmel, D., Darlow, A., Lempel, R., & Soffer, A. (2003). Topic Distillation with Knowledge Agents. *Proceedings of the11th Text Retrieval Conference (TREC 2002)*, 263-272.

Bartell, B. T., Cottrell, G. W., & Belew, R. K. (1994). Automatic combination of multiple ranked retrieval systems. *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*.

Buckley, C., Salton, G., & Allan, J., & Singhal, A. (1995). Automatic query expansion using SMART: TREC 3. *Proceeding of the 3rd Text Rerieval Conference (TREC-3)*, 1-19.

Buckley, C., Singhal, A., & Mitra, M. (1997). Using query zoning and correlation within SMART: TREC 5. *Proceeding of the 5th Text REtrieval Conference (TREC-5)*, 105-118.

Craswell,N., & Hawking, D. (2003). Overview of the TREC-2002 web track. *Proceedings of the 11th Text Retrieval Conference (TREC 2002)*, 86-95.

Frakes, W. B., & Baeza-Yates, R. (Eds.). (1992). *Information retrieval: Data structures & algorithms*. Englewood Cliffs, NJ: Prentice Hall.

Fox, E. A., & Shaw, J. A. (1995). Combination of multiple searches. *Proceeding of the3rd Text Rerieval Conference (TREC-3)*, 105-108.

Gurrin, C., & Smeaton, A.F. (2001). Dublin City University experiments in connectivity analysis for TREC-9. *Proceedings of the 9th Text Retrieval Conference (TREC-9)*, 179-188.

Hawking, D., & Craswell, N. (2002). Overview of the TREC-2001 web track. *Proceedings of the 10th Text Retrieval Conference (TREC 2001)*, 25-31

Hawking, D., & Craswell, N., Thistlewaite, P., & Harman, D. (1999). Results and challenges in web search evaluation. *Proceedings of the 8th WWW Conference*, 243-252.

Hawking, D, Voorhees, E, Craswell, N., & Bailey, P. (2000). Overview of the TREC-8 web track. *Proceedings of the 8th Text Retrieval Conference (TREC-8)*, 131-148.

Kraaij, W., Westerveld, T., & Hiemstra, D. (2002). The Importance of Prior Probabilities for Entry Page Search. *Proceedings of the 25th ACM SIGIR Conference on Research and Development in Information Retrieval*, 27-34

Lee, J. H. (1997). Analyses of multiple evidence combination. *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, 267-276.

Savoy, J., & Picard, J. (1998). Report on the TREC-8 Experiment: Searching on the Web and in Distributed Collections. *Proceedings of the 8th Text Retrieval Conference (TREC-8)*, 229-240.

Savoy, J., & Rasolofo, Y. (2001). Report on the TREC-9 experiment: Link-based retrieval and distributed collections. *Proceedings of the 9th Text Retrieval Conference (TREC-9)*, 579-516.

Singhal, A., Buckley, C., & Mitra, M. (1996). Pivoted document length normalization. *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, 21-29.

Singhal, A., & Kaszkiel, M. (2001). A case study in Web search using TREC algorithms. *Proceedings of the 11th International WWW Conference*, 708-716.

Thompson. P. (1990). A combination of expert opinion approach to probabilistic information retrieval, part 1: The conceptual model. *Information Processing & Management*, 26(3), 371-382.

Tomlinson, S. (2003). Robust, Web and Genomic Retrieval with Hummingbird SearchServer at TREC 2003. *The 12th Text REtrieval Conference (TREC 2003) Notebook*, 372-385.

Voorhees, E., & Harman, D. (2000). Overview of the Eighth Text Retrieval Conference. *Proceedings of the 8th Text Retrieval Conference (TREC-8)*, 1-24.

Yang, K. (2002a). Combining Text-, Link-, and Classification-based Retrieval Methods to Enhance Information Discovery on the Web. (*Doctoral Dissertation*. University of North Carolina).

Yang, K. (2002b). Combining Text- and Link-based Retrieval Methods for Web IR. *Proceedings of the 10th Text Rerieval Conference (TREC2001)*, 609-618.

Yang, K. & Maglaughlin, K. (2000). *IRIS at TREC*-8. *Proceedings of the 8th Text Rerieval Conference (TREC-8)*, 645-656.

Zhang, M., Song, R., Lin, C., Ma, S., Jiang, Z., Jin, Y., Liu, Y., & Zhao, L. (2003). THU TREC 2002: Web Track Experiments. *Proceedings of the 11th Text Retrieval Conference (TREC 2002)*, 591-594.