Emulating Query-Biased Summaries Using Document Titles*

Hideo Joho
Department of Computing Science, University of Glasgow, UK.
hideo@dcs.gla.ac.uk

David Hannah
Department of Computing Science, University of Glasgow, UK.
hannahd@dcs.gla.ac.uk

Joemon M. Jose
Department of Computing Science, University of Glasgow, UK.
jj@dcs.gla.ac.uk

ABSTRACT

Generating query-biased summaries can take up a large part of the response time of interactive information retrieval (IIR) systems. This paper proposes to use document titles as an alternative to queries in the generation of summaries. The use of document titles allows us to pre-generate summaries statically, and thus, improve the response speed of IIR systems. Our experiments suggest that title-biased summaries are a promising alternative to query-biased summaries.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Search process; H.5 [**Information Interfaces and Presentation**]: User Interfaces

General Terms

Design, Experimentation, Human Factors, Performance

Keywords

Title-biased summaries, Interactive Information Retrieval

1. INTRODUCTION

Early interactive information retrieval (IIR) systems presented the first few sentences of documents in search results. This changed when the advantages of query-biased summaries were shown to be significant [5]. The query-biased summary, or *QBS*, became a standard feature in the result presentation of search engines.

The methods of generating a QBS vary (e.g., [5, 6]). An early approach [5] was based on a mixture of factors such as the frequency of occurrence of query terms, location of sentences in a document, and/or weighting of sentence words based on HTML tags. A disadvantage of QBS is the generation-cost at query-time. The summary needs to be generated for every single document presented to a searcher in response to a potentially diverse range of queries. One way to reduce such a cost is to cache the summary for frequently submitted queries [6]. Search engines generate the QBS in a satisfactory response time with the machine power of thousands of PCs.

This paper proposes an alternative way of using existing QBS techniques to generate a summary without the cost at query-time. More specifically, we propose to use document titles as an alternative to queries. Since we use document titles, the summary can be pre-generated statically. When the summaries are pre-generated, presenting them to the users becomes a simple lookup in a database. We call it a *title-biased summary*, or *TBS* in this paper. For a given document, QBS and TBS will be identical when a query is the title of the document. Therefore, the relative effectiveness of TBS to QBS can be estimated by investigating the effect of query terms in search. To justify the title-biased approach of summary generation, we made three research hypotheses as follows.

- H_1: Top ranking documents tend to have a query term in the title;
- H_2: Searchers prefer to visit a document when a query term appears in the title;
- H_3: There is no significant difference between QBS and TBS in supporting search tasks.

H_1 is concerned with the behaviour of existing ranking techniques. If H_1 is supported, a document is more likely to be presented to users when query terms appear in the title. We call such a document a *query-in-title* (QIT) document. H_2 is concerned with the behaviour of searchers. If H_2 is supported, QIT documents are more likely to be accessed to complete search tasks. We found that this hypothesis has already been supported by the work of Clarke, et al. [2]. Their analysis shows that the presence of query terms in document titles of search results affects people's preference of click-through actions. Therefore, we decided to alter our experiment to omit H_2. Finally, H_3 is a summative evaluation of TBS. It should be noted that we do not expect TBS to outperform QBS. An equivalent performance of TBS will be sufficient due to the reduction of resources required at query-time of retrieval.

2. EXPERIMENTS

This section presents the results of our experiments carried out to address the hypotheses H_1 and H_3 as discussed above.

2.1 QIT documents in search results

H_1 addressed the behaviour of ranking techniques with regards to the QIT documents. Figure 1 shows the proportion of QIT documents in the top 100 documents retrieved by search engines (Google and Yahoo) and retrieval models (TFxIDF, BM25, and PL2). We used the title of the 100 topics developed in the HARD track 2005 and 2006 [1] as queries in this analysis. Note that the Aquaint collections were used for the retrieval models but not for the search engines.

As can be seen, a high proportion of the documents retrieved by search engines contain a query term in the title of search results.

*This work was supported by the MIAUCE project (Ref: IST-033715). Any opinions, findings, and conclusions described here are the authors and do not necessarily reflect those of the sponsor.

SIGIR'08, July 20–24, 2008, Singapore.
ACM 978-1-60558-164-4/08/07.

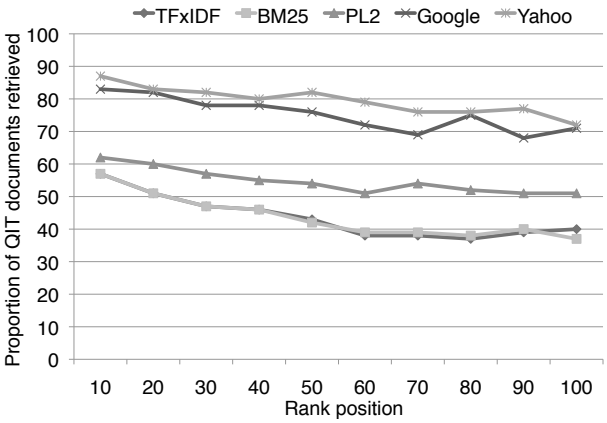

Figure 1: Proportion of QIT documents in the search results.

