Collaborative Annotation for Pseudo Relevance Feedback

Christina Lioma[1,2], Marie-Francine Moens[1], and Leif Azzopardi[2]

[1] Computing Science, Katholieke Universiteit Leuven, 3001, Belgium
[2] Computing Science, University of Glasgow, G12 8QQ, U.K.
{christina.lioma,sien.moens}@cs.kuleuven.be,
{xristina,leif}@dcs.gla.ac.uk

Abstract. We present a pseudo relevance feedback technique for information retrieval, which expands keyword queries with semantic annotation found in the freely available Del.icio.us collaborative tagging system. We hypothesise that collaborative tags represent semantic information that may render queries more informative, and hence enhance retrieval performance. Experiments with three different techniques of enriching queries with Del.icio.us tags, and also varying the number of tags used for expansion between 1-10, show small improvement in retrieval precision, over a baseline of short keyword queries.

1 Introduction

The field of Information Retrieval (IR) addresses the general problem of how to retrieve information, which is relevant to a user need, from a given repository of information, such as a document collection. A common example of IR systems is World Wide Web (Web) search engines, in which a short keyword query is used to generate a ranked list from a pre-indexed heterogeneous collection of documents. The matching between queries and documents is mostly term-based, i.e. the words within documents are used to describe the documents and to determine their relevance to a given query [25].

Often, the matching between queries and documents is enhanced by *relevance feedback*, which aims to render the initial query more informative, and resubmit it to the IR system, so that it can better match it to documents. There exist several ways for rendering queries more informative: (i) in *explicit relevance feedback* systems, users may expand their original query manually with potentially relevant terms suggested by the system [4, 6, 17, 21]; (ii) in *implicit relevance feedback* systems, logged user behaviour and/or search history can be used by the system to expand the original query automatically [2, 7, 8, 23]; (iii) in *pseudo relevance feedback* systems, term/document statistics can be used by the system to expand the original query automatically [16, 18]. Such term statistics can be extracted from documents already retrieved by the system (*local feedback*) [22], or from external sources of evidence, for instance Wikipedia entries [12] (*global feedback*) [11, 27].

In this paper we present a technique for expanding user queries with assumed relevant terms extracted from an external source of evidence, namely the Del.icio.us[3] collaborative annotation system. Del.icio.us is an online 'social tagging' system where users tag (= annotate), store and retrieve Web links. Given a query, Del.icio.us also suggests its most relevant tags. For example, given the query `holidays`, Del.icio.us suggests the related tags `travel, flights, calendar, hotels`[4]. We take advantage of this option, and expand a set of queries with their respective most related Del.icio.us tags. Our hypothesis is that such tags encode semantic information which may render the queries more informative and hence benefit retrieval performance.

We present three alternatives for selecting Del.icio.us tags: (i) on an individual term basis, (ii) on a phrase basis, (iii) on a whole query basis. Experimental evaluation of these techniques using the original (unexpanded) queries as baseline, on a standard Text REtrieval Conference (TREC[5]) dataset and with a robust model for matching documents to queries (Okapi's BM25 [15]) shows that our technique can improve retrieval precision for some but not all queries. This is a good starting point for further research into using collaborative annotation for IR. Given the free availability and increasing popularity (hence amount) of collaborative annotation, further research into incorporating this type of evidence in IR may be fruitful.

The remainder of this paper is organised as follows. Section 2 presents collaborative annotation systems and their use in IR. Section 3 presents our methodology for enriching queries with collaborative annotation. Section 4 presents and discusses our experiments. Section 5 summarises our findings and states intended future work.

2 Related studies

Broadly speaking, the underlying idea of semantic annotation is to identify interesting bits of metadata in documents (e.g. entities, relations, etc.). This type of annotation is becoming increasingly available online. For instance, the New York Times now uses rich headers metadata, while Reuters has launched the Open Calais[6] API for automatic semantic markup on HTML documents. Semantic annotation can be used in several ways to improve IR. For instance, knowledge of entities in text may be used to build sophisticated entity-based IR systems (sometimes referred to as *vertical search engines*). Another application is to automatically enrich textual content, for instance by inserting related links into raw text, as is done by the Inform[7] engine. Further applications include improving existing alert systems (e.g. RSS feeds), which are mostly based on keywords, and also incorporating on the fly text analysis into browsers. In brief,

[3] http://del.icio.us/
[4] Tags related to `holidays`, submitted to Del.icio.us on 14/02/2008.
[5] http://trec.nist.org/
[6] http://www.opencalais.com/
[7] http://www.inform.com/

semantic annotation is appealing because it can be seen as a way of enriching information (with more and structured data), which can result in improved processing; the question is whether this type of improved processing can result in improved system performance.

