Report on the TREC-9 Experiment: Link-Based Retrieval and Distributed Collections

Jacques Savoy, Yves Rasolofo

Institut interfacultaire d'informatique
Université de Neuchâtel (Switzerland)
E-mail: {Jacques.Savoy, Yves.Rasolofo}@unine.ch
Web page: http://www.unine.ch/info/

Summary

The web and its search engines have resulted in a new paradigm, generating new challenges for the IR community which are in turn attracting a growing interest from around the world. The decision by NIST to build a new and larger test collection based on web pages represents a very attractive initiative. This motivated us at TREC-9 to support and participate in the creation of this new corpus, to address the underlying problems of managing large text collections and to evaluate the retrieval effectiveness of hyperlinks.

In this paper, we will describe the results of our investigations, which demonstrate that simple raw score merging may show interesting retrieval performances while the hyperlinks used in different search strategies were not able to improve retrieval effectiveness.

Introduction

Due to the huge number of pages and links, browsing cannot be viewed as an adequate searching process, even with the introduction of tables of contents or other classifying lists (e.g., Yahoo!). As a result, effective query-based mechanisms for accessing information will always be needed. Search engines currently available on the web are not able to adequately access all available information [Lawrence 99], as they are inhibited by many drawbacks [Hawking 99].

In the first chapter, we will describe our experiments on the web track in which a large web text collection is divided into four sub-collections in order to keep inverted file size below the 2 GB limit. The second chapter will verify whether or not hyperlinks improved retrieval effectiveness based on four different link-based search models.

To evaluate our hypothesis, we used the SMART system as a test bed for implementing the OKAPI probabilistic model [Robertson 95]. This year our experiments were conducted on an Intel Pentium III/600 (memory: 1 GB, swap: 2 GB, disk: 6 x 35 GB) and all experiments were fully automated.

1. Distributed collections

To evaluate the retrieval effectiveness of various merging strategies, we formed four separate sub-collections (see Appendix 1). In this study, we assumed that each sub-collection used the same indexing schemes and retrieval procedures. A distributed context such as this more closely reflects local area networks or search engines available on the Internet than the meta search engines, where different search engines may collaborate to respond to a given user request [Le Calvé 00], [Selberg 99].

The following characteristics would more precisely identify our approach. A query was sent to all four text databases (no selection procedure were applied) and according to the four ranked lists of items produced, our search system merged them into a single result list presented to the user (Section 1.2). Before we describe the collection merging approaches, Section 1.1 will identify retrieval effectiveness measures achieved by various search models with the whole collection and with each of our four sub-collections.

1.1. Performance of sub-collections

From the original web pages, we retained only the following logical sections: <TITLE>, <H1>, <CENTER>, <BIG>, with the most common tags <P> (or <p>, together with </P>, </p>) being removed. Text delimited by the tags <DOCHDR>, </DOCHDR> were also removed. For long requests, various insignificant keywords were also removed (such as "Pertinent documents should include …"). Moreover, search keywords appearing in the Title part of the topics were considered to have a term frequency of 3 (this feature has no impact on short requests).

For the web track, we conducted different experiments using the OKAPI probabilistic model in which the weight w_{ij} assigned to a given term t_j in a document D_i was computed according to the following formula:

$$w_{ij} = \frac{(k_1 + 1) \cdot tf_{ij}}{K + tf_{ij}}$$

$$\text{with} \quad K = k_1 \cdot \left[(1 - b) + b \cdot \frac{l_i}{avdl} \right]$$

where tf_{ij} indicates the within-document term frequency, and b, k_1 are parameters. K represents the ratio between the length of D_i measured by l_i (sum of tf_{ij}) and the collection mean denoted by advl.

To index a request, the following formula was used:

$$w_{qj} = \frac{tf_{qj}}{k_3 + tf_{qj}} \cdot \ln[(N - df_j) / df_j]$$

where tf_{qj} indicates the search term frequency, df_j the collection-wide term frequency, N the number of documents in the collection, and k_3 is a parameter.

To adjust the underlying parameters of the OKAPI search model, we used advl = 900, b = 0.7625, k_1 = 1.5, and k_3 = 1000. These parameter values were set according to the best performance achieved on the WT2g (TREC-8). A slightly different parameter setting was suggested by Walker *et al.* [98] whereby advl = 900, b = 0.75, k_1 = 1.2, and k_3 = 1000. When using our parameter values, the corresponding label will be "OKAPI" while the second setting will be identified by adding an "R".

