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, the largest encyclopedia on the Web, is often seen as the most successful example of crowdsourcing. The encyclopedic knowledge it accumulated over the years is so large that one often uses search engines, to find information in it. In contrast to regular Web pages, Wikipedia is fairly structured, and 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. Built on the studies on "Wisdom of Crowd" and the effectiveness of the knowledge collected by a large number of people, we investigate the effect of incorporating the extent of review of an article in the quality of rankings of the search results. The extent of review is measured by the number of distinct editors 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.

1. INTRODUCTION

Web 2.0 is the second generation of the Web that promotes crowdsourcing, the process of outsourcing a task to a large group of people, in the form of an open call [21, 19]. Using wiki technology, Wikipedia has become the largest online encyclopedia, used as a reference for encyclopedic knowledge [10]. Wikipedia's model of knowledge creation is relatively novel, allowing anyone to enter and edit content. This has enabled its rapid expansion: since its inception in 2001, Wikipedia has grown to encompass 11.9 million articles in 265 languages generated from 615 million edits by 15 million contributors[1]. Its current position as the 7th most visited website[2] confirms its usefulness and popularity.

Wikipedia, as 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 a decade, and they succeed in searching Wikipedia too. However, the special features introduced by wiki technology make search in the domain of Wikipedia different from traditional Web content. For example, unlike traditional Web pages owned by individuals, wiki pages are edited and maintained by a large number of co–contributors thereby the reliability of the content is not correspondent to an individual or an organization. Hence, evaluating the reliability and quality of the collected knowledge by the crowd can play an important role in devising heuristics for search tailored for Wikipedia and other such knowledge repositories.

Some studies show that a broad spectrum of human activities requiring critical decisions illustrate the greater reliability of judgments by crowds than by experts [18, 20]. A recent study on Wikipedia [22] shows that high–quality articles in Wikipedia benefit from higher number of edits and distinct contributors. Other studies [16, 18] 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. 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 [17]. 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 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 not only results in considerable improvements for the two least-performing search engines, but it suggests that the future of search on the Web will need to take into account the social activities that are now taking place in it.

[1] http://meta.wikimedia.org/wiki/List_of_Wikipedias

[2] http://www.alexa.com

The reminder of this paper is organized as follows. In Section 2, we analyze the differences in search performance in Wikipedia between the three major Web search engines, namely Google, Yahoo! and Live. Section 3 shows how those search engines are affected by adding the additional review–based heuristic. Section 4 presents related work, and Section 5 concludes the paper.

2. CURRENT STATE OF SEARCH IN WIKIPEDIA

Major Web search engines have been successful in making the Web content easily accessible to users for more than a decade. Although wikis are built on top of the Web and are part of it, there are differences that make wiki content distinguishable from traditional Web content. The main differences are summarized as follows:

- *Meta–data information.* Each Wikipedia article is accompanied with history, talk, and category pages. History pages comprise old revisions of the wiki text as well as the record of the timestamp and username of the contributor. Category pages provide semantic knowledge about the concept which is presented by the article. This meta–data can provide significant insight into the content of the article and can be used to provide higher quality search results to users.
- *Wikipedia link structure.* The semantics of links that connect wiki pages can be different from the semantics of the links in traditional Web pages. Whereas in Web documents an author can arbitrarily link his page to any other page, whether there is a topical relation or not, a significant fraction of links in Wikipedia point to semantically related content [8, 9]. In addition, some links are inserted automatically by Wikipedia's registered bots[3] serving particular purposes. In aggregate, Wikipedia's links can not simply be interpreted as votes for authoritativeness of the target page as it is in traditional Web pages. For example, articles representing years like "2008" have a large number of in–links compared to most of the other articles. This issue suggests that link–based ranking algorithms like PageRank [15] might be less effective in the domain of Wikipedia.
- *Multi–ownership.* Wikipedia articles are created and edited by co–contributors. Thus, reliability of a wiki page is not necessarily correspondent the reputation of a single entity. The open editing model of Wikipedia allows users to contribute to any article regardless of their attributions or expertise.