A type of semantic annotation is collaborative annotation, also known as *social tagging* or *distributed classification*, which refers to users creating and aggregating their own metadata. Collaborative annotation is a relatively new area (until recently largely absent from academic literature) but rapidly gaining ground on the Web.

The idea of asking users to annotate terms freely was initially developed by [5], who saw the process as a possible way of indexing particularly subjective forms of information where full-text searching was either not possible or not useful, such as multimedia or fiction objects. They developed the idea of aggregating users' indexing terms to create a generalised overall view of the resources, which today has been adapted by working systems, such as Del.icio.us, Flickr[8], a photo-sharing Web site where users upload, annotate and share photographs, CiteULike[9], a similar system but oriented towards scholarly writing and journal articles in particular, YouTube[10] and Last.fm[11], collaborative annotation services of multimedia resources (often user-authored). A potential disadvantage of human semantic annotation is inter-annotator disagreement or inconsistency, a result of allowing users to freely tag content. Early studies on human indexing also noted this as a problem [10, 19, 20].

The emergence of collaborative semantic annotation has stirred research in various directions, such as social issues surrounding tagging, growth and dynamics of social networks, cognitive processes behind tagging, and so on. The field of IR in particular has also shown interest in collaborative annotation: several commercial IR systems now include recommendation functionalities which are based on collaborative annotation, e.g. Amazon[12] uses collaborative annotation to suggest relevant products to online buyers. In addition, analogies between users - products in such recommender systems and queries - documents in IR systems are currently researched [13, 26].

3 Methodology for expanding queries with collaborative annotation

We present the steps taken in order to test the hypothesis that collaborative annotation includes semantic evidence, which can be used to enrich queries and hence enhance retrieval performance. Given a set of queries, for each query separately:

[8] http://www.flickr.com/
[9] http://www.citeulike.org/
[10] http://www.youtube.com/
[11] http://www.last.fm/
[12] http://www.amazon.com/

– **Step 1:** we submit it to Del.icio.us;
– **Step 2:** we use the θ most relevant tags returned by Del.icio.us to expand the original query;
– **Step 3:** we submit the expanded query to the IR system for retrieval.

We use three alternatives for **Step 1**:

– **term-based** alternative: we split the original query into individual terms, and submit each term separately to Del.icio.us. For example, original query=`foreign minorities, Germany`. Three separate queries submitted to Del.icio.us: (1) `foreign`, (2) `minorities`, (3) `Germany`.
– **phrase-based** alternative: we split the original query into phrases, and submit each phrase separately to Del.icio.us. For example, original query=`foreign minorities, Germany`. Two separate queries submitted to Del.icio.us : (1) `foreign minorities`, (2) `Germany`. We define phrases as comma-separated groups of terms.
– **query-based** alternative: we do not split the original query at all, but submit it as it is to Del.icio.us. For example, original query=`foreign minorities, Germany`=query submitted to Del.icio.us.

4 Evaluation

4.1 Experimental settings

The experimental aim is to test the hypothesis that expanding queries with collaborative semantic annotation can improve retrieval performance. We expect expanded queries to be more informative (which may increase early retrieval precision). The experimental setting is an IR system that matches documents to queries using an established retrieval model and unexpanded queries (baseline). To test our hypothesis, we expand queries with collaborative annotation from Del.icio.us, and compare retrieval performance to that of the baseline. We realise three rounds of experiments, one for each technique used in Step 1 to obtain related tags from Del.icio.us: (i) term-based, (ii) phrase-based, (iii) query-based.

We retrieve documents from the WT2G (2GB) collection, from the 1999, 2000 and 2001 Small Web tracks of the TREC Web Track (see Table 1), using topics 40-450 (see Table 4 in the Appendix). We use a Web dataset because it is more representative of real Web search. TREC queries usually contain a title, description, and narrative portion. The title contains few keywords; the description includes a brief description of the information need; the narrative contains a longer description of the information need. We experiment with short queries (title portion) only, because they are more representative of real user queries on the Web. We evaluate retrieval performance in terms of Precision at 10 (P10) and 20 (P20) retrieved documents.