Two different query formulations were considered: (1) using only the Title section (T), or (2) all three logical sections (Title, Descriptive and Narrative, noted T-D-N). The data in Table 1 shows that retrieval effectiveness is significantly enhanced when topics include more search terms.

In order to build a single collection, we selected the first 500 retrieved items of 13 search strategies (corresponding to OKAPI and different vector-space approaches) and we added all relevant documents not retrieved by our various search models.

Table 1 provides a summary of the results of our various experiments. In this case, we reported the non-interpolated average precision at eleven recall values, based on 1,000 retrieved items per request. From this data we can see that the parameter setting used by Walker's *et al.* results in better performance (e.g., in the WEB9.1 sub-collection, the average precision increases from 19.47 to 20.30 (+4.3%)).

It is recognized that pseudo-relevance feedback (blind expansion) is a useful technique for enhancing retrieval effectiveness. In this study, we evaluated the OKAPI search model with and without query expansion in order to verify whether or not this technique might improve retrieval performance when faced with different query formulations.

In this study, we adopted Rocchio's approach [Buckley 96] where the parameter settings were chosen according to experiments done with the WT2g from the TREC data (TREC-8).

For a short request the values α=0.75, β=0.25 were assigned and the system was allowed to add to the original query those 50 search terms extracted from the 12-best ranked documents. For long queries, the parameters were set as follows: α=0.7, β=0.3 and the search engine was allowed to add to the original query those 40 search terms extracted from the 15 best-ranked documents. The resulting retrieval effectiveness is depicted in Table 1 under the label "XQ".

After examining sub-collections WEB9.1 and WEB9.3, there was some improvement in results, as depicted in Table 1. For example, based on our parameter setting and examining the WEB9.1 sub-collection, the average precision increased from 19.47 (label "OKAPI") to 21.44 (label "OKAPIXQ") (+10.1%). However, for the other two sub-collections, the average precision decreased (e.g. in WEB9.4, the average precision decreases from 19.26 to 18.24 (-5.3%)).

1.2. Merging procedure

Recent works have suggested solutions in which answer lists were merged in order to produce a unique ranked list of retrieved records. As a first approach, we might assume that each sub-collection contains approximately the same number of pertinent items and that the distribution of the relevant documents is the same across the answer lists. Based only on a ranking of the retrieved records, we might interleave the results in a round-robin fashion. According to previous studies [Voorhees 95], [Callan 95], the retrieval effectiveness of such interleaving schemes is around 40% below the performance achieved by a single retrieval scheme technique, with a single huge collection representing the entire set of documents. The third column of Table 2 confirms this finding but to a lesser extent (around -26.1% when dealing with short queries or -17.0% when examining Title, Descriptive and Narrative fields in the topics).

	Average Precision				
Query Title only Model	WEB9.1 46 queries 749 rel	WEB9.2 44 queries 600 rel	WEB9.3 43 queries 608 rel	WEB9.4 46 queries 660 rel	WEB9 50 queries 2,617 rel.
Okapi OkapiR	19.47 20.30	20.85 21.32	16.09 16.52	19.26 19.51	19.55 19.86
OkapiXQ OkapiNRXQ	21.44 21.70	20.89 20.67	17.73 18.98	18.24 18.33	19.43 19.31
Query T-D-N					
Okapi OkapiNR	32.61 33.25	30.26 30.19	28.09 29.01	28.44 28.49	27.25 27.52
OkapiXQ OkapiNRXQ	34.41	28.25	31.18	26.69	28.10 28.30

Table 1: Average precision of isolated sub-collections and the whole test collection

In order to account for the score achieved by the retrieved document, we might formulate the hypothesis that each sub-collection is managed by the same search strategy and that the similarity values are therefore directly comparable [Kwok 95]. Such a strategy, called raw-score merging, produces a final list, sorted by the retrieval status value computed by each separate search engine.

However, as demonstrated by Dumais [94], collection-dependent statistics in document or query weights may vary widely among sub-collections; and therefore, this phenomenon may invalidate the raw-score merging hypothesis.

The fourth column of Table 2 indicates the retrieval effectiveness of such merging approaches, depicting a relatively interesting performances in our case (degradation of around -5.3% for long requests or -14.9% for short queries). Thus, the raw-score merging seems to be a simple and valid approach when a huge collection is distributed across a local-area network and operating within the same retrieval scheme.

