External Query Reformulation for Text-based Image Retrieval

Jinming Min and Gareth J. F. Jones

Centre for Next Generation Localisation
School of Computing, Dublin City University
Dublin 9, Ireland
{jmin,gjones}@computing.dcu.ie

Abstract. In text-based image retrieval, the Incomplete Annotation Problem (IAP) can greatly degrade retrieval effectiveness. A standard method used to address this problem is pseudo relevance feedback (PRF) which updates user queries by adding feedback terms selected automatically from top ranked documents in a prior retrieval run. PRF assumes that the target collection provides enough feedback information to select effective expansion terms. This is often not the case in image retrieval since images often only have short metadata annotations leading to the IAP. Our work proposes the use of an external knowledge resource (Wikipedia) in the process of refining user queries. In our method, Wikipedia documents strongly related to the terms in user query ("definition documents") are first identified by title matching between the query and titles of Wikipedia articles. These definition documents are used as indicators to re-weight the feedback documents from an initial search run on a Wikipedia abstract collection using the Jaccard coefficient. The new weights of the feedback documents are combined with the scores rated by different indicators. Query-expansion terms are then selected based on these new weights for the feedback documents. Our method is evaluated on the ImageCLEF WikipediaMM image retrieval task using text-based retrieval on the document metadata fields. The results show significant improvement compared to standard PRF methods.

1 Introduction

The volume of online images has been expanding at an increasing rate in recent years. Searching for interesting and useful images from among the enormous number of images available generally relies either on content-based image retrieval using visual image features or text based search using text queries to search for images based on textual annotations of the images. Often it is difficult to find a sample image to use as a query image for visual search, and thus the text-based method is often the most commonly used by search engine vendors such as *Google*, *Bing* and *Yahoo!*. Where high quality detailed annotations are available, the text-based method can be very effective. However, annotations

Image News	Text News
 File image of Bill Clinton. Image: MSGT Renee Humble.	Former President Bill Clinton and New York City Mayor Michael Bloomberg have revealed plans to merge their climate leadership groups C40 and the Clinton Climate Initiative on Wednesday. Bloomberg serves as the chair for C40 Cities, an organization of cities devoted to tackling climate change. Together with the Clinton Climate Initiative, a program of the William J. Clinton Foundation, the two high profile environmentalists believe the alliance will assist cities to reduce greenhouse gas emissions through a range of energy efficient and clean energy programs. "I am pleased and honored to be here with my friend President Clinton. He has been a life-long champion of climate initiatives and I can think of no greater partner in this effort," said Bloomberg. ......

Fig. 1. Incomplete annotation problem example.

are unfortunately often found to be noisy or incomplete, e.g. Picasa[1], Flickr[2]. The annotations are generally provided by those contributing the images who often only provide very brief or sometimes inaccurate details. These issues of poor image annotation can greatly affect image retrieval effectiveness based on textual metadata. Without complete textual description of an image, it is difficult to reliably match the image with text queries, since relevant images may not contain useful annotation terms. Thus is is not possible for the retrieval system to return the relevant images with high accuracy. We refer to this effect as the *incomplete annotation problem (IAP)* in image retrieval. An example of this problem is shown on the left of Figure 1 which is an image example from Wikinews [3]. Compared to the text version of same content, there are many fewer terms used to describe content of the image in the annotation.

In ad-hoc information retrieval (IR) tasks, a popular method to address the more general problem of query-document mismatch is query expansion (QE). This seeks to add terms to the user query which will match with terms appearing in relevant documents. Standard QE methods can also be applied to improve retrieval effectiveness in image retrieval. In our research, we aim to address the question: is standard QE the most suitable method to address the IAP problem in text-based image retrieval? Classical QE methods greatly depend on the target collection to provide useful terms for QE. The IAP problem means that the assumption that the target collection provides enough information for feedback in image retrieval may often be violated. We propose that an effective solution to IAP for image search could be the introduction of an external resource in the feedback process. Furthermore, since the search query in image retrieval is usually a noun phrase for which there is a high chance that Wikipedia contains specific articles to describe it, Wikipedia is a suitable external source for QE in

[1] http://picasaweb.google.com/
[2] http://www.flickr.com/
[3] http://en.wikinews.org/

image search. In our work we refer to Wikipedia articles which directly describe the contents of a user query as *definition documents (DDs)*. Based on this analysis, we propose a definition document based relevance feedback (DRF) method for text-based image retrieval task.