We conduct experiments using the Terrier IR system [14]. Before retrieval, terms are tokenised on whitespace and punctuation marks, and lower-cased; stopwords are removed and terms are stemmed with the Porter stemmer. We

domain	size	#docs	#terms	#orig. query
Web	2GB	247,491	1,002,586	2.3

Table 1. The WT2G TREC dataset: domain = where it was crawled from; size = collection size; #docs = number of documents indexed; #terms = number of unique terms indexed; #orig. query = average length of original queries (unexpanded) measured in terms.

match documents to queries with the Okapi Best Match 25 (BM25) model [15]. BM25 includes certain parameters, which we set to default values. We use default values, instead of tuning these parameters, because our focus is to test our hypothesis, and not to optimise retrieval performance. If these parameters are optimised, retrieval performance may be further improved.

We expand the queries with θ terms suggested as relevant tags from Del.icio.us. We vary θ between 1-10, when possible (for some queries Del.icio.us offers < 10 relevant tags, see Table 5 in the Appendix). We treat Del.icio.us as a black box for suggesting relevant tags, i.e. we do not know how Del.icio.us estimates the relevance of the suggested tags, or whether these terms are ranked. We do not use queries numbered 407, 411, 414, 423, 427, 432, 438, 439, 440, 441, 442, 443, 449 because Del.icio.us does not suggest related tags for all of these query terms.

4.2 Experimental results

Figures 1 and 2 plot the number of collaborative tags used to expand queries (θ, x axis) against retrieval precision (y axis) for each of our three techniques used to select terms from Del.icio.us (term-based, phrase-based, query-based). We observe that overall, the term-based technique tends to perform worse, and that the phrase-based and query-based techniques perform approximately similarly, with the query-based technique giving slightly better results. This may be due to the compositional semantics of the whole query, which give a better representation of the user need than phrases or individual query terms. Figures 1 and 2 also show that precision seems to increase for lower θ values (= less expansion terms). In fact, the best performance is always associated either with $\theta = 0$, which corresponds to no query expansion (baseline), or with $\theta = 1$, which corresponds to the first relevant tag suggested by Del.icio.us (shown in Table 5 in the Appendix). Even though the estimation of relevance and the order of tag suggestion used by Del.icio.us is unknown to us, it seems that the most relevant tags come first, which may explain why $\theta = 1$ performs better.

Table 2 shows the P10 and P20 score of each query, separately for the original baseline queries (base) and for our three pseudo relevance feedback techniques (PRF_{term}, PRF_{phrase}, PRF_{query}), $\theta = 1$ (best θ, excluding the baseline). We observe that for about one third of the queries (13/36), the $\theta = 1$ term suggested by Del.icio.us is already a query term, e.g. term `Stirling` in query 447 (see Table 4). Out of the thirteen times that this happens, on one occasion P10 improves, and on two occasions P10 decreases, while for the remaining ten there is no change in performance. This is an interesting observation: the expansion

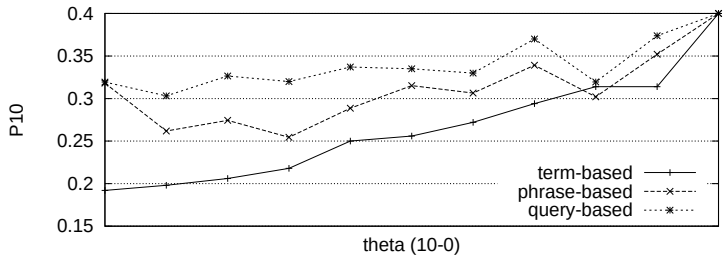

Fig. 1. Precision at 10 returned documents versus number of Del.icio.us terms used for query expansion. $\theta = 0$ is the baseline (no query expansion).