Based on these differences, we set out to study the effectiveness of major Web search engines in searching Wikipedia. In the reminder 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 and freshness of their index.

2.1 Effectiveness of Search Results

2.1.1 Method

[3]http://en.wikipedia.org/wiki/Wikipedia:Bot_policy

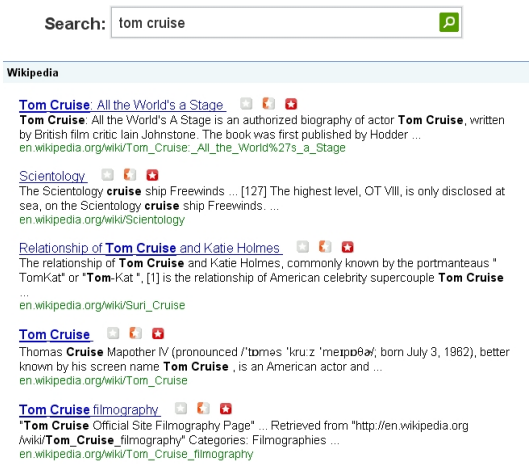

Figure 1: A screenshot of how results are presented to users for labeling

To compare the effectiveness of the rankings of search results, we use the evaluation metric called Normalized Discounted Cumulative Gain at top k (NDCG(k) for short) [6]. This metric measures the usefulness, or *gain*, of a document based on its position in the results list. The gain is accumulated cumulatively from the top of the results list to the bottom with the gain of each result discounted at lower ranks. Two assumptions are made in using NDCG:

1. Highly relevant documents are more useful when appearing earlier in a search engine results list (have higher ranks)
2. Highly relevant documents are more useful than marginally relevant documents, which are in turn more useful than irrelevant documents.

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 of the 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 [7]. 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 user for labeling. Figure 1 shows a screenshot of how results are presented to users for labeling. Users label search results by clicking on the appropriate star beside each search result. After each click, a message is sent to our logging server which records the labeling. A total of 240 queries were submitted and 3,410 results were labeled.

HR–labeled articles are considered more relevant than R–labeled articles so we adapt NDCG to distinguish these two different degrees of relevance:

$$NDCG_q(k) = \frac{1}{N_q(k)} \sum_{p=1}^{k} \frac{2^{r(p)} - 1}{\log(1+p)} \quad (1)$$

NDCG(k) for a query is computed by summing up the gains from position $p = 1$ to $p = k$ in results. $r(p)$ is an integer representing the amount of reward given to the article at position p. Reward is considered to be 2 for highly relevant results, 1 for relevant results, and 0 for non–relevant results.

The term N_q is a normalization factor for query q, derived from a perfect ordering of top k articles that would yield a $NDCG_q(k)$ of 1. Intuitively, the perfect ordering ranks all highly relevant articles before all relevant articles. Formally,

$$N_q(k) = \sum_{p=1}^{k} \frac{2^{s(p)} - 1}{\log(1+p)} \quad (2)$$

where,

$$s(p) = \begin{cases} 2 & if\ 1 \le p \le n_q^{HR} \\ 1 & if\ (n_q^{HR}+1) \le p \le (n_q^{HR}+n_q^{R}) \\ 0 & otherwise \end{cases}$$

where n_q^{HR} and n_q^{R} represent the number of highly relevant and relevant articles for query q.

2.1.2 Results

Figure 2 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 rest for the top 2 to top 10 results. Given that Live search treats Wikipedia pages different that other pages (e.g, see [14]), it seems that it is using some Wiki–specific information in ranking its search results.