The remainder of this paper is structured as follows: Section 2 overviews background and related work to our investigation; Section 3 presents our DRF method including identifying DDs using query key terms, feedback document weighting and feedback term selection; Section 4 describes our experimental setup and results, this compares standard PRF methods based on the target annotated search collection, external PRF which conducts QE on the external resource collection, feedback term selection from DDs only and our DRF method; and finally Section 5 gives conclusions and directions for further work.

2 Background and Related Work

The IAP problem in text-based image retrieval task is typically addressed by relevance feedback (RF) approach [1]. One standard approach to RF is QE where terms from top ranked documents from an initial search are added to the original query before performing another search run. A popular method of RF is pseudo relevance feedback (PRF) where top ranked documents are assumed to be relevant without being judged by the searcher. PRF via QE traditionally focuses on selecting expansion terms from top ranked documents from an initial retrieval on the target document collection. In recent research however, with the rapid growth of the web and other electronic document resources, QE from external resources has received increased attention. This approach, the initial retrieval run is carried out on an external corpus, and feedback terms are then selected from the top ranked documents in external corpus. The new expanded query is applied to the target corpus to conduct the final retrieval run. These research topics are strongly related to our query reformulation method for image retrieval tasks.

Various techniques and resources have been investigated for RF using external resources in existing work. Elsas et al. [2] utilize the link structure of Wikipedia for QE in a Blog distillation task and yield significant improvement for retrieval effectiveness, this work also showed that standard PRF does not perform well in a Blog distillation task. Yang et al. [3] classify user queries into three types: entity queries, ambiguous queries, and broader queries. In this work, for entity queries expansion terms are selected only from an entity page in Wikipedia. Their experiments show improvement on several TREC evaluation tasks. Yin et al. [4] compare two QE methods from an external resource, one selects QE terms from user search query logs, the other method selects feedback terms from snippets gathered from search engine output results. Their results show that the snippet approach was more effective. Kwok et al. [5] use a technique of collection enrichment for QE which is essentially QE from an external resource. Their system performs between 9% to 26% better than the initial retrieval as measured using mean average precision (MAP) as reference. Xu et

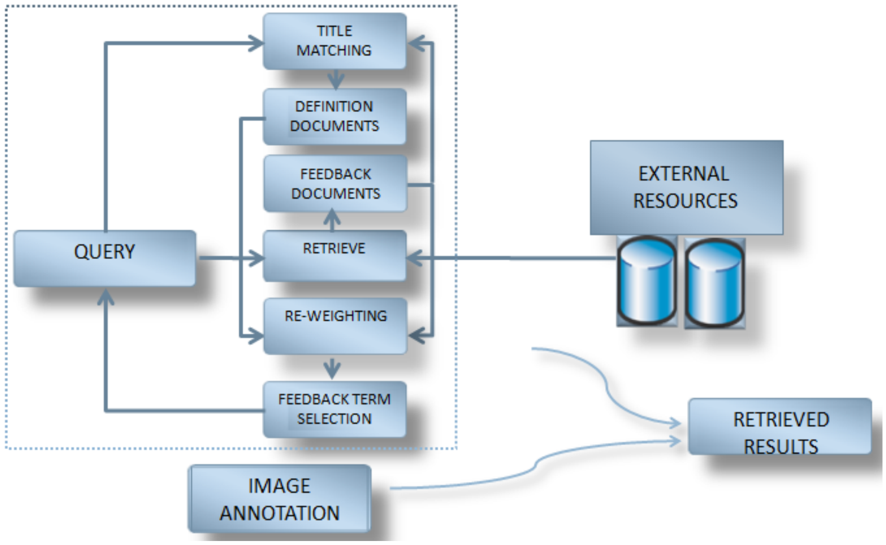

Fig. 2. Flowchart of the algorithm.