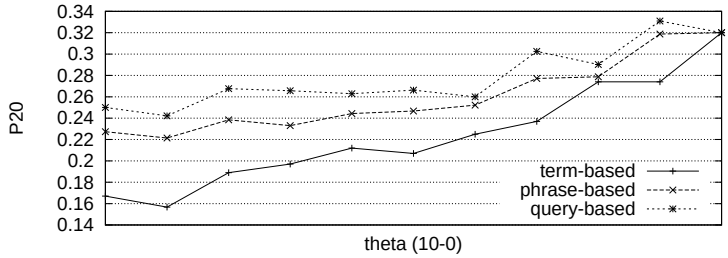

Fig. 2. Precision at 20 returned documents versus number of Del.icio.us terms used for query expansion. $\theta = 0$ is the baseline (no query expansion).

terms are providing redundant information, which explains why little change is observed overall in terms of performance. However, some interesting examples where performance did improve were on queries 404, 418, and 445, where the first Del.icio.us term seems helpful in emphasising the context of the query; e.g. in query 445 `women clergy`, the added term is `religion`, which is related to `clergy`; and in query 404 `Ireland peace talks`, the added term `activism` is again also related and on topic. While most of Del.icio.us terms seem to be related, not all are on topic. For instance, in query 404, Del.icio.us also suggests terms like `Iraq` and `Israel` where peace talks have been taking place, but are not on topic for the query. This suggests that the Del.icio.us tags might be better suited to aiding interaction, facilitating browsing or clustering data, instead of query expansion.

Table 3 compares baseline performance to the best performance marked by each of our three PRF techniques and also to a traditional PRF technique that expands queries with terms from the most relevant retrieved documents. For traditional PRF, we use the Bose Einstein 1 (Bo1) [1] term weighting model and add the 1-10 most relevant terms from the single top retrieved document. We observe that traditional PRF outperforms the baseline and our technique. This is expected, given that traditional PRF expands the query with 'local', 'weighted' relevant terms, while our technique expands the query with 'global', 'non-weighted' terms.

Results for best θ ($\theta = 1$)								
	Precision @ 10				Precision @ 20			
qid	base	PRF_{term}	PRF_{phrase}	PRF_{query}	base	PRF_{term}	PRF_{phrase}	PRF_{query}
401	0.30	0.20	**0.40**	0.30	0.40	0.20	0.25	0.20
402	0.50	**0.50**	**0.50**	**0.50**	0.50	**0.50**	**0.50**	**0.50**
403	0.90	**0.90**	**0.90**	**0.90**	0.90	**0.90**	**0.90**	**0.90**
404	0.50	**0.50**	0.40	**0.70**	0.55	0.50	0.30	**0.65**
405	0.00	**0.00**	**0.00**	**0.00**	0.05	0.00	**0.10**	**0.10**
406	0.20	0.00	0.00	0.00	0.20	0.05	0.15	0.15
408	0.60	**0.60**	**0.60**	**0.60**	0.35	**0.35**	**0.35**	**0.35**
409	0.30	0.20	0.20	**0.40**	0.25	**0.25**	**0.25**	**0.25**
410	0.80	0.70	0.50	0.50	0.75	0.65	**0.75**	**0.75**
412	0.70	**0.70**	**0.70**	**0.70**	0.70	**0.70**	**0.70**	**0.70**
413	0.00	**0.00**	**0.00**	**0.00**	0.10	0.00	0.00	0.00
415	0.40	0.30	**0.40**	**0.40**	0.35	0.20	**0.35**	**0.40**
416	0.60	0.50	0.30	0.30	0.50	**0.50**	**0.50**	**0.50**
417	0.40	**0.40**	**0.40**	**0.40**	0.30	**0.30**	**0.30**	**0.30**
418	0.40	0.60	**0.60**	**0.70**	0.30	**0.40**	**0.40**	**0.60**
419	0.20	0.10	**0.20**	**0.20**	0.20	**0.20**	0.15	**0.20**
420	0.60	**0.60**	**0.60**	**0.60**	0.35	**0.45**	**0.35**	**0.35**
421	0.60	0.10	0.20	0.20	0.35	0.20	0.20	0.20
422	0.00	**0.20**	**0.20**	**0.10**	0.10	**0.15**	**0.15**	0.05
424	0.20	**0.20**	**0.20**	**0.20**	0.20	**0.20**	**0.20**	**0.20**
425	0.40	0.20	**0.50**	**0.50**	0.35	**0.35**	**0.35**	**0.35**
426	0.10	0.00	0.00	**0.10**	0.05	**0.05**	**0.05**	**0.05**
428	0.10	**0.10**	**0.10**	**0.10**	0.05	**0.05**	**0.05**	**0.05**
429	0.40	**0.40**	0.30	0.30	0.30	0.25	**0.30**	**0.30**
431	0.20	**0.20**	0.10	0.10	0.25	**0.25**	**0.25**	**0.25**
433	1.00	0.40	0.40	0.60	0.55	0.25	0.25	0.55
434	1.00	0.90	0.90	**1.00**	0.65	0.60	0.60	**0.65**
435	0.10	**0.10**	**0.10**	**0.10**	0.05	**0.20**	**0.10**	**0.10**
436	0.00	**0.00**	**0.00**	**0.10**	0.10	**0.10**	0.05	0.05
437	0.10	**0.10**	**0.10**	**0.10**	0.05	**0.05**	**0.05**	**0.05**
444	1.00	**1.00**	**1.00**	**1.00**	1.00	**1.00**	**1.00**	**1.00**
445	0.00	**0.30**	**0.30**	**0.30**	0.15	**0.15**	**0.15**	**0.15**
446	0.40	**0.40**	**0.40**	**0.40**	0.30	**0.40**	**0.40**	**0.30**
447	1.00	0.80	0.80	0.80	0.60	0.55	0.55	0.55
448	0.00	**0.00**	**0.00**	**0.00**	0.00	**0.00**	**0.00**	**0.00**
450	0.40	0.00	**0.40**	**0.40**	0.30	0.10	**0.30**	**0.30**
Avg	0.40	0.34	0.35	0.38	0.32	0.31	**0.32**	**0.33**