Wikipedia pages are organized into namespaces. A namespace is a part of Wikipedia including all pages of a particular kind. The *main namespace* contains all encyclopedia articles along with their history revisions. The *user namespace* contains the personal pages of registered users in Wikipedia. Typically, when users search in the domain of Wikipedia, they are looking for encyclopedic knowledge, which is in the main namespace. Hence, a search engine should give more preference to main articles rather than Talk pages, Templates, User pages, or the pages in other non–main namespaces. To see how current search engines consider this issue, table 1 shows the percentage of different non–article results in the top 10 results of the three search engines. Live search gives more preference to articles in the main namespace compared to the other two ones. About 94% of the top 10 results

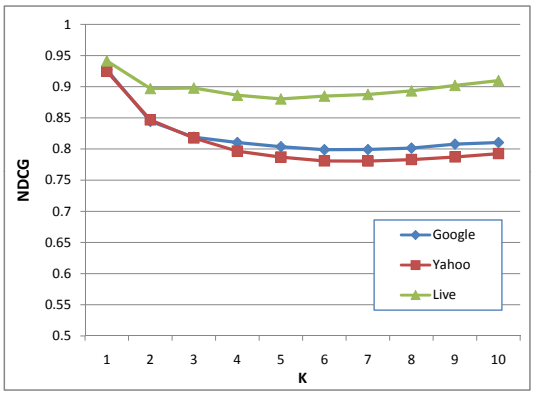

Figure 2: Quality of search results of major Web search engines in domain of Wikipedia

	Google	Yahoo!	Live Search
User Pages	1.7%	1.3%	0%
Talk Pages	0.8%	8.6%	0.7%
Category Pages	3.7%	3.1%	3.8%
Old Revisions	1.1%	0.2%	0.7%
Special Pages	3.2%	1.6%	0.8%
Templates	0.7%	0.3%	0.4%
Media Pages	0.3%	0.3%	0.2%
Sum	**10.4%**	**15.2%**	**5.9%**

Table 1: Percentage of Non-article results in search results returned by the three major Web search engines.

returned by Live search are in Wikipedia's main namespace.

Our study also shows the diversity of the algorithms used by the three major search engines to retrieve and rank relevant results to users' queries, which makes their results different even on Wikipedia as a small fraction of the Web. Table 2 shows that the three search engines agree on the top search result for 61.7% of the queries. It also shows percentage of common pages in top-k results returned by the the search engines. Considering the top 10 results of each search engine, only 23.4% of results are common among them.

2.2 Freshness of Search Index

Being edited by millions of users, Wikipedia articles are updated more frequently compared to traditional static web pages. Therefore, the content indexed by general Web search

	Top 10	Top 5	Top 1
Google & Yahoo!	34.3%	35.7%	66.5%
Google & Live	37.9%	38.1%	69.6%
Yahoo! & Live	36.5%	37.3%	65.1%
All	23.4%	24.9%	61.7%

Table 2: Percentage of common pages in top–k search results returned by the three major Web search engines.

engines may not reflect the current content of Wikipedia articles. It is not trivial for a general Web search engine to keep its index fresh. However, based on the small size of Wikipedia compared to the whole Web, it is much easier for a Wikipedia specific search engine to keep its index up–to–date. Search engines can also monitor Wikipedia's log of recent changes[4] and only crawl articles when they are updated. To study the freshness of the indices of Web search engines for Wikipedia, we conducted an experiment on the three major Web search engines.

2.2.1 Methods

In addition to search results, the search API of Yahoo! and Live also return the last timestamp that pages are indexed and cached. Google API only returns a link to the cached version of the page; following the link, one can extract the the cache timestamp which is displayed at the top of the cached version.

During our previous experiment for evaluating the effectiveness of Web search engines in Wikipedia, we also recorded the timestamps of the cached version of pages in each search engine. For each of the 3,410 pages, we used the Wikipedia API[5] in order to find the timestamp of the first submitted revision after the page was indexed by each search engine. We define *out–dated* period as the time interval between the submission of that revision and the current time (Figure 3). This is the time interval in which the index of the search engine does not reflect the current revision of the page. If no revision is submitted after the page was indexed, out–dated period is considered as zero.

2.2.2 Results

Figure 4 shows the out–dated period for each of the three search engines. Pages are grouped by their PageRank scores. For Google search engine, the out–dated period is less than one day regardless of the PageRank level. Despite the large volume of the World Wide Web, Google keeps its index of Wikipedia articles pretty fresh. Our conjecture is that Google can do it by monitoring the change log of Wikipedia and crawling the pages when they are updated; instead of wasting resources for crawling all pages that have not been updated since the last index time.