As a third merging strategy, we may normalize each sub-collection's similarity value (SIM(D_i, Q)) by dividing it by the maximum value in each result list. The fifth column in Table 2 shows its average precision, depicting surprisingly poor retrieval effectiveness (average reduction of -19.6% for short queries and -16.2% for long requests).

As a fourth merging strategy, Callan *et al.* [95] suggest using the CORI approach, which will first compute a score s_i for each sub-collection as follows:

$$\text{score}(t_j \mid db_i) = \text{defB} + (1-\text{defB}) \cdot \frac{df_i}{df_i + K} \cdot \frac{\log\left[\frac{db + 0.5}{cf_j}\right]}{\log(db+1)} \quad \text{with } K = k \cdot \left[(1-b) + b \cdot \frac{ldb_i}{avldb}\right]$$

where t_j indicates a search keyword, db_i the ith collection, df_i the number of documents in the ith collection containing term t_j, cf_j the number of collections containing term t_j, db the total (number of collections equals to four in our case), ldb_i the number of indexing terms included in the ith corpus, avldb the mean value of ldb_i, where defB, b and k are three parameters. Xu & Callan [98] suggest assigning values to these constants (defB=0.4, k=200, and b=0.75, values used in this study). The previous equation is defined for one search term, and the score for a given collection is simply the sum over all keywords included in the current request.

The sub-collection score (noted s_i) is the first component used to merge the retrieved items. To obtain the score of a given retrieved item of the ith collection, the similarity between the request and the document is multiplied by a coefficient w_i computed as follows:

$$w_i = 1 + dbs \cdot [(s_i - S_m) / S_m]$$

	Average Precision (% change)				
Query Title Model	50 queries 2617 rel one coll	merge 50 queries 2617 rel round	merge 50 queries 2617 rel raw-score	merge 50 queries 2617 rel norm. score	merge 50 queries 2617 rel CORI
Okapi	19.55	13.88 (-29.0%)	17.59 (-10.0%)	15.94 (-18.5%)	15.83 (-19.0%)
OkapiR	19.86	14.44 (-27.3%)	17.81 (-10.3%)	16.37 (-17.6%)	15.99 (-19.5%)
OkapiXQ	19.43	14.54 (-25.2%)	15.96 (-17.9%)	15.07 (-22.4%)	15.31 (-21.2%)
OkapiNRXQ	19.31	14.89 (-22.9%)	15.87 (-17.8%)	15.44 (-20.0%)	15.28 (-20.9%)
		-26.1%	-14.9%	-19.6%	-20.2%
Query T-D-N					
Okapi	27.25	22.82 (-16.3%)	26.56 (-2.5%)	23.39 (-14.2%)	26.81 (-1.6%)
OkapiNR	27.52	23.19 (-15.7%)	26.75 (-2.8%)	23.94 (-13.0%)	26.87 (-2.4%)
OkapiXQ	28.10	23.09 (-17.8%)	25.99 (-7.5%)	23.28 (-17.2%)	25.94 (-7.7%)
OkaNRXQ	28.30	23.22 (-18.0%)	25.93 (-8.4%)	22.57 (-20.2%)	25.84 (-8.7%)
		-17.0%	-5.3%	-16.2%	-5.1%

Table 2: Average precision of different merging procedures

where dbs indicates the number of the selected collections (all in our case), s_i the score achieved by the ith collection and S_m the mean score over all collections. According to our evaluation, the mean average precision results in a degradation of around 20.2% for short queries and 5.1% for long requests. It is interesting to note that both the raw-score merging and the CORI approach result in good performances when dealing with long requests yet a decrease in performance when using short requests.

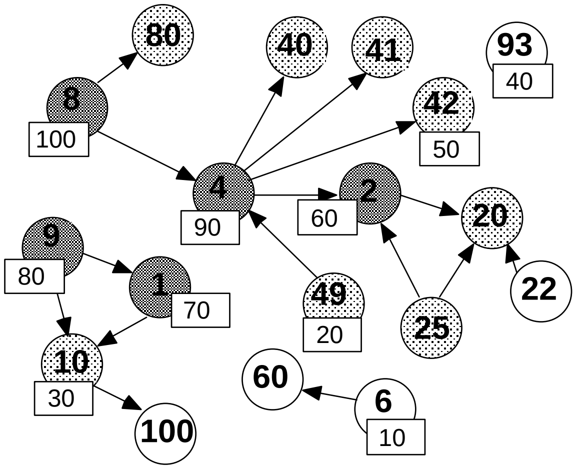

Figure 1: Starting situation for our link-based approaches

2. Link-based retrieval

Various retrieval strategies have been suggested in order to take account of hyperlinks, based on the assumption that links between documents indicate useful semantic relationships between related web pages [Kleinberg 98], [Brin 98], [Chakrabarti 99]. For example, Chakrabarti et al. [99] stated:

> "Citations signify deliberate judgment by the page author. Although some fractions of citations are noisy, most citations are to semantically related material. Thus the relevance of a page is a reasonable indicator of the relevance of its neighbors, although the reliability of this rule falls off rapidly with increasing radius on average." [Chakrabarti 99, p. 550-551]

With small variations, similar hypotheses are also cited by other authors [Kleinberg 98]. In order to verify the retrieval effectiveness of such assumptions, we have evaluated four different search strategies, namely our spreading activation approach in Section 2.1, our PAS search model in Section 2.2, Kleinberg's algorithm in Section 2.3 and the PageRank approach in Section 2.4. These search strategies will be described briefly using a small example.

As a first step, the search strategy computes the similarity between the given query and the documents, with values noted as $SIM(D_i, Q)$. These values are depicted inside a rectangle in Figure 1. In this case, we can see that the first five retrieved documents are D_8, D_4, D_9, D_1 and D_2. At this point various retrieval schemes will take note of the hyperlinks so that the retrieval effectiveness might hopefully be improved.

2.1. Spreading activation

In a first link-based strategy, we chose the spreading activation (SA) approach [Crestani 00]. In that method, the degree of match between a web page D_i and a query, as initially computed by the IR system (denoted $\mathrm{SIM}(D_i, Q)$), is propagated to the linked documents through a certain number of cycles using a propagation factor. We used a simplified version with only one cycle and a fixed propagation factor λ for all links. In that case, the final retrieval status value of a document D_i linked to m documents is computed according to the following equation:

$$\mathrm{RSV}(D_i) = \mathrm{SIM}(D_i, Q) + \lambda \cdot \sum_{j=1}^{k} \mathrm{SIM}(D_j, Q)$$

Using all the incoming and outgoing links, and for different values of the parameter λ, in most cases did not result in retrieval improvement within the WT2g corpus [Savoy 01]. In order to be more selective in the spreading phase, we only consider in this study the best outgoing and the best incoming link for each of the k best-ranked documents (the constant k was fixed to 15 in this paper and the parameter λ to 0.05). But, what do we mean by the best link?

Instead of considering the m web pages linked to a given document, we only consider the incoming link coming from the best ranked document. For the outgoing links, we adopt a similar point of view, taking into account only the link starting from the given document to the best rank web page.

For example, based on Figure 1, we do not follow all outgoing from D_4 but we activate only the hyperlink to D_2 (the rank of this document is better than for the others). Similarly, the best incoming link is the link between D_8 to D_4. Fixing the parameter λ to 0.1 and k to 5, the final retrieval status value of D_4, noted $\mathrm{RSV}(D_4)$, will be :

$$\mathrm{RSV}(D_4) = \mathrm{SIM}(D_4, Q) + \lambda \cdot \mathrm{SIM}(D_2, Q) + \lambda \cdot \mathrm{SIM}(D_8, Q) = 90 + 0.1 \cdot 60 + 0.1 \cdot 100 = 106$$

The similarity value of non-retrieved documents (e.g., D_{20} in our example) will be set according the similarity achieved by the last retrieved item (10 in our example, 1,000 in the evaluation). The evaluation of other web pages included in our example is given in Table 3.

2.2. Probabilistic argumentation system

In a second set of experiments, we used our probabilistic argumentation systems (PAS) [Picard 98], in which we used a simplified version of our approach, whereby the final retrieval status value of a document (or its degree of support, denoted $\mathrm{DSP}(D_i)$) might only be affected by its direct neighbors. In this case we do not need to keep track of inferences, and can derive a simple formula which might be considered to be a more refined spreading activation method. Instead of propagating a document's similarity value, we propagated its probability of being relevant.