Table 2. Precision at 10 and 20 relevant documents retrieved. Base = baseline (original queries). PRF_{term} = pseudo relevance feedback with Del.icio.us tags in response to individual terms. PRF_{phrase} = pseudo relevance feedback with Del.icio.us tags in response to phrases. PRF_{query} = pseudo relevance feedback with Del.icio.us tags in response to whole queries. Avg = average of all values. Bold = equal to or better than the baseline.

no PRF vs traditional PRF vs our PRF	P10	P20
baseline (no PRF)	0.40	0.30
traditional PRF (from relevant documents)	0.40	0.35
PRF_{term} (from Del.icio.us)	0.34	0.31
PRF_{phrase} (from Del.icio.us)	0.35	0.32
PRF_{query} (from Del.icio.us)	0.38	0.33

Table 3. Precision at 10 and 20 relevant documents retrieved. Baseline = original queries, no PRF. Traditional PRF = using the Bose Einstein 1 [1] term weighting model to expand queries with terms from the first relevant retrieved document. PRF_{term} = pseudo relevance feedback with Del.icio.us tags in response to individual terms. PRF_{phrase} = pseudo relevance feedback with Del.icio.us tags in response to phrases. PRF_{query} = pseudo relevance feedback with Del.icio.us tags in response to whole queries. For all PRF methods, we show values for the best number of expansion terms (θ between 1-10).

Overall, we observe that for most queries, our pseudo relevance feedback technique is either equal to or slightly better than the baseline. This seems to indicate that the contribution of the Del.icio.us semantic annotation is marginal. This may be due to our small dataset, or the techniques used for selecting terms for expansion without weighting them, but simply by considering them on a term- or phrase- basis. Perhaps more principled ways of selecting terms by weighting them, for instance by looking at their inverse document frequency in the collection, may benefit retrieval performance even more. The fact that there was no overall significant decrease of performance is encouraging, and indicates that this technique might be beneficial to retrieval on a selective basis, as has been shown with other forms of pseudo-relevance feedback [3].

5 Conclusion and future work

We presented a technique for pseudo relevance feedback, which expands queries with semantic annotation found in freely available collaborative tagging systems, and specifically Del.icio.us. We hypothesised that collaborative tags can represent semantic information that might be used to enrich queries, and hence enhance retrieval performance. We experimented with three different techniques of enriching queries with collaborative semantic annotation: (i) based on individual terms, (ii) based on phrases, and (iii) based on whole queries. We also experimented with the number of terms used for expansion, ranging it between 1-10. Out of the three techniques, the ones conveying context (phrase-based and query-based) behaved generally similarly; better performance was associated with the query-based technique and fewer expansion terms. Experiments with 36 Web queries showed no significant difference in retrieval performance between the original queries and the expanded queries. Some queries benefited from our technique, yet others did not; overall results are inconclusive. Collaborative semantic annotation seems to be broader than or quite general with respect to the user query, suggesting that perhaps better applications for it would be in

aiding user interaction, facilitating browsing and serendipitous search, or clustering documents, for instance. Further experimentation is needed in this direction, and particularly with regards to the selection of the most appropriate terms from the Del.icio.us related tags (e.g. by looking at their term statistics, or comparing their distribution in a general document collection to the distribution of query terms in the same collection, to identify discriminative terms).

