A Study on Improving Web Search Ranking with Social Tagging

Shihn-Yuarn Chen

Ph.D. Student
Department of Computer Science, National Chiao Tung University
E-mail: sychen@cs.nctu.edu.tw

Hao-Ren Ke

Professor
Graduate Institute of Library and Information Studies, National Taiwan Normal University
E-mail: clavenke@ntnu.edu.tw

Chuen-Tsai Sun

Professor
Department of Computer Science, National Chiao Tung University
E-mail: ctsun@cs.nctu.edu.tw

Abstract

With increasing volume of Internet content, search engines have been used extensively by average users. Improving search quality of search engines is always an important issue that has attracted much interest from the research community. With the booming of Web 2.0, social tags are becoming widely available online, as tagging systems enable users to describe Web pages interesting for themselves with personal annotations and community wisdom. Social tags have much potential to help improve Web search, because a bookmarked Web page must be interesting or useful for the user or similar users, and tags of the bookmark provides complementary or concise representation of the Web page. To leverage the community wisdom, this study uses social tags to improve the ranking of search results. In consideration of time efficiency, document expansion with tagging information was employed at indexing time, and document re-ranking with tagging information was applied at query time. Two real world datasets, MSN Searchlog 2006 and Web pages crawled via Yahoo! API, were used for experiments and the results showed significant improvements on NDCG (>10%). Besides, for more thorough evaluation, this study constructed a new PETS dataset, and the evaluation results showed that the proposed approach improved precision about 23%.

Keywords: Social tagging, information retrieval, document expansion, reranking

1 Introduction

Given enormous online information, search engines are important tools for assisting users to locate valuable online information. In spite of remarkable achievements on search engines made by the information retrieval community, users' expectation of better search quality always keeps on raising [1]. Search quality covers various respects, such as ranking of results, response time, precision, recall, etc.

Due to the prevalence of Web 2.0 applications, social tags have become a new type of data widely available on the Internet. Social tagging enables users to describe and annotate a resource easily and directly by free keywords. It has been used by many Web 2.0 applications, such as online shopping (amazon.com), photo sharing (flickr.com), and social bookmarking (delicious.com). When a user thinks a Web resource is useful or interesting, he/she can freely add some tags to it. One major goal of social tagging is to help a user quickly re-find the useful or interesting information in the future.

Researchers have attempted to use social tagging to improve search. Schenkel et al. [2] presented a framework to improve search, and the data was collected from librarything.com. They used community wisdom and scored documents based on social relations and semantic relations among tags and items. Heymann et al. [3] compared new Web pages bookmarked by users of delicious.com with pages indexed by Yahoo! search engine. They concluded that social bookmarking had not yet exerted immense impact on Web search, because the scale of social bookmarks was much smaller than the whole Web pages (~1/1000).

Bischoff et al. [15] evaluated the effectiveness of applying tags on Web search, Bao et al. [16] used a dataset crawled from delicious.com to show social tagging benefits search. Noll and Meinel [17] and Xu et al. [18] further applied social tagging information to facilitate personalizing Web search. However, these approaches only collected data from social tagging service sites. Thus, most Web pages in such kind of dataset are tagged, this cannot reflect the real world distribution of tagged and un-tagged Web pages.

Although social tags may benefit to Web search, according to the authors' knowledge and survey, there are no reported experiment results, which are based on users' search logs to show the effectiveness of social tags improving Web search. Thus, we proposed an approach to improve search result ranking quality by document expansion and re-ranking, and carried out the experiments presented in the previous work [4] and this paper.

This study focuses on applying social tagging data to solve the vocabulary mismatch problem and to improve search quality. This study modifies search engines at indexing phase and query phase. At the indexing phase, the original documents are expanded with social tagging data. At the query phase, re-ranking technique is applied on the search result list based on tagging information. The reset of the paper is organized as follows. The vocabulary mismatch problem is discussed in Section 2. Section 3 introduces the datasets and evaluation metrics. The proposed approach is detailed in Section 4. Section 5 depicts the evaluation results and discussions, and the conclusion and future work are in Section 6.