al. [6] identify an entity page and reformulate the query with phrases from the entity page in Wikipedia. Their results show improvement in several TREC evaluation tasks compared to a language modelling IR baseline. Weerkamp et al. [7] explore different ways of using external corpora to expand the original query in a Blog post retrieval task. They achieve their best results when using external expansion on a combination of news, Wikipedia and blog posts. Custis et al. [8] apply language modelling keyword search augmented with Berger and Lafferty's (1999) translation model for QE to formulate three QE methods using word co-occurrence statistics from a large external corpus and user clickthrough data. Results show that QE using the translation model is effective for retrieval in the legal domain. Weerkamp et al. [9] propose a generative model for expanding queries using external collections in which dependencies between queries, documents, and expansion documents are explicitly modeled. Results using two external collections (news and Wikipedia) show texternal expansion for retrieval of user generated content to be effective. Hersh et al. [10] expand the query from web pages online in a genomic IR task. Our own previous work [11] reports initial experiments on QE from Wikipedia for a text-based image retrieval tasks, and shows improvement compared with the QE from the target corpus. We extend this earlier work in this paper.

3 Definition Documents based Relevance Feedback

In this section, we introduce our definition documents (DDs) based RF method. This method utilises DDs identified by key-term title matching to re-weight the feedback documents. Our document-based relevance feedback (DRF) algorithms consist of the following steps as shown in Figure 2:

1. The user query is applied to Wikipedia to conduct an initial retrieval run to produce a ranked list;
2. The user query is applied to the top ranked documents retrieved in stage 1 to conduct key-term title matching to find the DDs for this query;
3. The DDs identified in the second stage are used as indicators to compute the similarity score with the top k ranked documents from the initial retrieval run in stage 1 by the Jaccard coefficient;
4. The similarity score from DDs is used to form a new weight for every feedback document;
5. Feedback terms are selected from the feedback documents from stage 2 with new associated weights;
6. The new query updated with feedback terms is applied to the target search collection to carry out the final retrieval run.

In the following subsections, we introduce the standard PRF method used in our work in subsection 3.1, then the key-term title matching method is described in subsection 3.2, and the feedback term weighting method is addressed in subsection 3.3.

3.1 Pseudo Relevance Feedback

PRF is the standard method used for QE. It has been found to improve average search effectiveness in many ad-hoc text search tasks. Equation 2 is a typical method for selection of feedback terms in PRF [12].

$$RW(i) = log[\frac{(r+0.5)(N-n-R+r+0.5)}{(n-r+0.5))(R-r+0.5)}] \tag{1}$$

$$Weight_{PRF}(t) = r * RW(i) \tag{2}$$

where r is the number of top-ranked feedback documents which contain the term t and $RW(i)$ is computed using Equation 1 where N is the total number of documents in the corpus and n is the number of documents where the term t appears. R is the number of known relevant documents for a query. Another simple version for PRF is Equation 3 where *idf* can be computed using Equation 4.

$$Weight_{PRF}(t) = r * idf(t) \tag{3}$$

$$idf(t) = \log \frac{N}{n} \tag{4}$$

3.2 Identifying Definition Documents by Key-term Title Matching

In this section, we address the problem of how to find the DDs for a query. For a query, some documents very strongly related to the query can be found in Wikipedia, we can refer to these documents as "relevant" to the query in the sense that they essentially describe one or more concepts contained in the query. For example, given a query "Ferrari" a Wikipedia DD will appear among the top ranked list after a prior retrieval run. Figure 3 illustrates an example user query with a DD in Wikipedia. Since exact DDs may not be found for all queries, our DRF method allows a more relaxed matching approach in these cases. We use a partial matching approach with regard to Wikipedia documents in top ranked documents from the prior retrieval whose title contains the key term of the user query as the DD of this query.