In the future, we wish to experiment with larger datasets and more retrieval models (e.g. Inference Network Models [24] or Language Models [9], which allow for a straight-forward integration of evidence into the retrieval process and for weighting the effect of this integration), and with alternative ways of using collaborative semantic annotation to IR (e.g. to enrich documents, as opposed to queries only, a technique that might help to discriminate better between documents in a collection, and hence enhance retrieval performance).

References

1. Amati, G.: Probabilistic Models for Information Retrieval Based on Divergence from Randomness. PhD thesis, University of Glasgow, 2003.
2. Anick, S., P.: Using Terminological Feedback for Web Search Refinement: a Log-Based Study. In: ACM Conference on Research and Development in Information Retrieval (SIGIR 2003) 88-95.
3. Cronen-Townsend, P., Zhou, Y., Croft, B.:: A Framework for Selective Query Expansion. In: ACM Conference on Information and Knowledge Management (CIKM 2004) 236-237.
4. Harman, D.: Towards Interactive Query Expansion. In: ACM Conference on Research and Development in Information Retrieval (SIGIR 1988) 321-331.
5. Hidderley, R., Rafferty, P.: Democrating Indexing: an Approach to the Retrieval of Fiction. Information Services & Use, 17(23). (1997) 101-109
6. Koenemann, J., Belkin, N.: A Case for Interaction: a Study of Interactive Information Retrieval and Effectiveness. In: Conference on Human Factors in Computing Systems (SIGCHI 1996) 205-212.
7. Komlodi, A., Soergel, D., Marchionini, G.: Search Histories for User Support in User Interfaces. JASIS, 57(6). (2006) 803-807
8. Lau, T, Horvitz, E.: Patterns of Search: Analyzing and Modeling Web Query Refinement. Proceedings of the 7th International Conference on User Modelling. (1999) 119-128.
9. Lavrenko, V., Croft, B.: Relevance-Based Language Models. In: ACM Conference on Research and Development in Information Retrieval (SIGIR 1999) 120-127
10. Lewis, D. D., Sparck Jones, K.: Natural Language Processing for Information Retrieval. In: Communcations of the ACM, (1996) 39(1):92-101.
11. Mandala, R., Tokunaga, T., Tanaka, H.: Combining Multiple Evidence from Different Types of Thesaurus for Query Expansion. In: ACM Conference on Research and Development in Information Retrieval (SIGIR 1999) 191-197
12. Milne, D., Witten, I. H., Nichols, D. M.: A Knowledge-Based Search Engine Powered by Wikipedia. In: CIKM, (2007) 445-454.
13. Mishne, G.: AutoTag: a Collaborative Approach to Automated Tag Assignment for Weblog Posts. In: Proceedings of the 15th international conference on World Wide Web, (2006) 253-254.

