Athens Institute for Education and Research
ATINER

ATINER's Conference Paper Series
COM2014-0939

Using Text Highlight and Page Reviews to Improve Search Engine Result Page

Vimala Balakrishnan
Senior Lecturer, University of Malaya
Malaysia

Xinyue Zhang
Master Student, University of Malaya
Malaysia

An Introduction to
ATINER's Conference Paper Series

ATINER started to publish this conference papers series in 2012. It includes only the papers submitted for publication after they were presented at one of the conferences organized by our Institute every year. The papers published in the series have not been refereed and are published as they were submitted by the author. The series serves two purposes. First, we want to disseminate the information as fast as possible. Second, by doing so, the authors can receive comments useful to revise their papers before they are considered for publication in one of ATINER's books, following our standard procedures of a blind review.

Dr. Gregory T. Papanikos
President
Athens Institute for Education and Research

This paper should be cited as follows:

Balakrishnan, V. and Zhang, X., (2014) **"Using Text Highlight and Page Reviews to Improve Search Engine Result Page"** Athens: ATINER'S Conference Paper Series, No: **COM2014-0939.**

Using Text Highlight and Page Reviews to Improve Search Engine Result Page

Vimala Balakrishnan
University of Malaya
Malaysia

Xinyue Zhang
University of Malaya
Malaysia

Abstract

Studies have shown that user interactions with search results can provide relevance information on the documents displayed in the results. Therefore this study was undertaken with the aim to improve search results, particularly the document relevancy by using two implicit feedback techniques, namely text highlight and the number of times a page is reviewed. Laboratory experiments were conducted and results showed improved relevancy when both these techniques were integrated. To be specific, the mean average precision (MAP) was found to be approximately 74.2%, and the average precision at level 100 (11-standard levels) was 56.8%, compared to baseline (55% and 44%, respectively). Additionally, the normalized discounted cumulative gain also exhibits improved relevancy for the combined techniques (81%) than the baseline (69%). Overall results also indicate that more relevant documents were returned when both implicit feedback techniques were integrated compared to single technique.

Keywords: Document relevancy; user interaction; implicit feedback; page review count; text highlight;

Introduction

The amount of textual information in electronic form has been growing rapidly over time since the emergence of computers. As a consequence, user's demand for retrieving the stored information available in the large collection of texts has become more and more urgent as people has realized to make the efficient use of the information. This resulted in the birth of information retrieval – a science of searching for information from a large set of documents collection (Baeza-Yates & Ribeiro-Neto, 1999). One classic example of information retrieval is the web search engine, such as Google Search.

One of the concepts that are closely related to information retrieval would be relevance feedback – which obtains additional information beyond users' initial requests. Relevance feedback can be categorized into two, i.e., explicit and implicit feedbacks. Explicit feedback requires the users to rank or rate documents explicitly based on their relevancy. This technique forces the users to engage in additional activities, hence it can be difficult to collect the required data. On the other hand, implicit feedback estimates relevancy based on user's behaviour such as reading time, saving, and printing. Apart from removing the cost to the user of providing feedback, implicit feedback enables one to collect a large quantity of data without any additional user activity. Designing feedback techniques is still a research challenge, and various studies are being conducted to capitalize the benefits of explicit (Bulo et al., 2011) and implicit (Fox et al., 2005; White et al., 2006; Lee et al., 2008) techniques. With this in mind, this paper aims to improve document relevancy by integrating two implicit feedback techniques, namely text highlight and page review counts. This will be achieved by comparing the results with the baseline results (i.e. without any feedbacks).

The remainder of the paper is organized as follows: the following section presents the background for this study, followed by the research methodology. The results and discussions are presented next. Finally, the paper is concluded.

Background

Implicit feedback has been used to retrieve, filter and recommend a myriad of items, from academic articles to movies and books. As highlighted previously, implicit feedback technique monitors user's search behaviour, unobtrusively and removes the need for the users to explicitly indicate the relevant documents.

Common methods used in gathering implicit feedback include document reading time, scrolling, printing, clicking and browsing history, among others. For instance, studies have found document reading or display time to indicate users' interest (Kelly & Belkin, 2004) and (White et al., 2003). Additionally, click data (i.e. a series of pages selected for viewing in a search session) were used to train retrieval algorithms to re-rank results based on users' clicks. For example, reference (Smyth et al., 2005) re-ranked results based on the selection

history of previous searchers and found this technique to improve search performance. Similarly, click data gathered from a search session were found to increase both precision and recall of the feedback data (Jung et al., 2007). Studies combining several implicit feedback techniques also showed improved relevancy. For example, time spent on a page and scrolling were found to predict relevance in Internet browsing (Claypool et al., 2001).

Recently, new techniques were explored by some researchers to capture richer interactions with users including cursor movements and text selections (White & Buscher, 2012) and (Buscher et al., 2012). Text selection or highlight refers to users marking blocks of text for further manipulation. Generally users select or highlight text to copy-and-paste to another application, or as a query for another search. This may indicate that the particular document is relevant to the user, and researchers in (White & Buscher, 2012) showed that using text highlight as implicit feedback can significantly improve search result relevance. Similarly, researchers in (Golovchinsky et al., 1999) reported a significant improvement of the annotation-based feedback technique compared to explicit feedback. The annotations in their study included highlighting, underlining, circling and making notes in the document margin.