> *Ferrari is an Italian sports car manufacturer based in Maranello, Italy. Founded by Enzo Ferrari in 1929, as Scuderia Ferrari, the company sponsored drivers and manufactured race cars before moving into production of street-legal vehicles as Ferrari S.p.A. in 1947. Throughout its history, the company has been noted for its continued participation in racing, especially in Formula One, where it has had great success.*

Fig. 3. Definition Document Example.

Given a query Q: $\{q_1,q_2,...,q_m\}$ and a document D with title T: $\{t_1, t_2, ..., t_n\}$, the key term q_k of the Q is the term with highest *idf* score given by Equation 4. Document D whose title contains q_k is called the DD of query Q. There may be more than one DD for a given query in Wikipedia. We use the *idf* values of the terms in the target collection to identify the key term in the query.

3.3 Feedback Term Weighting

Our RF method is built on the simple version of PRF in Equation 3. In Equation 3, PRF assigns all the top documents from the prior retrieval the same importance, which is usually not actually true. This is built on the assumption that the top k feedback documents in the prior retrieval are all relevant to the query. Of course, this assumption is generally not true for most retrieval tasks.

When some DDs have already been identified by title matching, we assume those feedback documents which are similar to the DDs have a higher probability of providing useful external knowledge to the query. The similarity of the feedback documents and the DDs is computed using a pairwise method using the Jaccard coefficient in Equation 5 where V_i, V_j is the vocabulary set of document i and j [13].

$$sim(D_i, D_j) = \frac{V_i \bigcap V_j}{V_i \bigcup V_j} \tag{5}$$

By knowing which document is more useful to the query, new document weights are assigned to the feedback documents using Equation 9. DDs are used as indicators for the top-ranked feedback documents in the prior retrieval as shown in Equation 7. For each feedback document, we have an initial retrieval score for query S_i. We normalize the scores into the range [0, 1] using Equation 6.

$$S_{nm}(i) = \frac{S_i - S_{min}}{S_{max} - S_{min}} \tag{6}$$

In Equation 7, J is the set of all DDs and $j \in J$; $sim(D_i, D_j)$ is the similarity score of document i for DD j calculated using Equation 5; $S_{nm}(j)$ is the normalized retrieval score of DD j; $sim(J)_{avg}$ is the average similarity score for all definition documents J with all feedback documents;

$$G(i) = \frac{\sum_J (sim(D_i, D_j) - sim(j)_{avg}) S_{nm}(J)}{\sum_J S_{nm}(j)} \tag{7}$$

Before combining these scores into the final weight score for a feedback document, $G(i)$ is normalized into the range [0, 1] using Equation 8 where G_{max} and G_{min} are the highest and lowest $G(i)$ scores calculated using Equation 7 for all feedback documents.

$$G_{nm}(i) = \frac{G_i - G_{min}}{G_{max} - G_{min}} \tag{8}$$

$$W_{new}(D_i) = \alpha * \overline{S_{nm}} + \beta * G_{nm}(i) \tag{9}$$

where $\overline{S_{nm}}$ is the average normalized retrieval score for all the feedback documents. α and β are parameters to adjust the rating system ($\alpha \geq 0$, $\beta \geq 0$). In Equation 9, the new weight of a FD is combined from two parts: one is the average normalized retrieval score which is identical for every FD; the other is from the rating scores of different DDs. If we set $\beta = 0$ and $\alpha \neq 0$, our method automatically falls into the simple version of PRF method; if we set $\beta \neq 0$ and $\alpha = 0$, the weights of FDs are all decided by the rating scores of the DDs.

