Leveraging Crowdsourcing Heuristics to Improve Search in Wikipedia

Yasser Ganjisaffar, Sara Javanmardi and Cristina Lopes
School of Informatics & Computer Science
University of California, Irvine
Irvine, CA, USA
{yganjisa, sjavanma, lopes}@ics.uci.edu

ABSTRACT

Wikipedia articles are usually accompanied with history pages, categories and talk pages. The meta–data available in these pages can be analyzed to gain a better understanding of the content and quality of the articles. We analyze the quality of search results of the current major Web search engines (Google, Yahoo! and Live) in Wikipedia. We discuss how the rich meta–data available in wiki pages can be used to provide better search results in Wikipedia. We investigate the effect of incorporating the extent of review of an article into ranking of search results. The extent of review is measured by the number of distinct editors who have contributed to the articles and is extracted by processing Wikipedia's history pages. Our experimental results show that re–ranking search results of the three major Web search engines, using the review feature, improves quality of their rankings for Wikipedia–specific searches.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Information filtering; H.3.5 [**Online Information Services**]: Web based Services

1. INTRODUCTION

Wikipedia, a massive repository of knowledge, is most useful when its articles are well–organized and easily accessible. Web search engines have been successful in making the Web content accessible for more than a decade, and they succeed in searching Wikipedia as well. However, the special features introduced by wiki technology make search in the domain of Wikipedia different from the traditional Web content.

A recent study on Wikipedia [5] shows that high–quality articles in Wikipedia benefit from higher number of edits and distinct contributors. Other studies on prediction in crowdsourcing systems show that the average of predicted scores by the crowd becomes more reliable as the size of the crowd increases [2, 4]. Some interpret this fact by the "law

of large numbers" in which the mean of a sample of independent observations from a given population approaches the population mean as the sample size increases [3]. According to these observations, we expect the quality of the Wikipedia entries to improve as they go through iterations of edits by different users. We propose a review–based ranking algorithm to improve quality of search in the domain of Wikipedia. We show that the quality of the rankings by the current major Web search engines can be improved by incorporating the proposed heuristic in their ranking schemes.

The contributions of this work are twofold. First, the empirical study of search performance by the three major search engines in Wikipedia provides valuable evidence that not all search engines are equal. Second, the review–based heuristic proposed here results in considerable improvements for the two least–performing search engines.

2. CURRENT STATUS

Based on the differences between wiki content and general Web content, we set out to study the effectiveness of major Web search engines in searching Wikipedia. In the remainder of this section, we present the current state of the three major Web search engines, namely Google, Yahoo! and Live search, in terms of retrieval effectiveness in Wikipedia.

2.1 Method

To compare the effectiveness of the rankings of search results, we use the evaluation metric called Normalized Discounted Cumulative Gain (NDCG) [1].

We asked seven graduate students in different majors to use our interface for searching Wikipedia, and to label search results as highly relevant (HR), relevant (R), and irrelevant (IR). Students were asked to search for both special topics related to their major and general topics. All three search engines have application programming interfaces (API) that allow programs to submit queries and get the search results. After a query is submitted through our interface, it is submitted to Google, Yahoo!, and Live APIs. Queries are appended with "site:en.wikipedia.org" to restrict domain of search to English Wikipedia.

For presenting search results to users for labeling, we used the *pooling* method. Each query is submitted to the three search engines and the top 10 results from each search engine are added to a pool. Duplicates in search results are removed and the final set of results is *randomly* presented to the user for labeling. A total of 240 queries were submitted and 3, 410 results were labeled.