In this approach, we must therefore first compute the relevance probability of a document D_i. To achieve this, we suggest using logistic regression methodology [Bookstein 92] and the natural logarithm of its rank as an explanatory variable. Such a computation will be noted $p(D_i \mid \mathrm{rank})$ [Le Calvé 00] and in accordance with the following formula:

$$P\left[D_i \mid \mathrm{rank}\right] = \frac{e^{\alpha + \beta \cdot \ln(\mathrm{rank})}}{1 + e^{\alpha + \beta \cdot \ln(\mathrm{rank})}}$$

in which α et β are parameters set to 0.7 and -0.8 respectively.

In a second step, this probability of relevance will be modified according to the neighbors of a given document. The individual contribution of a linked document D_j to D_i is given by $[p(D_j \mid \mathrm{rank}) \cdot p(\mathrm{link})]$, instead of the $[\mathrm{SIM}(D_j, Q) \cdot \lambda]$ used with the spreading activation technique.

Just as with the spreading activation experiments, using all incoming or outgoing links did not demonstrate any improvement, except in some cases when using the WT2g test collection [Savoy 01]. We then decided to include only the most important sources of evidence, the same way as for spreading activation. For example, we considered the initial rank of document D_i, the best incoming document D_{in} and the best outgoing document D_{out}.

This link-based retrieval approach will thus multiply the probability of linked document relevance by the probability of the link, denoted $p(link_{in})$ for incoming hyperlinks or $p(link_{out})$ for outgoing links. The final degree of support corresponding to document D_i is computed as follow:

$$DSP(D_i) = 1 - (1 - p(D_i \mid rank)) \cdot [1 - p(D_{in} \mid rank) \cdot p(link_{in})] \cdot [1 - p(D_{out} \mid rank) \cdot p(link_{out})]$$

Fixing $p(link_{in})=0.1$ and $p(link_{out})=0.2$, and based on the situation depicted in Figure 1, computation of degree of support for Document 1 as follows:

$$DSP(D_1) = 1 - (1 - p(D_1 \mid rank)) \cdot [1 - p(D_9 \mid rank) \cdot p(link_{in})] \cdot [1 - p(D_{10} \mid rank) \cdot p(link_{out})] =$$
$$1 - (1 - 0.3991) \cdot [1 - 0.4554 \cdot 0.1] \cdot [1 - 0.2762 \cdot 0.2] =$$
$$1 - (0.6009) \cdot [0.95446] \cdot [0.94476] =$$
$$1 - 0.5418 = 0.4582$$

Table 4 lists other results pertaining to the best ten retrieved items of Figure 1. For the results based on the web test collection, link probabilities are fixed as $p(link_{in}) = 0.062$, $p(link_{out}) = 0.051$, probability estimates are defined in [Savoy 01]. Finally, documents not belonging to the top 1000 have a similarity value equal to the similarity value obtained for the 1000th retrieved item.

2.3. Kleinberg’s algorithm

As a third link-based approach, we have applied Kleinberg's algorithm [Kleinberg 98]. In this scheme, a web page pointing to many other information sources must be viewed as a "good" hub while a document with many web pages pointing to it is a "good" authority. Likewise, a document that points to many "good" authorities is an even better hub while a web page pointed to by many "good" hubs is an even better authority.

For document D_i after c+1 iterations, the updated formulas for the hub and authority scores $H^{c+1}(D_i)$ and $A^{c+1}(D_i)$ are:

$$A^{c+1}(D_i) = \sum_{D_j = parent(D_i)} H^c(D_j)$$

$$H^{c+1}(D_i) = \sum_{D_j = child(D_i)} A^c(D_j)$$

Rank	D_i	$SIM(D_i, Q)$	D_i	$RSV(D_i)$
1	8	100	8	109
2	4	90	4	106
3	9	80	9	87
4	1	70	1	78
5	2	60	2	69
6	42	50	42	50
7	93	40	93	40
8	10	30	10	37
9	49	20	49	20
10	6	10	20	16
			6	10

Table 3: Retrieval status value obtained by the spreading activation

Rank	D_i	$SIM(D_i,Q)$	$p(D_i \mid rank)$	D_i	$DSP(D_i)$
1	8	100	0.6682	8	0.7038
2	4	90	0.5363	4	0.5982
3	9	80	0.4554	9	0.4989
4	1	70	0.3991	1	0.4582
5	2	60	0.3572	2	0.4211
6	42	50	0.3244	42	0.3244
7	93	40	0.2980	10	0.3051
8	10	30	0.2762	93	0.2980
9	49	20	0.2577	20	0.2690
10	6	10	0.2419	49	0.2577
11			0.2419	6	0.2419

Table 4: Computation of the degree of support of our PAS search model

which is computed for the k best-ranked documents (defined as the root set) retrieved by a classical search model, together with their children and parents (which defined the base set). The hub and authority scores were updated for five iterations (while the ranking did not change after this point), and a normalization procedure (dividing each score by the sum of all square values) was applied after each step.