Another idea that is still considered new is the number of times a page is reviewed or reexamined. A user who finds a particular document to be interesting or relevant will most probably return back to the same document in future. Therefore, we can safely assume that the higher the page review count, the more relevant the page is to the particular user (or users with similar interests).

Based on the reviews above, the novelty of this study would be to examine the effect on search results when both the implicit feedback techniques are combined.

Research Methodology

Text Highlight and Page Review Count

Our technique re-ranks the original search results by taking into consideration users' behaviour, particularly text highlighting and page reviewing. Equation one below shows how the re-ranking is done.

$$NS = Source * (\alpha + \beta * \sum_{j=1}^{n}(TH + PRC)) \quad (1)$$

In the above equation, *NS* refers to the re-ranked score which includes text highlight (TH) and the page review count (PRC) for a retrieved document. Source refers to the original weight of *tf-idf*, α and β are constants to control and justify the weight of the ranking algorithm and implicit feedbacks. The initial values for α is 0.5 and β is 0.15. If a document is not relevant for a given query, *NS* then equals to 0.

Evaluation Metrics

Our proposed technique was evaluated based on the top 25 retrieved documents using three popular evaluation metrics: eleven-point interpolated average precision, mean average precision (MAP) and normalized discount cumulative gain (NDCG). The total number of queries was twenty.

Eleven-Point Interpolated Average Precision

This metric measures the interpolated precision over all 20 queries in this study at 11 recall levels [0.0 - 1.0]. Then, for each recall level the arithmetic mean of the interpolated precision is calculated.

Mean Average Precision

The mean average precision (MAP) measures the query average precision across all recall levels for top 25 documents in this study, based on equation two below:

$$\text{MAP}(Q) = \frac{1}{|Q|} \sum_{j=1}^{|Q|} \frac{1}{m_j} \sum_{k=1}^{m_j} \text{Precision}(R_{jk}) \quad (2)$$

In the formulation, Q refers to the number of queries in the search (i.e. 20), m_j presents the number of relevant documents which were retrieved in top 25 documents, and R_{jk} is the precision for query j.

Normalized Discounted Cumulative Gain

Finally, the discounted cumulative gain (DCG) is a metric that combines graded relevance assessments effectively. DCG works on the premise that highly relevant documents are preferable at the top of the ranking than mildly relevant ones. As search results vary according to queries, the DCG should be normalized across all queries, as depicted in equation three:

$$NDCG_p = \frac{DCG_p}{IDCG_p} \quad (3)$$

Evaluation Setup

The CACM test collection is deemed to be suitable for small-scale experiments and easy to use for full-text based information retrieval systems. Therefore, CACM was used to evaluate the information retrieval system in this study. A total of 20 topics were selected.

A web based system was developed to gather implicit feedback from users in a laboratory. A total of 23 users were given a pre-fixed list of keywords and queries. They were also informed to browse through the relevant documents and to highlight the text that was deemed to be relevant. The web based system automatically tracks the number of times texts have been highlighted and also the number of times pages have been reviewed. These collected information train the system automatically and re-ranks the results based on users' feedbacks. No explicit feedbacks were collected.

Results and Discussion

Eleven-point Interpolated Average Precision

Table 1 presents the average interpolated precision at each of the eleven recall levels. Text highlights and page review counts techniques have better precisions compared to the baseline at every recall level. Similarly, the combination of text highlight and page review also produced better precision as opposed to the baseline.

Table 1. *Average Interpolated Precision for Top 20 Documents*

Recall (%)	Baseline Precision (%)	TH Precision (%)	PRC Precision (%)	TH+PRC Precision (%)
0	71.4	89.9	73.5	87.6
10	66.4	89.9	73.5	87.6
20	55.9	77.0	62.0	74.2
30	56.8	71.7	58.5	71.3
40	45.5	68.6	56.1	69.1
50	43.5	62.5	54.6	62.6
60	43.4	61.0	51.1	62.4
70	44.2	60.1	47.4	56.8
80	44.2	59.3	47.4	56.8
90	44.2	59.3	47.4	56.8
100	**44.2**	**55.0**	**47.4**	**56.8**

Mean Average Precision

Table 2 above shows the MAP results for the top 25 documents used in the experiments. It can be noted that in general, the implicit feedback techniques improved the document relevancy as opposed to the baseline results. Additionally, the results also indicate that combining both text highlight and number of times a page is reviewed has a higher precision compared to single techniques.