With the new weights for the all feedback documents in the prior retrieval, the top feedback terms are selected using Equation 10 where r is the set of feedback documents which contain term t.

$$Weight_{DRF}(t) = idf(t) \cdot \sum_{D_i \in r} W_{new}(D_i) \tag{10}$$

4 Experimental Setup and Results

In this section, we describe our experimental setup and results. The data and retrieval model used in experiments are described in subsection 4.1, while the manual evaluation of precision of DDs is described in subsection 4.2. The effect of parameter setting on DRF and PRF is presented in subsection 4.3, with results of comparing DRF and PRF in subsection 4.4, and a comparison with our method of selecting feedback terms only from DDs in Section 4.5.

Table 1. Data Average Length.

Data	Average Length (by terms)
Topics	2.8
Annotation Documents	24.4
English DBpedia Documents	99.7

Table 2. Overview on the definition documents.

No. of topics	120
No. of overall definition documents	421
Average No. of DDs per topic	3.5
DDs with total match	77
Topics with total match DDs	46

4.1 Experimental Setup

In this section, we describe our experimental setup. Experiments were conducted using the collection from the ImageCLEF WikipediaMM 2008 and 2009 tasks. The corpus is taken from the (INEX MM) Wikipedia image collection and includes 151,519 images [14]. Every image is associated with a metadata file. These metadata documents are typically very short, meaning that there is a high chance of IAP problems in this collection.

All our experimental results are based on the 120 official queries in this collection. Another important resource we use is the English Wikipedia abstract collection (DBpedia) including 2,452,726 documents which is used as the external resource for QE. We chose the English DBpedia collection as the external resource for QE in this study since: 1) the DBpedia dataset contains only the abstract documents of Wikipedia terms and so contains less noise than full articles; 2) the DBpedia corpus covers many topics which holds the promise that we can find relevant documents for a large number of queries. The average length of data are shown in Table 1. We use the Okapi BM25 model in the Lemur toolkit[4] for retrieval tasks.

4.2 Evaluation on Definition Documents

Our DDs are selected by key term matching from document titles. A further question in this process is how good are DDs as indicators for feedback? We manually evaluated the DDs for the official queries from WikipediMM tasks. These DDs are selected from the top 30 Wikipedia documents in the prior retrieval run. In Table 2, DDs with "total match" means those DD's with titles which exactly match the query terms after removal of stop words. We also manually evaluate the relevance of the DDs with the original topics. The results are shown in Table 3. As shown in Table 3, all the total match DDs and most partial match DDs are relevant to the original topics. The results indicate that DDs are a good feedback source for text queries in image retrieval.

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

Table 3. Evaluation on the definition documents.

	Relevant	Non-relevant
total match definition documents	100%	0
partial match definition documents	85.5%	14.5%

Table 4. Parameters Choice for DRF Method.

Parameters Setting	MAP	NDCG	P@10	R-Prec
$\alpha = 1$, $\beta = 0$	0.2529	0.5322	0.3157	0.2899
$\alpha = 1$, $\beta = 1$	0.2619	0.5409	0.3386	0.2986
$\alpha = 1$, $\beta = 2$	0.2623	0.5413	0.3400	0.2980
$\alpha = 1$, $\beta = 5$	0.2641	**0.5414**	0.3414	0.2987
$\alpha = 0$, $\beta = 1$	**0.2650**	0.5404	**0.3457**	**0.2995**
$\alpha = 2$, $\beta = 1$	0.2568	0.5350	0.3343	0.2900
$\alpha = 5$, $\beta = 1$	0.2503	0.5147	0.3157	0.2803

4.3 Parameters Setting

To find suitable parameters for our DRF method for Equation 9, several combinations of α, β values were tested in our experiments as shown in Table 4. Firstly we set $\alpha = 1$, the results show that larger values of β give better results; secondly we set $\beta = 1$, the results show that smaller α gives better results. From Table 4, we can see that $\alpha = 0$ and $\beta = 1$ gives the best result in our experiments. In Table 4, the number of feedback documents is 30 (a higher number than is typically used for standard PRF, but is more effective when using DBpedia) and the number of feedback terms is 10 (a typical value for QE using PRF).