As an example, we will refer to the initial situation shown in Figure 1. We fixed k = 5 and our root set was $\{D_8, D_4, D_9, D_1, D_2\}$, leading to the following base set $\{D_8, D_4, D_9, D_1, D_2, D_{80}, D_{40}, D_{41}, D_{42}, D_{20}, D_{25}, D_{49}, D_{10}\}$. Initially, the hub and authority score for each document is set to 1. In the first iteration, the hub score for D_4 corresponds to the sum of the authority values for its children (D_{40}, D_{41}, D_{42}, D_2) while its authority score is the sum of the hub scores of its parents (D_8, D_{49}). For other items belonging to the basic set, computation of these scores is depicted in Table 5.

After five iterations and using the normalization procedure, we obtained the ranked list depicted in Table 6. Taking the five best-ranked documents obtained by the traditional search engine into account and the top five documents retrieved according to the authority scores, we note that the intersection included only one item, namely D_2.

2.4. PageRank algorithm

Brin & Page [98] suggest a link-based search model called PageRank that first evaluated the importance of each web page based on its citation pattern. As for the spreading activation approach, the PageRank algorithm reranked the retrieved pages of a traditional search schemes according to the PageRank values assigned to the retrieved items.

In this approach, a web page will have a higher score if many web pages point to it. This value increases if there are highly scoring documents pointing to it. The PageRank value of a given web page D_i, value noted as $PR(D_i)$, having D_1, D_2, … D_m pages pointing to D_i, is computed according to the following formula:

$$PR(D_i) = (1 - d) + d \cdot [(PR(D_1) / C(D_1)) + \ldots + (PR(D_m) / C(D_m))]$$

where d is a parameter (set to 0.85 as suggested by [Brin 98] and $C(D_j)$ are the number of outgoing links for web page D_j.

The computation of the PageRank value can be done using an iterative procedure (five iterations were computed in our case). After each iteration, each PageRank value was divided by the sum of all PageRank values. Finally, as initial values, $PR(D_i)$ were set to 1/N where N indicates the number of documents in the collection.

Based on our example, the result list achieved by using the PageRank algorithm is depicted in Table 8.

2.5. Evaluation

The retrieval effectiveness of the four link-based search model are shown in Table 9. From this table, it seems clear that links do not seem an appropriate source of information about document contents, and they seem to provide less information than do the bibliographic references or co-citation schemes used in our previous studies [Savoy 96]. The poor results depicted by Kleinberg's approach or PageRank algorithm raise some questions: Is our implementation without bugs? Can other teams confirm these findings? Have the underlying parameters the good values?

Our official runs were produced using the raw-score merging, where three were based only on the Title portion of the requests (NEtm, NENRtm, NENRtmLpas) and three were based on all logical sections of the queries (NEnm, NEnmLpas, NEnmLsa). Three of them were link-based retrievals (ending by Lpas or Lsa indicating the PAS or spreading activation approach).

For the two types of requests, our official runs included a spelling check performed automatically by the Smalltalk-80 system. This feature has a positive effect for short queries (e.g., 15.96 vs. 17.54 (+9.9%)) but not for long ones (25.99 vs. 24.99 (-3.8%)).

Conclusion

The various experiments carried out within the web track demonstrated that:

- Hyperlinks do not result in any significant improvement (at least as implemented in this study). Link information seems to be marginally useful when the retrieval system produces relatively high retrieval effectiveness;

- Pseudo-relevant feedback techniques (blind query expansion) usually result in significant improvement but setting the underlying parameters based on another test collection may lead to a decrease in retrieval effectiveness;
- Longer topic descriptions (Title, Description and Narrative) improve the retrieval performance significantly over short queries built only from the Title section;
- It seems that the raw-score approach might be a valid first attempt for merging result lists provided by the same retrieval model.

Acknowledgments

The authors would like to thank C. Buckley from SabIR for allowing us the opportunity to use the SMART system. This research was supported by the SNSF (Swiss National Science Foundation) under grant 21-58'813.99.