2 Vocabulary Mismatch and Document Expansion

Vocabulary mismatch is a well-known problem for search engines. Vocabulary mismatch means that two phrases (one query phrase and one document phrases) are semantically equivalent, but they shares no terms in common. For example, "MLB" and "Major League Baseball" have totally different terms, but are semantically the same. The information retrieval community has developed several techniques to solve this problem, include stemming [5][6], LSI [7], translation [8], and vocabulary expansion.

Research of vocabulary expansion can be categorized into two major classes: query expansion and document expansion [9-11]. Query expansion is executed at query run time. While the search engine receives the query, terms related to the original query are firstly extracted, then added these terms into the original query, and finally the expanded query is executed. For example, if a user submits "car" as a query, a related word "automobile" can be added. The query after expansion becomes "car automobile". In this manner, a document which contains the word "automobile", but does not contains "car", can be retrieved and returned to the user. Query expansion benefits the quality of information retrieval, but the time of query execution is sacrificed. Document expansion is very similar to query expansion, except document expansion expands a document at indexing phase by adding relevant words. Singhal and Pereira [10] expanded transcribed documents with related terms from a side corpus, and gained a relative average precision improvement of 12% compared with other techniques.

Although the effectiveness of query expansion is better than document expansion, document expansion is widely used because the query response time is affected less than query expansion. There are two main categories of document expansion approaches, document centric expansion and term centric expansion [9]. For document centric approaches, a document, d, is run as a query to retrieve relevant documents from a corpus. The top n significant terms of returned results are appended to d. For term centric approaches, a query q is submitted to retrieve relevant documents from a corpus. The top n significant terms of return results are then collected. Finally, documents which contain these n terms are expanded with the query q. Term centric approaches are considerably faster than document centric approaches, but while the corpus grows, the previously expanded terms probably become sub-optimal.

Billerbeck and Zobel [9] showed that both document expansion and query expansion improved the mean average precision (MAP), precision at 10 (P@10) and precision at the number of relevant documents (R-Pr.). Although query expansion brings more significant improvement than document expansion, in consideration of the query efficiency, document expansion may be more desirable for a practical system.

3 Approach

This study applies document expansion and search results re-ranking with tagging information to improve search quality. The proposed document expansion approach only expands Web pages bookmarked in a social bookmark service provider, such as delicious.com. A bookmarked page p can be represented in the vector space model as $\{(t_1, m_1), (t_2, m_2), \ldots, (t_j, m_j)\}$, where t_j is a term appearing in page p, m_j is the frequency of t_j in page p. The tags used to tag page p are $T = \{(tg_1, n_1), (tg_2, n_2), \ldots, (tg_k, n_k)\}$, where tg_k is a tag tagging page p, and n_k is the tagging count of tg_k for p (i.e. n_k is the total number of users who use tg_k to tag page p). This study considers tags T as potential keywords to expand the original page p. The expanded page $p_{ex} = \{(t_1, m_1), (t_2, m_2), \ldots, (t_j, m_j), (tg_1, n_1), (tg_2, n_2), \ldots, (tg_k, n_k)\}$ is indexed for retrieval. The retrieval module scores the expanded page p_{ex} as $S_{pex,Q}$ for a query Q, and the scores of the original page p is $S_{p,Q}$.

The other method to improve search quality is reranking with social tagging information. If a tag t_i of page p is used as a term in a query Q, the result list would be re-ranked according to the tagging weight of t_i of page p, $W_{i,p}$, which is similar to the well-known tf-idf measure, as follows:

$$W_{i,p} = \left(n_{i,p} \Big/ \sum_k n_{k,p}\right) \times \ln\left(|P| \Big/ \left|\{p_j : t_i \in p_j\}\right|\right) \tag{1}$$

where $n_{i,p}$ is the frequency that t_i is used to tag page p, P is the set of all bookmarks in the social bookmark service provider, and p_j is a page tagged by t_i. If the query term t_i is not used as a tag for page p, $W_{i,p}$ is 0.