For Yahoo! search, the out–dated period is 17.7 days on average for PageRank level of zero and it is 0.9 days for PageRank level of 9. It seems that the crawling policy of Yahoo! is biased towards pages with higher PageRank. For Live search, the out–dated period varies from 25 days to 51 days with the exception of 5 days for pages with PageRank level as 9. These results show that index of Live search for Wikipedia pages is more out–dated compared to Yahoo! and Google search.

3. IMPROVING SEARCH IN WIKIPEDIA

Unlike many traditional models of knowledge and publishing, which attempt to limit content creation to a relatively small group of approved editors in order to exercise strong quality control, Wikipedia articles are published without

[4]http://en.wikipedia.org/w/api.php?action=query&list=recentchanges

[5]http://en.wikipedia.org/w/api.php

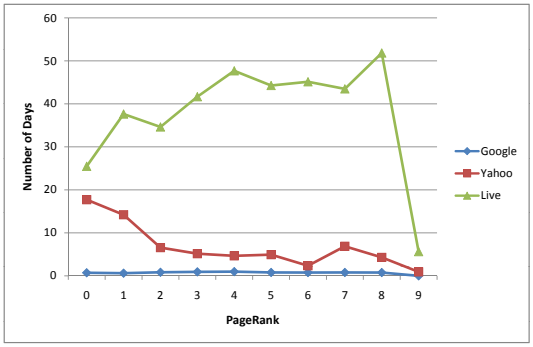

Figure 4: Average length of out–dated period of cached version of Wikipedia pages in major Web search engines

strict prior quality checking. The open editing model of Wikipedia allows users to *review* and edit previously contributed content by other users in order to improve its quality. Because of the radical openness in Wikipedia, the peer reviewing process is scaled up orders of magnitude compared to peer reviewing process of scholarly articles. 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 [16, 18]. 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 [22] that show Wikipedia's featured articles[6] 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.

There is a positive correlation between distinct number of contributors (review scores) and content quality, and also between content quality and PageRank scores [22, 12]. Hence, we expect to see a positive correlation between PageRank and review scores. To verify this hypothesis, we processed English Wikipedia. We crawled the English Wikipedia domain (en.wikipedia.org) in January 2009. The crawlers were programmed to discard redirected pages (about 35% of the pages) and eliminate HTML templates from the downloaded pages. After elimination of duplicates, a total of 2,389,715 pages were indexed. We stored the internal link structure of Wikipedia while crawling. A total of 256,304,639 links were extracted. We used the PageRank algorithm to find the relative importance of articles according to their in–links. The PageRank values converged after 29 iterations. We scaled the PageRank values to a value between 0 and 10 using a logarithmic scale. We downloaded the dump of the Wikipedia history released in October, 2008[7] and extracted number of distinct editors contributed in each article. We assigned a review score between 0 and 10 to each of the articles by

[6]Featured articles are considered to be the best articles in Wikipedia. Before being listed as featured, each article is reviewed by some experts familiar with the subject for accuracy, neutrality, completeness, and style (http://en.wikipedia.org/wiki/Featured_Article).

[7]http://download.wikimedia.org/enwiki/20081008/

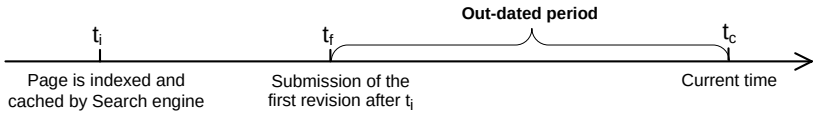

Figure 3: Definition of the out–dated period

scaling the number of editors of the article on a logarithmic scale.

Figure 5 shows the average review scores of articles as a function of PageRank. Review scores increase with the rise of PageRank values which supports the positive correlation between review scores and article quality.