D_i	$H^0(D_i)$	Author comput	$A^1(D_i)$	$A^0(D_i)$	Hub comput	$H^1(D_i)$
8	1		0	1	1 + 1	2
4	1	1 + 1	2	1	1 + 1 + 1 + 1	4
9	1		0	1	1 + 1	2
1	1	1	1	1	1	1
2	1	1 + 1	2	1	1	1
40	1	1	1	1		0
41	1	1	1	1		0
42	1	1	1	1		0
49	1		0	1	1	1
80	1	1	1	1		0
20	1	1 + 1	2	1		0
25	1		0	1	1 + 1	2
10	1	1 + 1	2	1		0

Table 5: Computation of the hub and authority scores for our example

Rank	D_i	$SIM(D_i, Q)$	D_i	$A^5(D_i)$	D_i	$H^5(D_i)$
1	8	100	2	0.1239	4	0.1501
2	4	90	42	0.0762	25	0.0723
3	9	80	41	0.0762	9	0.0241
4	1	70	40	0.0762	8	0.0241
5	2	60	20	0.0667	2	0.0222
6	42	50	4	0.0413	49	0.0148
7	93	40	10	0.0413	1	0.0148
8	10	30	80	0.0254	80	0
9	49	20	1	0.0254	42	0
10	6	10	9	0	41	0

Table 6: Computation of the hub and authority scores after five iterations

Rank	D_i	$SIM(D_i, Q)$	Rank	D_i	$PR(D_i)$
1	8	100	1	10	0.2710
2	4	90	2	4	0.2548
3	9	80	3	2	0.2146
4	1	70	4	1	0.1849
5	2	60	5	42	0.1797
6	42	50	6	93	0.15
7	93	40	7	49	0.15
8	10	30	8	9	0.15
9	49	20	9	8	0.15
10	6	10	10	6	0.15

Table 8: Ranked list obtained in our example by the traditional and the PageRank approach

	without normalizat.		with normalization	
D_i	$PR^1(D_i)$	$PR^5(D_i)$	$PR^1(D_i)$	$PR^5(D_i)$
1	0.1736	0.2138	0.1925	0.1849
2	0.1854	0.2863	0.2138	0.2146
4	0.2208	0.3413	0.2775	0.2548
6	0.15	0.15	0.15	0.15
8	0.15	0.15	0.15	0.15
9	0.15	0.15	0.15	0.15
10	0.2208	0.3954	0.2775	0.2710
20	0.2681	0.5846	0.3625	0.3547
22	0.15	0.15	0.15	0.15
25	0.15	0.15	0.15	0.15
40	0.1618	0.2225	0.1713	0.1797
41	0.1618	0.2225	0.1713	0.1797
42	0.1618	0.2225	0.1713	0.1797
49	0.15	0.15	0.15	0.15
60	0.1972	0.2775	0.235	0.2198
80	0.1736	0.2138	0.1925	0.1849
93	0.15	0.15	0.15	0.15
100	0.1972	0.4861	0.235	0.2762

Table 7: Computation of the PageRank values with and without normalization

	Average Precision				
Query Title Model	merge raw-score	SA	PAS	Kleinberg	PageRank
Okapi	17.59	14.64	17.57	0.18	2.82
OkapiR	17.81	14.59	17.76	0.19	2.79
OkapiXQ	15.96	13.43	15.91	0.17	2.37
OkapiNRXQ	15.87	13.48	15.85	0.17	2.69
Query T-D-N					
Okapi	26.56	23.80	26.43	0.36	3.09
OkapiNR	26.75	24.10	26.65	0.25	3.14
OkapiXQ	25.99	22.27	25.87	0.31	3.11
OkaNRXQ	25.93	22.57	25.82	0.25	3.13

Table 9: Average precision of different link-based approaches

Official run name	Corresponding run name	Average Pre.	# ≥ Median	# Best
NEtm	OKAPIXQ	17.54	41	3
NENRtm	OKAPIRXQ	17.43	41	2
NENRtmLpas	OKAPIRXQ + PAS	17.36	40	1
NEnm	OKAPIXQ	24.99	45	4
NEnmLpas	OKAPIRXQ + PAS	24.88	43	0
NEnmLsa	OKAPIRXQ + SA	21.85	41	0

Table 10: Summary of our official runs for the web track

References

[Bookstein 92] A. Bookstein, E. O'Neil, M. Dillon, D. Stephens: Applications of loglinear models for informetric phenomena. Information Processing & Management, 28(1), 1992, 75-88.