After $W_{i,p}$ is computed, $W_{i,p}$ is added to the score of page p, and then the result list is re-ranked based on the new score, $S_{rerank,p,Q}$.

$$S_{rerank,p,Q} = S_{p,Q} + \sum_i W_{i,p} \tag{2}$$

Document expansion adds tags as potential keywords into the original documents at indexing time, while the re-ranking technique attempts to improve the ranking of documents at query time. These two techniques are used at different phases of the retrieval process and they are complementary to each other. This study combines these two methods linearly as a hybrid approach, hoping to gain better improvements.

$$S'_{p,Q} = (1-\alpha)\times S_{pex,Q} + \alpha \times \sum_i W_{i,p} \qquad (3)$$

$S_{pex,Q}$ is the score of expanded pages computed by the retrieval module, $S'_{p,Q}$ is the new score, and α is the weighted parameter (where $0 \le \alpha \le 1$).

The PETS dataset contains the personal social tagging information, so this study utilizes this information to further improve the search quality. The personal social tagging information is also applied for document expansion and re-ranking. For document expansion by personal social tagging information, the expanded page p_{ex} is expanded again. The final expanded page $p_{ex,pt}$ is represented as $\{(t_1, m_1), (t_2, m_2), \ldots, (t_j, m_j), (tg_1, n_1), (tg_2, n_2), \ldots, (tg_k, n_k), (tgp_1, np_1), (tgp_2, np_2), \ldots, (tgp_l, np_l)\}$ in the vector space model, where tgp_i is a personal social tag, np_i is the tagging count, and $np_l = n_k$ if $tgp_l = tg_k$. For re-ranking by personal social tagging information, the formula is similar to eq. (3) and (4), but only personal tags are considered.

4 Dataset and Measure Metrics

4.1 MSN Searchlog 2006 Dataset

This study uses two datasets to evaluate the feasibility of the proposed approach. The first one is MSN Searchlog 2006 and the other one is the Web pages crawled using Yahoo! API. MSN Searchlog 2006 collects MSN query logs sampled over one month in 2006, and each log record contains both the query (Sessionid, Queryid, Query, Time, and resultCounts) and the clicks (Queryid, Query, Time, URL and the Position of the URL in the result list). Each Queryid represents a query in a query session, and result click-through logs are stored in the clicks data. MSN Searchlog 2006 collects about 7.5 million distinct Sessionids and 15 million Queryids (6.6 million distinct queries).

Because MSN Searchlog 2006 does not have the full result list of each query, using this dataset for experiments may not be reliable. Therefore, this study re-submitted queries of MSN Searchlog 2006 to Yahoo! API, and collected the top 100 result pages of each query. These collected Web pages and the clicked URLs of MSN Searchlog 2006 were merged as the second dataset in our experiments.

However, MSN Searchlog 2006 does not record the unclicked URLs of a query, and the result lists retrieved via Yahoo! API must be different from the original result lists of MSN search engine in 2006. Thus, it is impossible to get the completely original result list and relevance judgments for each query. Traditional measure metrics such as precision and recall could not be reliably estimated given the existing data set. Instead, nDCG (normalized discounted cumulative gain) [12] is used as evaluation measures in this study. nDCG is computed as:

$$DCG = rel_1 + \sum_{i=2}^{p} \frac{rel_i}{\log_2 i}$$
$$nDCG = \frac{DCG}{IDCG} \qquad (4)$$

where rel_i is the graded relevance of i-th result, and IDCG is the DCG value of sorted result list according to rel_i.

The social bookmarking and tagging information, including URL of a bookmarked page, tags of the bookmark and the count of each tag, are collected from delicious.com.

4.2 Personal Tagging and Searchlog Dataset

Due to the major shortcoming of MSN Searchlog 2006 – no full result list, Personal Tagging and Searchlog (PETS[1]) dataset was created in this study. PETS covered 14 subjects' search log (Apr. 2009 - Jul. 2009) collected from Google Web History[2]. Google Web History includes the queries submitted by the subjects, the clicked URLs of each query, and the time of each query and each click. These 14 subjects were not in the same domain background. PETS contained 10,191 queries and 11,603 click-through data. Besides, each query was re-submitted to Google search engine to get the full result list (top 100 results at most). As a result, PETS contains 348,843 URLs (278,248 distinct), and these Web pages were mostly in Chinese.