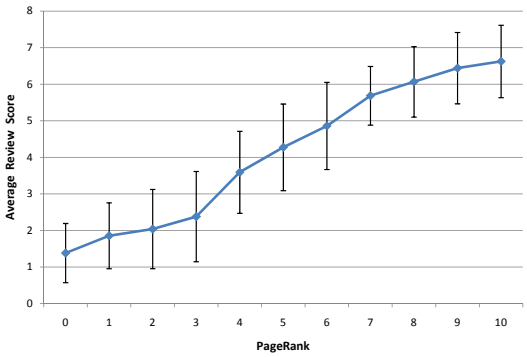

Figure 5: Average review scores of the articles as a function of PageRank values

Table 3 lists the top English Wikipedia articles based on PageRank and Review scores. The list of top articles based on PageRank values is dominated with country pages and Wikipedia's special pages like "Wikipedia Commons" and "Geographic Coordinate System". These pages have a large number of in–links compared to other pages. For example, in most of the articles about United States cities, the latitude and longitude of the city is specified in the page and there is a link to "Geographic Coordinate System" article near these values. Some of these links are inserted by Wikipedia's registered bots that make automated edits in articles. However, all of the edits by bots are counted only once in the list of most reviewed articles.

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 240 queries data set labeled during our previous experiments to a 120 queries training set and a 120 queries test set. We used the training set to train a support vector machine (SVM) classifier[8] in order to see if doc_i should be ranked higher than doc_j, for the query q according to the following features:

- Position of doc_i among the top 10 results returned by search engine.

[8]http://svmlight.joachims.org/

Most Reviewed	United States, 2007, United Kingdom, Canada
	England, New York City, World War II, 2006
	India, Germany, World War I, Wikipedia
	The Beatles, Adolf Hitler, George W. Bush
	Ronald Reagan, Jesus, Bill Clinton, Wiki
	Harry Potter, Led Zeppelin, Hurricane Katrina
	Albert Einstein, Michael Jackson, Star Wars
	The Simpsons, Xbox 360, Wii, Metalica
High PageRank	United States, 2007, United Kingdom, Canada
	England, France, Wikipedia Commons
	Geographic Coordinate System, New York City
	World War II, 2006, India, Germany, Russia
	Scotland, Norway, 2008, English Language, 2005
	Australia, New Zealand, Europe, London, Spain
	Poland, China, Italy, Sweden, Netherlands, Japan
	Brazil, California, 2004, Census, Public domain

Table 3: Top most reviewed and high PageRank articles in English Wikipedia

- Position of doc_j among the top 10 results returned by search engine.
- Difference between positions of doc_i and doc_j.
- Review score of doc_i.
- Review score of doc_j.
- Difference between review scores of doc_i and doc_j on an exponential base[9].

Figure 6 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 it is more apparent for Google and Yahoo! search engines. In case of Live search engine, review scores improve the quality of ranking only when we consider more than 7 positions in 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. RELATED WORK

Web search has been studied as a classic information retrieval problem. Link analysis techniques such as PageRank [15] and HITS [11] measure the popularity of Web pages

[9]Review scores are scaled on a logarithmic scale. Therefore, it is necessary to measure the difference of review scores on an exponential basis.