[Brin 98] S. Brin, L. Page: The anatomy of a large-scale hypertextual web search engine. WWW8, 1998, 107-117.

[Buckley 96] C. Buckley, A. Singhal, M. Mitra, G. Salton: New retrieval approaches using SMART. TREC-4, 1996, 25-48.

[Callan 95] J. P. Callan, Z. Lu, W. B. Croft: Searching distributed collections with inference networks. ACM-SIGIR'95, 21-28.

[Chakrabarti 99] S. Chakrabarti, M. Van den Berg, B. Dom: Focused Crawling: A new approach to topic-specific web resource discovery. WWW8, 1999, 545-562.

[Crestani 00] F. Crestani, P. L. Lee: Searching the web by constrained spreading activation. Information Processing & Management, 36(4), 2000, 585-605.

[Dumais 94] S. T. Dumais: Latent semantic indexing (LSI) and TREC-2. TREC'2, 1994, 105-115.

[Hawking 99] D. Hawking, N. Craswell, P. Thistlewaite, D. Harman: Results and challenges in web search evaluation. WWW8, 1999, 243-252.

[Kleinberg 98] J. Kleinberg: Authoritative sources in a hyperlinked environment. Proceedings of 9th ACM-SIAM Symposium on Discrete Algorithms, 1998, 668-677.

[Kwok 95] K. L. Kwok, L. Grunfeld, D. D. Lewis: TREC-3 ad-hoc, routing retrieval and thresholding experiments using PIRCS. TREC-3, 1995, 247-255.

[Lawrence 99] S. Lawrence, C. Lee Giles: Accessibility of information on the web. Nature 400 (6740), 1999, 107-110.

[Le Calvé 00] A. Le Calvé, J. Savoy: Database merging strategy based on logistic regression. Information Processing & Management, 36(3), 2000, 341-359.

[Picard 98] J. Picard: Modeling and combining evidence provided by document relationships using PAS systems. ACM-SIGIR'98, 182-189.

[Robertson 95] S. E. Robertson, S. Walker, M. M. Hancock-Beaulieu: Large test collection experiments on an operational, interactive system: OKAPI at TREC. Information Processing & Management, 31(3), 1995, 345-360.

[Savoy 96] J. Savoy: Citation schemes in hypertext information retrieval. In Information retrieval and hypertext, M. Agosti, A. Smeaton (Eds), Kluwer, 1996, 99-120.

[Savoy 01] J. Savoy, J. Picard: Retrieval effectiveness on the web. Information Processing & Management, 2001, to appear.

[Selberg 99] E. W. Selberg: Towards comprehensive web search. Ph.D. Thesis, University of Washington, 1999.

[Voorhees 95] E. M. Voorhees, N. K. Gupta, B. Johnson-Laird: Learning collection fusion strategies. ACM-SIGIR'95, 172-179.

[Walker 98] S. Walker, S. E. Robertson, M. Boughamen, G. J. F. Jones, K. Sparck Jones: Okapi at TREC-6: Automatic ad hoc, VLC, routing, filtering and QSDR. TREC-6, 1998, 125-136.

[Xu 98] J. Xu, J. P. Callan: Effective retrieval with distributed collections. ACM-SIGIR'98, 112-120.

Appendix 1: Statistics describing our various sub-collections

Collection	WEB9.1	WEB9.2	WEB9.3	WEB9.4	WEB9
Size (in MB)	2,799 MB	2,754 MB	2,790 MB	2,690 MB	11,032 MB
# of documents	414,914	423,965	442,711	410,506	1,692,096
# of relevant doc.	749	600	608	660	2617
# of queries	46	44	43	46	50
mean	16.2826	13.6364	14.1395	14.3478	52.34
standard error	25.0986	21.826	21.3637	22.1873	84.1405
maximum	157	105	133	124	519
for # query	(#q:495)	(#q:495)	(#q:495)	(#q:495)	(#q:495)
minimum	1	1	1	1	1
for # query	(#q:461)	(#q:461)	(#q:464)	(#q:456)	(#q:473)
size invert. file doc.nnn	674.2 MB	642.1 MB	655.6 MB	635.4 MB	
# indexing terms	3,428,795	2,827,067	3,607,359	3,537,393	
max df	189,386	207,892	228,922	191,208	
Indexing time (real)	1:05:17	1:00:28	1:00:18	1:00:49	

Table A.1: Some statistics about the four sub-collections of the Web corpora