PETS collected the social bookmark information of these 14 subjects from delicious.com and hemidemi.com (a social bookmark service provider in Taiwan) as well. The social bookmark data covered the bookmarks of each subject and the tags (annotated by the subject and other users) of each bookmark.

5 Evaluation

5.1 MSN Searchlog 2006 Evaluation

Before evaluation, we manually checked whether each query session has exact one information need or not. If a query session contains more than one information need, the evaluation result might be affected. Two experts were asked to review 1031 randomly selected query sessions, and each session contained at least two queries. Only the sessions which were labeled as single information need by both experts were selected for evaluation.

Among the 1031 randomly selected query sessions in MSN Searchlog 2006, only 524 query sessions (about 51%) had single information need. Table 1 shows two example sessions with multiple information needs. This study randomly chose 56 sessions from the 524 sessions for experiments. These 56 query sessions contained 210 queries and 19,954 clicked Web pages. This study used the Lemur Indri module[3] [13], which is a language model retrieval approach, as the retrieval tool of this MSN Searchlog 2006 dataset evaluation.

[1] http://www.cs.nctu.edu.tw/~sychen/PerSearch_SocialTag_Dataset.html

[2] http://www.google.com/history/

[3] http://www.lemurproject.org/indri.php

Table 1 Example Query Sessions with Multiple Information Need in MSN Searchlog 2006

Session ID	Queries
000082fbfed64b62	• doberman pinscher • doberman pinscher health • doberman pinscher health concerns • google
000098be3a604395	• Stephen Colbert • patterson dental company • patterson dental • First Florida Accounting • First Florida Accounting services

Table 2. Average nDCG of queries for each method

	baseline	DE	DE_{log2}	DE_{log10}	R_{ln}
average nDCG	0.1496	0.1768	0.1706	0.1688	0.1736
improve (%)	-	18.2%	14.0%	12.9%	16.1%
# of improved queries	-	20	17	16	23

Table 3. Average nDCG of sessions for each method

	baseline	DE	DE_{log2}	DE_{log10}	R_{ln}
average nDCG	0.1610	0.1924	0.1829	0.1828	0.1892
improve (%)	-	19.6%	13.6%	13.6%	17.5%
# of improved sessions	-	19	17	16	16

The first experiment evaluates the nDCG value of each query, and the average nDCG of different approaches are reported in Table 2. The baseline is the results returned by Indri module without document expansion and reranking with tagging information. Compare to the baseline, the values of average nDCG are improved by 18.2% and 16.1% while document expansion (DE) and re-ranking (R_{ln}) are applied, respectively. The second experiment evaluates each query session with nDCG, and Table 3 shows the improvements of DE and R_{ln} (compare to the baseline) are 19.6% and 17.5% respectively.

One might worry about the situation that document expansion approach might expand too many tags into the original documents, and it may hurt the retrieval performance. Thus, this study conducted two alternative document expansion approaches, DE_{log2} and DE_{log10}. DE_{log2} expands the original page p with $(1+\lfloor \log_2 n_i \rfloor)$ times of tag tg_i, and DE_{log10} expands $(1+\lfloor \log_{10} n_i \rfloor)$ times, where n_i is tg_i's tagging count of page p. However, these two alternatives improve less than the DE approach in nDCG, and the amount of improved queries and sessions are also less than the DE approach. In other words, the DE approach does not excessively expand the original page, and does not hurt the retrieval performance.

Table 4. Average nDCG of queries for combining DE and reranking with different α value

	Average nDCG	Improve (%)
baseline	0.1496	
α=0.1	0.1906	27.42%
α=0.2	0.2022	35.16%
α=0.3	0.2081	39.16%
α=0.4	0.2165	44.75%
α=0.5	0.2164	44.68%
α=0.6	0.2144	43.35%
α=0.7	0.2088	39.61%
α=0.8	0.2104	40.66%
α=0.9	0.2018	34.91%

The third experiment evaluates the hybrid method, which combines document expansion (DE) and reranking with tagging information (R_{ln}) approaches with a weighted parameter α. We repeated this experiment with various α values to find out the optimal performance.

Table **4** shows that we can always improve the performance when α goes from 0.1 to 0.9, and the optimal nDCG is achieved when α is about 0.4.