The TFxIDF and BM25 models had a smaller proportion of QIT documents but it still went over 50% at the top 20 documents. PL2 [3], one of the advanced models, had a higher proportion of QIT documents than TFxIDF and BM25. We speculate that such models which are designed to incorporate field weighing (e.g., [4]) will have an a higher proportion of QIT documents, closer to the level of search engines. In summary, H_1 was well supported for the search engines, and partially supported for the retrieval models.

2.2 Task performance

H_3 addressed the effectiveness of two summary generation methods in supporting search tasks. To address H_3 we recruited 9 people (All were research assistants or research students in Computer Science) and asked them to evaluate the quality of search results presented with the title and summary.

Each subject was presented the description of a HARD Track topic and a randomised list of 10 documents relevant to the topics. They were asked to browse the document list first, then to select a single document that looked most promising to find relevant information, without accessing the fulltext. By changing the number of relevant documents shown in the list, we simulated the difficulty of tasks such as $P@10 = .2$, .4, and .6. We used the first N (2, 4, or 6) relevant documents retrieved by PL2 [3] for each topic. The non-relevant documents were selected in the same fashion. Subjects were informed that the document list was randomised so that the first document did not imply the most relevant.

Each subject was asked to perform the selection task for 12 topics randomly selected from 47 topics of the HARD Track 2005 [1] (Three topics had less than six relevant documents, thus, excluded). QBS was generated for six of 12 topics, and TBS was generated for the rest of the topics. We used a version of the summary generation software developed in [7]. The title of the topic descriptions was used for QBS and title of each document was used for TBS. In the search results, the terms appeared in the title of the topic descriptions were highlighted regardless of the summary methods. Finally, the order of summary methods, topics, and baseline levels was randomised for each subject.

The results of the experiment are shown in Table 1. Every subjects performed the task in each block twice, and thus, the sample size was 18. We measured the probability of relevant documents selected by subjects as an accuracy, and time taken to make the selection. The standard deviation of the average is shown in the brackets. As can be seen, the accuracy was found to be higher than the baseline, suggesting that subjects were able to identify relevant documents from the search results. The average time taken to complete the task suggests that they carefully examined the results to make the selection but did not vary much across the baseline levels.

	Accuracy		Time (sec)	
Baseline	QBS	TBS	QBS	TBS
$P@10 = .2$	.44 (.51)	.72 (.46)	115 (54)	107 (53)
$P@10 = .4$	.56 (.51)	**.89** (.32)	111 (71)	115 (70)
$P@10 = .6$	.78 (.43)	.83 (.38)	106 (68)	118 (73)
Total	.59 (.50)	**.81** (.39)	111 (64)	114 (65)

Table 1: Accuracy of click-through documents and time taken for the selection (N=18)

We ran a χ^2 test to measure the correlation between the accuracy and summarisation methods. The accuracy of TBS was found to be significantly better than QBS in $P@10 = .4$ ($p = .03$) and in the total ($p = .01$). This was unexpected. While we must take this result with care due to the sample size, a possible explanation was that the document title could be a good source to generate a summary when it was carefully written by the author (i.e., might not apply to some web pages). Clearly, more studies with different search tasks are needed to support H_3. However, our results suggest that the title-biased summary is a promising alternative to query-biased summary in an ad-hoc search task.

3. CONCLUSION

This paper proposed using document titles as an alternative to queries in the generation of document summaries. This allowed us to pre-generate document summaries statically, thus, the response time of a search interface can be reduced to a simple lookup on a database. Our experimental results suggest that title-biased summaries are a promising alternative to query-biased summaries, due to the behaviour of existing retrieval systems as well as searchers' information seeking behaviour. We plan to extend our experiments to other test collections, retrieval models, and search tasks.

4. REFERENCES

[1] J. Allan. Hard track overview in trec 2005 high accuracy retrieval from documents. In *NIST Special Publication: SP 500-266 The Fourteenth TREC*, 2005.

[2] C. L. A. Clarke, E. Agichtein, S. Dumais, and R. W. White. The influence of caption features on clickthrough patterns in web search. In *Proceedings of the 30th ACM SIGIR conference*, pages 135–142, 2007.

[3] I. Ounis, C. Lioma, C. Macdonald, and V. Plachouras. Research directions in terrier. *Novatica/UPGRADE Special Issue on Web Information Access*, 2007.

[4] S. Robertson, H. Zaragoza, and M. Taylor. Simple bm25 extension to multiple weighted fields. In *Proceedings of the thirteenth ACM CIKM conference*, pages 42–49, 2004.

[5] A. Tombros and M. Sanderson. Advantages of query-biased summaries in information retrieval. In *Proceedings of the 21st ACM SIGIR Conference*, pages 2–10, 1998.

[6] A. Turpin, Y. Tsegay, D. Hawking, and H. E. Williams. Fast generation of result snippets in web search. In *Proceedings of the 30th ACM SIGIR conference*, pages 127–134, 2007.

[7] R. White, J. M. Jose, and I. Ruthven. A task oriented-study on the influencing effects of query-biased summarisation in web searching. *IP&M*, 9(5):707–733, 2003.