14. Ounis, I., Lioma, C., Macdonald, C., Plachouras, V.: Research Directions in Terrier: A Search Engine for Advanced Retrieval on the Web. In: Ricardo Baeza-Yates et al. (Eds.) Novatica/UPGRADE Special Issue on Web Information Access, invited paper, 2007.
15. Robertson, S., Walker, S.: Some Simple Approximations to the 2-Poisson Model for Probabilistic Weighted Retrieval. In: ACM Conference on Research and Development in Information Retrieval (SIGIR 1994) 232-241
16. Rocchio, J.: Relevance Feedback in Information Retrieval. In: The SMART Retrieval System, (1971) 313-323.
17. Ruthven, I.: Human Interaction: re-examining the potential effectiveness of interactive query expansion. In: ACM Conference on Research and Development in Information Retrieval (SIGIR 2003) 213-220.
18. Salton, G, Buckley, C.: Improving Retrieval Performance by Relevance Feedback. In: JASIS (1990), 41:288-297.
19. Salton, G, McGill, M. J.: Introduction to Modern Information Retrieval. MacGraw-Hill, New York (1983).
20. Soergel, D.: Indexing and Retrieval Performance: The Logical Evidence. In: JASIS (1994), 45(*):588-599.
21. Spink, A.: Term relevance feedback and query expansion: relation to design. In: ACM Conference on Research and Development in Information Retrieval (SIGIR 1994) 81-90.
22. Tan, B., Atulya, V., Fang, H., Zhai, C. X.: Term Feedback for Information Retrieval with Language Models. In: ACM Conference on Research and Development in Information Retrieval (SIGIR 2007) 263-270.
23. Teevan, J., Adar, E., Jones, R., Potts, M. A. S.: Information Re-Retrieval: Repeat Queries in Yahoo's Logs. In: ACM Conference on Research and Development in Information Retrieval (SIGIR 2007) 151-158.
24. Turtle, H. R., Croft, W. B.: Inference Networks for Document Retrieval. In: ACM Conference on Research and Development in Information Retrieval (SIGIR 1990) 1-24.
25. van Rijsbergen, C., J.: Information Retrieval. Butterworths, London (1979)
26. Wang, J., De Vries, A., Reinders, M. j. T.: Unified Relevance Models for Rating Prediction in Collaborative Filtering. In: ACM Transactions on Information Systems, in press.
27. Xu, J., Croft, B.: Query Expansion using Local and Global Document Analysis. In: ACM Conference on Research and Development in Information Retrieval (SIGIR 1996) 4-11.

qno	TREC query title	TREC query description
401	foreign minorities, Germany	**language, cultural**, differences, impede, integration, foreign, minorities, Germany
402	behavioral genetics	happening, field, **behavioral, genetics**, study, relative, influence, **genetic** environmental, factors, individual's, **behavior**, personality
403	osteoporosis	find, information, effects, dietary, intakes, potassium. magnesium, fruits, vegetables, determinants bone, mineral, density, elderly, men, women, preventing, **osteoporosis**, bone, decay
404	Ireland, peace talks	often, peace, talks, ireland, delayed, disrupted, result, acts, violence
405	cosmic events	unexpected, unexplained, cosmic, events, celestial, phenomena, radiation, supernova, outbursts, new, comets, detected
406	Parkinson's disease	being, done, treat, symptoms, **parkinson's, disease**, keep, patient, functional, long, possible
408	tropical storms	**tropical, storms, hurricanes**, typhoons, caused, significant, property, damage, loss, life
409	legal, Pan Am, 103	legal, actions, resulted, destruction, pan am, flight, 103, lockerbie, scotland, december 21 1988
410	Schengen agreement	involved, **schengen**, agreement, eliminate, border, controls, western, europe, hope, accomplish
412	airport security	**security**, measures, effect, proposed, go, effect, **airports**
413	steel production	new, methods, producing, **steel**
415	drugs, Golden Triangle	**drugs**, known, trafficking, golden, triangle, area, burna, thailand, laos, meet
416	Three Gorges Project	status, **three, gorges**, project
417	creativity	find, ways, measuring, **creativity**
418	quilts, income	ways, quilts, used, generate, income
419	recycle, automobile tires	new, uses, developed, old, automobile, tires, means, tire, recycling
420	carbon monoxide poisoning	widespread, **carbon**, monoxide, global, scale
421	industrial waste disposal	disposal, industrial, **waste**, being, accomplished, industrial, management, world
422	art, stolen, forged	incidents, stolen, forged, art
424	suicides	give, examples, alleged, **suicides**, aroused, suspicion, **death**, actually, being, murder
425	counterfeiting money	counterfeiting, **money**, being, done, modern, times
426	law enforcement, dogs	provide, information, use, **dogs**, worldwide, **law**, enforcement, purposes
428	declining birth rates	countries, U.S., china, declining, birth, rate
429	Legionnaires' disease	identify, outbreaks, legionnaires', disease
431	robotic technology	latest, developments, **robotic, technology**
433	Greek, philosophy, stoicism	contemporary, interest, **greek, philosophy, stoicism**
434	Estonia, economy	state, **economy, estonia**
435	curbing population growth	measures, taken, worldwide, countries, effective, curbing, **population**, growth
436	railway accidents	causes, railway, **accidents**, world
437	deregulation, gas, electric	experience, residential, utility, customers, following, deregulation, gas, electric
444	supercritical fluids	potential, uses, **supercritical, fluids**, environmental, protection, measure
445	women clergy	countries, United, states, considering, approved, **women, clergy**, persons
446	tourists, violence	tourists, likely, subjected, acts, **violence**, causing, bodily, harm, death
447	Stirling engine	new, developments, applications, **stirling, engine**
448	ship losses	identify, instances, weather, main, contributing, factor, loss, ship, sea
450	King Hussein, peace	significant, figure, years, late, jordanian, king, hussein, furthering, peace, middle, east