Table 5 and Table 6 show the average nDCG of queries and sessions by applying this hybrid method and comparing with the query expansion (QE) approach. The QE approach feedbacks top 10 results to the original baseline query, and resubmits the query to get the new result list by the Indri retrieval module. The results reveal significant improvements over the baseline method. Besides, the ratio of improved queries and sessions are more than 25% and 45%, respectively. These improvements are affirmative evidences that social tagging can improve the quality of search.

However, surprisingly, the QE approach hurts the average nDCG by -2.1% for queries and only improves insignificantly (2.3%) for sessions. This is due to that the dataset is randomly chosen from the clicked URLs of MSN Searchlog 2006, which does not contain the complete result list of each query. As a result, the URLs used for expanding the original query are probably not very relevant to the original query. In view of this, this study uses the other dataset collected by Yahoo! Search API for experiments.

As described in subsection 4.1, the new Yahoo! dataset is created by resubmitting the 210 queries to Yahoo! Search API. The results (top 100 at most) of each query were collected as the new Yahoo! dataset. Yahoo! dataset also contains the clicked URLs of MSN Searchlog 2006, and it contains 13259 distinct URLs (about 20k pages in total), among which 1823 URLs are bookmarked.

For this Yahoo! dataset, this study also increases α from 0.1 to 0.9 to find out the optimal performance of the hybrid approach on average nDCG, and find out α=0.4 is the optimal setting. Thus, the hybrid approach with α=0.4 is compared with the query expansion (QE) again. The results are shown in Table 7 and Table 8. According to the results, our hybrid approach still provides statistically significant improvement (>10%) in the search quality.

Table 5. Average nDCG of queries by the hybrid method and compare with Query Expansion (QE)

	baseline	QE	DE + R_{ln}
average nDCG	0.1496	0.1464	0.2165
improve(%)	-	-2.1%	44.7%
# of improved queries	-	31	47

Table 6. Average nDCG of sessions by the hybrid method and compare with Query Expansion (QE)

	baseline	QE	DE + R_{ln}
Average nDCG	0.1610	0.1647	0.2469
improve(%)	-	2.3%	53.4%
# of improved sessions	-	23	25

Table 7. Average nDCG of queries (Yahoo! Dataset)

	Baseline	QE	DE+R_{ln}
Average nDCG	0.1601	0.1576	0.1838
improve(%)	-	-1.6%	14.8%

Table 8. Average nDCG of sessions (Yahoo! Dataset)

	Baseline	QE	DE+R_{ln}
Average nDCG	0.1745	0.1701	0.2050
improve(%)	-	-2.5%	17.5%

However, the query expansion (QE) approach still hurts the performance. To understand this, we checked some queries and found that some clicked URLs were homepages of websites. For example, http://www.bmw.com/ is the homepage of BMW inc. These homepages usually only contain a few text, but more images, hyperlinks and javascripts, which are not indexed by the Indri retrieval module. A commercial search engine would heavily emphasize the title of a page or promote a website's homepage with a better score, but a language model retrieval module, such as Indri, does not. Thus, query expansion benefits pages with more relevant terms and phrases, such as a page introducing BMW inc., but does not benefit the homepages of a website. This is interesting, as it suggests that query expansion may not be useful for navigational queries [14], although it is useful for informational queries [14].

Fortunately, homepages of high quality websites are usually bookmarked in social bookmarking services and annotated with many relevant tags; hence, the performance of the proposed hybrid approach is not hurt. For example, the homepage of BMW inc. (http://www.bmw.com/) is bookmarked 258 times in delicious.com, and the many relevant terms are used as tags. For example, "bmw", "cars" and "auto" are used 84, 76 and 37 times, respectively. Besides, pages related to BMW inc. are bookmarked less times and annotated with less tags than the homepage. These tags can be used to expand the homepage of BMW inc. and to promote its ranking more effectively than other pages related to BMW inc. That is why our proposed approach still works well in this new Yahoo! dataset.