Table 2. *MAP Results for Top 25 Documents*

Baseline	TH	PRC	TH+PRC
55.01	71.36	63.93	74.17

Normalized Discounted Cumulative Gain
Table 3. *Evaluation Results of NDCG Top 25 Documents*

Topic no.	1	2	3	4	5	6	7	8	9
Baseline	0.55	0.5	0.48	0.49	0.51	0.51	0.54	0.56	0.57
TH+PRC	0.75	0.73	0.67	0.65	0.67	0.67	0.69	0.71	0.72
Topic no.	10	11	12	13	14	15	16	17	18
Baseline	0.58	0.59	0.6	0.61	0.62	0.62	0.63	0.63	0.64
TH+PRC	0.72	0.73	0.74	0.74	0.76	0.77	0.77	0.77	0.78
Topic no.	19	20	21	22	23	24	25		
Baseline	0.66	0.67	0.67	0.68	0.68	0.69	0.69		
TH+PRC	0.79	0.79	0.79	0.79	0.8	0.81	0.81		

Table 3 shows the NDCG results for the top 25 documents for combined techniques and the baseline. It can be noted that the combination of text highlight and the number of times a page is reviewed produce improved document relevancy, with a final precision of 81% as opposed to 69% for the baseline.

Precision Histogram

Finally, the histogram in Figure.1 below depicts the relevancy improvement when text highlights and page review counts were combined over the baseline. The histogram shows the integrated techniques results in approximately 55% (11/20) precision improvement compared to the baseline.

Figure 1. *A Precision Histogram (TH+PRC/ Baseline) for 20 Topics*
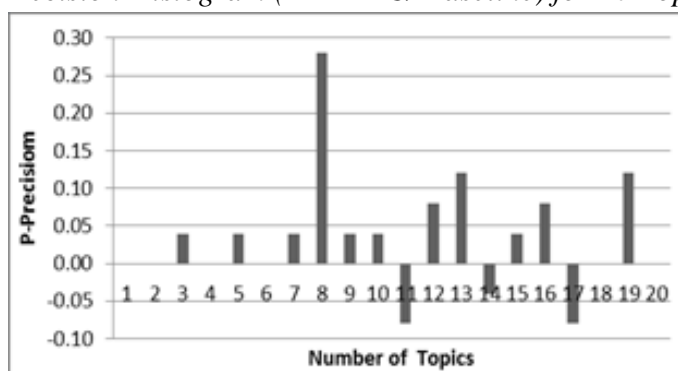
The results in the study clearly indicate that combining both text highlights and page review counts improve the relevancy of document search results. Furthermore, using any of these techniques separately also showed better relevancy precisions than baseline results.

Conclusions

Finding relevant documents from an enormous amount of search results can be a daunting task to many, therefore many studies attempted to utilize data gathered from user interactions. Similarly, in this study two implicit feedback techniques were examined and integrated to improve the document relevancy to the user. Experiments were conducted and data were collected based on their activities, specifically the amount of text being highlighted and also the number of page reviews. Analyses of these data revealed that document relevancy were improved compared to the baseline results.

References

Baeza-Yates, R. and Ribeiro-Neto, B. 1999. Modern Information Retrieval. New York: US: Addison-Wesley.

Bulo, S. R., Rabbi, M. and Pelillo, M. 2011. Content-based image retrieval with relevance feedback using random walks. Pattern Recognition, vol. 44, pp. 2109–2122.

Buscher, G., White, R. W., Dumais, S. and Huang, J. 2012. Large-scale analysis of individual and task differences in search result page examination strategies. In Proceedings of the fifth ACM International Conference on Web Search and Data Mining, p. 373–382.

Claypool, M., Lee, P., Wased, M. and Brown.D. 2001. Implicit interest indicators. In Proceedings of the 6th International Conference on Intelligent User Interfaces, p. 33–40.

Fox, S., Kamawat, K., Mydland, K. M., Dumais, S., and White, T. 2005. Evaluating implicit measures to improve the search experiences. ACM Transactions on Information Systems, vol. 23, pp. 147–168.

Golovchinsky, G., Price, M. N., and Schilit, B.N. 1999. From reading to retrieval: freeform ink annotations as queries. In Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, p. 19–25.

Jung, S., Herlocker, J. L., and Webster, J. 2007. Click data as implicit relevance feedback in web search. Information Processing and Management, vol. 43, pp. 791–807.

Kelly,D. and Belkin, N.J. 2004. Display time as implicit feedback: understanding task effects. In Proceedings of the 21 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, p. 377–384.

Lee,T. Q., Park,Y. and Park, Y.T. 2008. A time-based approach to effective recommender systems using implicit feedback. Expert Systems with Applications, vol. 34, pp. 3055–3062.

Smyth, B., Balfe, E., Freyne, E., Briggs, P., Coyle, M., and Boydell, O. 2005. Exploiting query repetition & regularity in an adaptive community-based web search engine. User Modeling and User-Adapted Interaction: The Journal of Personalization Research, vol. 14, pp. 383–423.

White, R.W and Buscher.G. 2012. Text Selections as Implicit Relevance Feedback. In Proceedings of the 35th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval , p. 1151 - 1152.

White, R.W., Jose, J. M., and Ruthven, I. 2006. An implicit feedback approach for interactive information retrieval. Information Processing and Management, vol. 42, pp. 166–190.

White,R. W., Jose,J. M. and Ruthven, I. 2003. A task-oriented study on the influencing effects of query-biased summarization in the web searching. Information Processing and Management, vol. 39, pp. 669–807.