Table 4. Original queries used in the experiments; qno = query number. Titles and descriptions as provided by TREC. Bold = description terms also suggested by Del.icio.us as relevant (see Table 5).

qno	original query	expansion terms (most common del.icio.us annotation)
401	foreign minorities, Germany	culture, language, languages, linguistics, reference
402	behavioral genetics	genetics, science, psychology, evolution, research, biology, behavior, sociology, economics, philosophy
403	osteoporosis	osteoporosis, health, nutrition, medical, food, medicine, calcium, arthritis, fitness
404	Ireland, peace talks	activism, audio, campaign, charity, environment, eu, green, Iraq, Israel, media
405	cosmic events	science, astronomy, space, news, interesting, physics, article, daily, cool, future
406	Parkinson's disease	health, parkinsons, science, parkinson's, brain, research, disease, medicine, parkinson, politics
408	tropical storms	weather, hurricane, hurricanes, news, maps, science, storm, reference, tropical, noaa
409	legal, Pan Am, 103	reference, dictionary, google, language, map, maps, thesaurus, travel, visualization
410	Schengen agreement	schengen, eu, politics, international, travel, visa, wiki, wikipedia
412	airport security	security, travel, airport, politics, terrorism, wifi, mac, tsa, wireless, osx
413	steel production	design, business, art, steel, diy, reference, technology, tools, engineering, hardware
415	drugs, Golden Triangle	drugs, ajax, asia, homepage, news, police, politics, portal, rss, strange
416	Three Gorges Project	china, environment, energy, dam, 3gorges, bbc, gorges, news, photos, three
417	creativity	creativity, design, inspiration, productivity, art, blog, lifehacks, innovation, writing, business
418	quilts, income	art, crafts, design, handmade, shopping
419	recycle, automobile tires	architecture, cool, destruction, environment, fun, green, shredding, sustainability, sustainable, video
420	carbon monoxide poisoning	suicide, carbon, crossover, health, poisoning
421	industrial waste disposal	recycling, recycle, waste, market, plastic, environment, management, scrap, photography, art
422	art, stolen, forged	audio, free, manifesto, music, opensource, pandora, radio, software, technology, web2.0
424	suicides	suicide, funny, humor, comics, death, war, bunny, politics, iraq, news
425	counterfeiting money	money, security, politics, privacy, counterfeit, currency, crime, economics, economy, printer
426	law enforcement, dogs	law, police, accessibility, activism, ada, censorship, disability, dog, doghouse, dogs
428	declining birth rates	articles, parenting
429	Legionnaires' disease	uk
431	robotic technology	technology, robotics, robots, robot, science, art, video, design, diy, electronics
433	Greek, philosophy, stoicism	philosophy, stoicism, classics, religion, books, epictetus, ethics, greek, history, jesus
434	Estonia, economy	estonia, politics, economics, statistics, economy, culture, europe, freedom, tax, bots
435	curbing population growth	development, environment, health, population, poverty
436	railway accidents	accident, activism, alcohol, article, design, disasters, madd, politics, risk, technology
437	deregulation, gas, electric	energy, engineering, engineers, organization, organizations
444	supercritical fluids	chemistry, fluids, nature, news, physics, science, supercritical, water
445	women clergy	religion, islam, feminism, politics, christianity, philosophy, women
446	tourists, violence	blog, china, crime, culture, drugs, egypt, humor, literature, police, violence
447	Stirling engine	stirling, engine, energy, science, solar, diy, power, technology, environment, howto
448	ship losses	music, riaa, business, apple, businessmodel, filesharing, future, hardware, lies, mac
450	King Hussein, peace	animation, bush, engine, flash, funny, google, israel, politics, search, searchengine

Table 5. Queries used in the experiments and their respective relevant terms in the order suggested by Del.ico.us and used in our query-based technique; qno = query number.