5.2 PETS Dataset Evaluation

Same as the evaluation of MSN Searchlog 2006, we asked two experts to review the information need of each query in PETS dataset, and determine the candidate-sessions of queries in advance. After discarding candidate-sessions with only one queries and extracting the strict overlaps of candidate-sessions, 255 query sessions (764 queries) are selected.

This study randomly selects 23 sessions (80 queries) from these 255 sessions for evaluation. The result list of each query is reviewed by the two experts for relevance judgment. These 23 sessions cover 3677 results. Among them, 1278 results, which are labeled as relevant by both experts judgment, are used as the standard for evaluation.

Table 9. Average precision of sessions for each method

	baseline	QE	DE	DE_{log2}	DE_{log10}	R_{ln}	$DE+R_{ln}$	$(DE+R_{ln})_{pt}$
average precision	0.277351	0.2655	0.341275	0.341025	0.341025	0.277351	0.321025	0.341275
improve (%)	-	-4.27%	23.05%	22.96%	22.96%	0%	15.75%	23.05%

Table 10. Average nDCG of sessions for each method

	baseline	QE	DE	DE_{log2}	DE_{log10}	R_{ln}	$DE+R_{ln}$	$(DE+R_{ln})_{pt}$
average nDCG	0.591454	0.609586	0.659744	0.659688	0.659712	0.591459	0.657104	0.659751
improve (%)	-	3.07%	11.546%	11.536%	11.540%	~ +0%	11.10%	11.547%

Because the pages in the PETS dataset are mostly in Chinese, this study does not use Indri language model module, which is weak in processing Chinese data, as the retrieval tool for evaluation. Instead, Lemur[4] KL-divergence retrieval method was used. Indri language module is one retrieval method of Lemur toolkit, along with InQuery, Okapi, and KL-divergence.

Concerning the real world scenario, a user usually browses a few pages of Web search results, so this study only retrieves at most top 50 results of each query by Lemur KL-divergence retrieval method for evaluation. Besides, PETS contains the full result list, and relevance judgment is conducted in advance, so this evaluation conducts the precision measure. The precision evaluation results are shown in Table 9, and the nDCG evaluation results are shown in Table 10.

In Table 9, the baseline is retrieved by Lemur toolkit, with KL-divergence method. The DE approach improves 23.05%, and is better than DE_{log2} and DE_{log10}. The hybrid method, $DE+R_{ln}$, also improves the precision by 15.75%. However, the QE approach, which uses top 10 results as feedback, surprisingly hurts the performance. The main reason why the QE approach does not improve the search quality is similar to that mentioned in subsection 5.1.

Another interesting issue is that the re-ranking approach, R_{ln}, gains no improvements on precision, and it only improve slightly in nDCG (almost 0%). After checked the result list of baseline and R_{ln}, we found both are quite similar, except a few result pages have different ranks. We further checked the tags of these pages and the queries, and find that most queries (more than 2/3) are in Chinese, but only a few tags in PETS dataset are in Chinese (about

[4] http://www.lemurproject.org/

7.3%). The language discrepancy between queries and tags leads to insignificant improvement made by re-ranking.

The other interesting thing is the $(DE+R_{ln})_{pt}$ method, which additionally applies personal social tagging information to document expansion and re-ranking. However, in comparison with the DE approach, the precision is the same, and only improves insignificantly (<1%) on nDCG measurement. The reason is similar to the preceding paragraph; only a few tags given by those subjects in PETS are in Chinese (~8%).

6 Conclusion and future work

This study attempts to use the tagging information to improve Web search ranking and find a significant improvement based on search engine user logs, even the scale of social bookmarks is still relatively small. The proposed approach, which is based on document expansion and re-ranking with tagging information, is applied to improve the search quality. The real world search log data, MSN Searchlog 2006 and dataset collected by using Yahoo! API, are used in this work to verify the effectiveness of the proposed approach. Besides, for more thorough evaluation and the effect of personal social tagging, this study constructs an additional dataset, PETS, which contains a lot of real world search log, click through records, the full result lists of queries, and personal social tagging information.