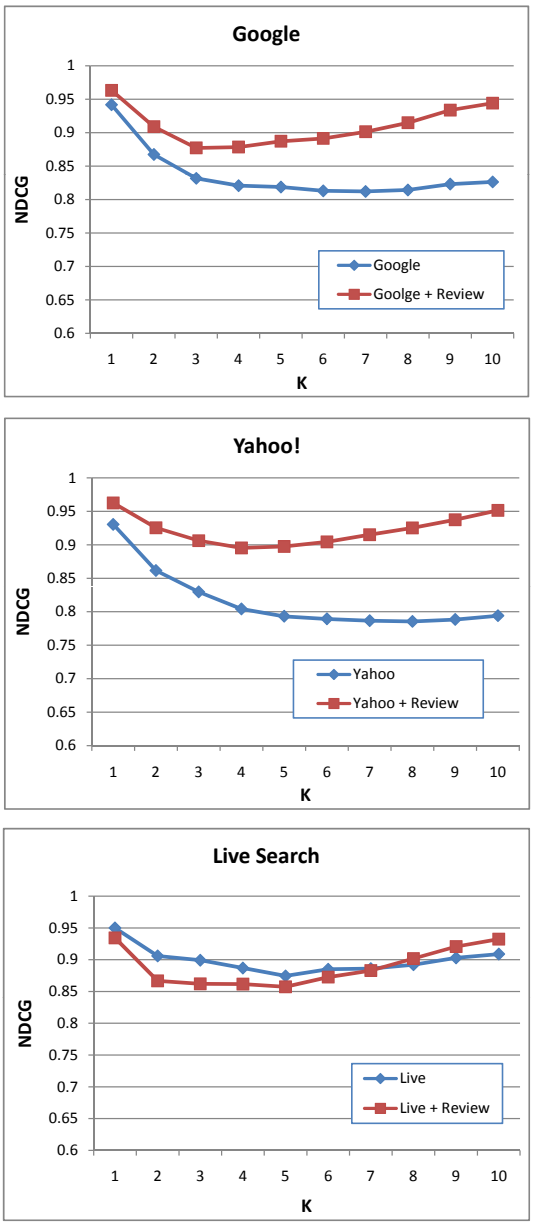

Figure 6: Quality of search results of major Web search engines in domain of Wikipedia and impact of re–ranking their results using the review–based ranking algorithm

based on their interlinking structure, and this popularity is used in ranking search results to yield better search performance. The PageRank score of a page is derived from the scores of pages linking to that page. In HITS algorithm, each page is assigned a hub and an authority score. A page deserves a high hub score when it provides links to authoritative pages and high authority score for being referenced by good hub pages. High PageRank, hub and authority pages are likely to be high quality pages [1, 24].

Besides PageRank and HITS, numerous metrics have been studied in literature to measure Web page quality [13, 1, 23, 24]. In particular, in [24], Zhu and Gauch have studied six metrics for assessing Web page quality, namely *currency*, *availability*, *information–to–noise ratio*, *authority*, *popularity* and *cohesiveness*, and found that incorporating quality metrics generally improved search effectiveness. However, these proposed quality metrics may not work for Wikipedia because most Wikipedia articles follow similar page design and offer equal accessibility.

While it is difficult to measure Wikipedia's overall quality in a definitive way, some studies have tried to assess it in several ways. Some characteristics such as factual accuracy [4] and credibility [3], have been used to compare small samples of Wikipedia articles to their parallel articles in other reputable sources.

In [12], the number of edits and unique editors to an article were suggested as metrics for quality. Built on this study, Wilkinson and Huberman [22] proposed a statistical analysis of Wikipedia and showed that featured articles are distinguishable from the rest by high number of edits and distinct number of contributors. They also verified the validity of this observation for articles with different age and visibility.

In [5], Hu *et al.* have proposed a framework that re–ranks Wikipedia search results considering article quality. They have developed two quality measurement models, Basic and PeerReview. Article quality is derived based on co–authoring data gathered from articles' edit history. According to their experimental results, compared with Google, Wikiseek[10] and Wikipedia's internal search engine, rankings generated based on their quality models are less accurate. However, they showed some improvements in rankings of Wikipedia's search engine and Wikiseek by combining their PeerReview Model to their rankings. The PeerReview model requires processing the whole history of each article for assessing its quality. Consecutive revisions should be compared to find out who has added each word and who has reviewed it. Given the high dynamics of Wikipedia which has resulted in more than 615 million revisions as of this writing, this model is computationally expensive and does not seem to scale well.

5. CONCLUSION & FUTURE WORK

Using Wiki technology, Wikipedia has become a massive online knowledge base. Its articles become most useful when they are well–organized and easily accessible. Although current search engines help users locate content in Wikipedia, there are some new features introduced by wiki technology

[10]A Wikipedia search engine which is no longer active as of 2008

that can be taken into account for ranking the results.