2.2 Results

Figure 1 shows NDCG values for positions 1 through 10 for Google, Yahoo!, and Live search engines. For the top 1 search results, the three search engines have similar gains. However, Live outperforms the other two for the top 2 through 10 results. Given that Live search treats Wikipedia pages differently than other pages[1], it seems to be using some Wiki–specific information in ranking its search results.
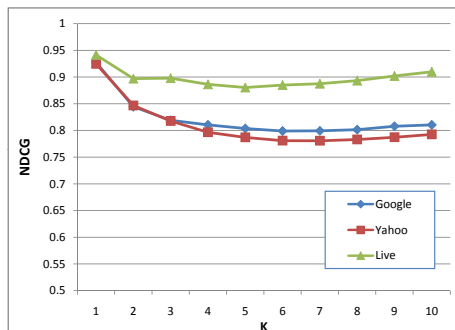

Figure 1: Quality of search results of major Web search engines in the domain of Wikipedia

3. IMPROVING SEARCH IN WIKIPEDIA

The open editing model of Wikipedia allows users to *review* and edit previously contributed content by other users in order to improve its quality. Some studies on prediction in crowdsourcing systems show that the average of predicted scores by the crowd becomes more reliable as the size of the crowd increases [2, 4]. Similarly in Wikipedia, we expect the quality of the content to improve as the number of contributors expands. This is also consistent with the results reported in [5] that show Wikipedia's featured articles benefit from higher number of edits and distinct contributors. Based on these observations, we investigate if the extent of review of articles can improve the quality of rankings of Wikipedia articles in search results. Extent of review of articles is measured by the number of contributors who have edited these articles.

We downloaded the dump of the Wikipedia history released in October, 2008[2] and extracted the number of distinct editors contributing to each article. We assigned a review score between 0 and 10 to each of the articles by placing the number of editors on a logarithmic scale.

We conducted an experiment to analyze if adding extent of review as a new feature can help improve rankings of the three search engines for Wikipedia articles. We divided the data that had been labeled from the 240 queries of our previous experiment, into two groups: 120 queries as a "training set" and 120 queries as a "test set". We used the training set to train a support vector machine (SVM) classifier[3] in order to see if doc_i should be ranked higher than doc_j, for the query q based on the position of these two documents among the top 10 results returned by search engine and the assigned review score of each of these two documents.

---

[1]http://blogs.msdn.com/livesearch/archive/2008/05/30/wikipedia-gets-big.aspx
[2]http://download.wikimedia.org/enwiki/20081008/
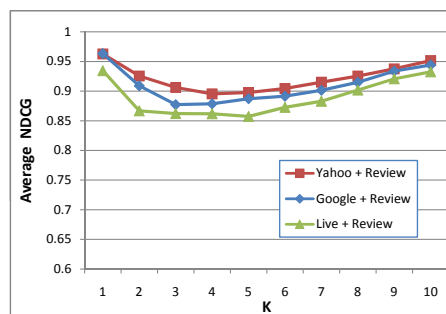[3]http://svmlight.joachims.org/

Figure 2: Quality of search results of major Web search engines in the domain of Wikipedia and impact of re–ranking their results using the review–based algorithm

Figure 2 plots the average NDCG for k from 1 to 10 for the three search engines on our test set and compares results with those gained by using the SVM classifier to rank search results. The results show that incorporating the review score improves the quality of ranking; but this is more apparent for Google and Yahoo! search engines. In the case of Live search engine, review scores improve the quality of ranking only when we consider more than 7 positions in the search results. Since review score is an extremely simple feature to calculate, totalling the number of editors of the article, it is very promising to see such improvement.

4. CONCLUSION

In this paper, we explored the differences between wiki pages and traditional web pages and studied the effectiveness of search in the domain of Wikipedia between three major Web search engines: Google, Yahoo! and Live Search. We introduced a very simple review–based ranking algorithm to rank Wikipedia articles in search results. We concluded from the results of our experiment that the number of editors, which can be easily extracted from the history of articles, can be used to improve the quality of ranking of search result.

5. REFERENCES

[1] K. Järvelin and J. Kekäläinen. IR evaluation methods for retrieving highly relevant documents. In *Proceedings of the ACM SIGIR conference*, pages 41–48, New York, NY, USA, 2000. ACM.

[2] D. Reeves and D. Pennock. How and when to listen to the crowd. Electronically.

[3] D. V. Schroeder. Book reviews. *American Journal of Physics*, 74(3), March 2006.

[4] J. Surowiecki. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. Doubleday, 2004.

[5] D. M. Wilkinson and B. A. Huberman. Assessing the value of cooperation in wikipedia. *Firstmonday*, 2007.