However, the approach has some limitations; for example, URLs that are relevant but not stored in the social bookmark service providers cannot be benefitted. Even though, in MSN Searchlog 2006 data, document expansion adds tags into Web pages during the indexing time, and it gives 18.2% and 19.6% improvements for each query and session on average. Reranking uses tagging information to re-rank the documents after the results for a query are returned, and it gains 16.1% and 17.5% improvements for queries and sessions respectively. Combining document expansion and reranking, the improvements of average nDCG jump to 44.7% for queries and 53.4% for sessions. More than 25% of queries and 45% of query sessions are improved. Besides, our hybrid approach (combining document expansion and re-ranking) also provides significant improvements (>10%) on Yahoo! dataset. These experimental results are strong evidences to conclude that social tagging can improve search quality.

For the PETS dataset, the proposed document expansion approach still improves the precision by 23.05%, compared with the baseline. Unfortunately, due to the language difference between query terms and tags, the re-ranking approach and the personal social tagging information benefit nothing in precision and only have minor improvement in nDCG. To ameliorate this issue in the future, multi-language information retrieval is promising, or a new uni-language dataset might help to show the effect of personal social tagging information. Besides, retrieving the Google pagerank data of each URL for comparing with Google search results is the future challenge.

REFERENCES

1. Hansell, S.: Google Keeps Tweaking Its Search Engine. New York Times (2007)
2. Schenkel, R., Crecelius, T., Kacimi, M., Neumann, T., Parreira, J. X., Spaniol, M., Weikum, G.: Social Wisdom for Search and Recommendation. IEEE Data Engineering Bulletin, 31(2), 40-49 (2008)
3. Heymann, P., Koutrika, G., Garcia-Molina, H.: Can Social Bookmarking Improve Web Search?. In: First ACM International Conference on Web Search and Data Mining (2008)

4. Chen, S.-Y., Zhang, Y.: Improve Web Search Ranking with Social Tagging. In: 1st International Workshop on Mining Social Media, 13th Conference of the Spanish Association for Artificial Intelligence, 9th November, Sevilla, Spain (2009)
5. Krovetz, R.: Viewing morphology as an inference process. In Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 191-202 (1993)
6. Porter, M. F.: An algorithm for suffix stripping, Program, 14(3), 130-137 (1980)
7. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by latent semantic analysis. Journal of the American Society for Information Science, 41(6), 391-407 (1990)
8. Berger, A., Lafferty, J.: Information retrieval as statistical translation. In: Proceedings of the 22nd Annual Conference on Research and Development in Information Retrieval (ACM SIGIR). Berkeley, CA., pp. 222-229 (1999)
9. Billerbeck, B., Zobel, J.: Document Expansion versus Query Expansion for Ad-hoc Retrieval. In: Tenth Australasian Document Computing Symposium (2005)
10. Singhal, A., Pereira, F.: Document Expansion for Speech Retrieval. In: Proceeding of 22nd ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 34-41 (1999)
11. Xu, J., Croft, W. B. Query Expansion Using Local and Global Document Analysis. In: 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (1996)
12. Järvelin, K., Kekäläinen, J.: Cumulated gain-based evaluation of IR techniques. ACM Transactions on Information Systems 20(4), 422-446 (2002)
13. Strohman, T., Metzler, D., Turtle, H., Croft, W. B.: Indri: A language model-based search engine for complex queries. In: IA ’05: Proceedings of the International Conference on Intelligence Analysis (2005)
14. Broder, A.: A taxonomy of Web search. SIGIR Forum, 36(2), 3-10 (2002)
15. Bischoff, K, Firan, C.S., Nejdl, W., Paiu, R.: Can all tags be used for search?. In: Proceedings of the 17th ACM conference on Information and Knowledge Management (CIKM'08), 193-202, New York, NY, USA (2008)
16. Bao,S., Xue, G., Wu, X., Yu, Y., Fei, B., Su, Z.: Optimizing web search using social annotations. In: Proceedings of WWW 2007, 501-510 (2007)
17. Noll, M.G., Meinel, C.: Web search personalization via social bookmarking and tagging. In: Proceedings of ISWC 2007, 367-380 (2007)
18. Xu, S., Bao, S., Fei, B., Su, Z., Yu, Y.: Exploring folksonomy for personalized search. In: Proceedings of SIGIR 2008, 155-162 (2008)