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. Based on the studies on crowdsourcing systems and hypothesis of wisdom of crowd, we introduced a very simple review–based ranking algorithm to rank Wikipedia articles in search results. We described results of our experiments that show the number of distinct editors of an article that can be easily extracted from history of articles can improve ranking of articles presented to users.

More sophisticated information can be extracted from history, category, and talk pages. This meta–data can be used to better evaluate quality of articles and provide better rankings of Wikipedia search results. In the future, we plan to develop a custom search engine for Wikipedia to use the additional information available in Wikipedia to provide better search results. For example, Category information can be used to provide categorized search results. Organizing search results allows users to focus on items in categories of interest rather than having to browse through all the results sequentially. Results of the study in [2] show that users are 50% faster at finding information when search results are organized into categories.

Overall, our study suggests that as the Web becomes more structured around social activities, the heuristics for search will likely have to adapt.

6. REFERENCES

[1] B. Amento, L. Terveen, and W. Hill. Does "authority" mean quality? predicting expert quality ratings of web documents. In *SIGIR '00: Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 296–303, New York, NY, USA, 2000. ACM.

[2] H. Chen and S. Dumais. Bringing order to the web: automatically categorizing search results. In *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 145–152, New York, NY, USA, 2000. ACM.

[3] T. Chesney. An empirical examination of wikipedia's credibility. *Firstmonday*, 11(11), November 2006.

[4] J. Giles. Internet encyclopaedias go head to head. *Nature*, 438(7070):900–901, December 2005.

[5] M. Hu, E.-P. Lim, A. Sun, H. W. Lauw, and B.-Q. Vuong. On improving wikipedia search using article quality. In *WIDM '07: Proceedings of the 9th annual ACM international workshop on Web information and data management*, pages 145–152, New York, NY, USA, 2007. ACM.

[6] K. Järvelin and J. Kekäläinen. Ir evaluation methods for retrieving highly relevant documents. In *SIGIR '00: Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 41–48, New York, NY, USA, 2000. ACM.

[7] K. S. Jones and C. van Rijsbergen. Report on the need for and provision of an "ideal" information retrieval test collection. Technical Report British Library Research and Development Report 5266, University of Cambridge, Computer Laboratory, 1975.

[8] J. Kamps and M. Koolen. *Advances in Information Retrieval*, chapter The Importance of Link Evidence in Wikipedia, pages 270–282. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2008.

[9] J. Kamps and M. Koolen. Is Wikipedia link structure different? In *WSDM '09: Proceedings of the second ACM international conference on Web search and data mining*, New York, NY, USA, February 2009. ACM.

[10] B. Keim. News feature: Wikimedia. *Nature Medicine*, 13:231–233, March 2007.

[11] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.

[12] A. Lih. Wikipedia as participatory journalism: Reliable sources? metrics for evaluating collaborative media as a news resource. In *Proceedings of the International Symposium on Online Journalism*, pages 16–17, 2004.

[13] T. Mandl. Implementation and evaluation of a quality-based search engine. In *HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia*, pages 73–84, New York, NY, USA, 2006. ACM.

[14] T. O. B. of the Live Search Team. Wikipedia gets big. Electronically.

[15] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998.

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

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

[18] 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.

[19] D. Tapscott and A. D. Williams. *Wikinomics: How Mass Collaboration Changes Everything*, pages 70–77. Portfolio Hardcover, December 2006.

[20] P. E. Tetlock. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton University Press, 2005.

[21] M. Waldrop. Big data: Wikiomics. *Nature News*, 455:22–25, September 2008.

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

[23] Y. Zhou and W. B. Croft. Document quality models for web ad hoc retrieval. In *CIKM '05: Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 331–332, New York, NY, USA, 2005. ACM.

[24] X. Zhu and S. Gauch. Incorporating quality metrics in centralized/distributed information retrieval on the world wide web. In *SIGIR '00: Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 288–295, New York, NY, USA, 2000. ACM.