To further investigate the impact of parameter setting, we compare the performance of DRF on external resource (Run: DRF) to the PRF on external resource (Run: PRF2) and PRF on target annotation collection results (Run: PRF1) with different parameter settings. On the left side of Figure 4, the number of feedback terms in all Runs is set as 10; on the right side, the number of feedback documents in all Runs is set as 30. As shown in Figure 4, DRF outperforms PRF1 and PRF2 when the number of feedback documents is larger than 15, where the number of feedback terms is fixed at 10; DRF outperforms PRF1 and PRF2 for all choices of number of feedback terms for a fixed number of feedback documents.

4.4 Comparing DRF with PRF

Table 5 shows results comparing DRF, PRF1 (baseline Run) and PRF2 with their best performance. The results in Table 5 indicate that PRF from DBpedia achieves higher retrieval effectiveness than the baseline based on the criterion of MAP. Furthermore, DRF outperforms PRF for all retrieval criteria. A paired t-test was applied to compare MAP for PRF2 and DRF ($p = 0.0069 < 0.05$; significant improvements are indicated by * in Table 5). Comparing the DRF

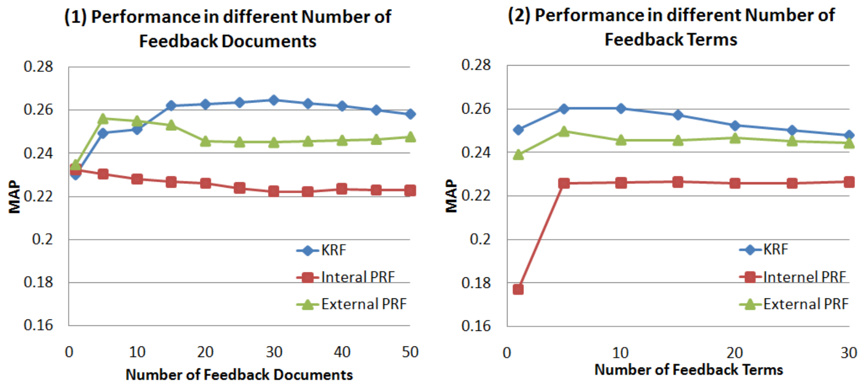

Fig. 4. Performance in Different Parameters Setting.

Table 5. Results Comparison.

Runs	MAP		NDCG	P@10	R-Prec
PRF1	0.2373		0.5055	0.3200	0.2772
PRF2	0.2529	+6.58%	0.5322	0.3157	0.2899
DRF	**0.2650**	**+11.67%**	**0.5404**	**0.3457**	**0.2995**

method with PRF on the target annotation collection, the result gains 11.67% in terms of MAP.

4.5 Comparing DRF with feedback from DDs

Since DDs play a very important role in the feedback process, one question is why not directly select terms from the DDs as the feedback terms. We carry out the PRF method to select feedback terms from DDs only using Equation 3. The results of this experiment are shown in Table 6. This experiment shows that using DRF is more effective in call cases.

4.6 Discussion

The key issue in QE is selecting feedback terms from the top ranked documents from the prior retrieval run. As stated previously, PRF assumes that all the top ranked documents are relevant, which will generally not be true. The identified relevant documents from Wikipedia help to judge which documents are more relevant to the query. Our results show that the DRF method can be effective for queries for which the DDs can be found in Wikipedia. However, feedback terms selected from non relevant documents can introduce a query drift problem for in the QE process.

Our results show that directly selecting feedback terms from DDs only does not perform better than our proposed method. The main reason for this is the

Table 6. Compare DRF with term selection from DDs only.

Parameters Setting	MAP	NDCG	P@10	R-Prec
DRF	**0.2650**	**0.5404**	**0.3457**	**0.2995**
DDs only	0.2403	0.5221	0.3180	0.2831

fact that the number of DDs is very small and cannot not provide enough information in the feedback process. Our term weighting method fully utilizes the characteristic of queries in image retrieval where all queries are noun phrases. We assume that it is easy to find DDs among the Wikipedia dataset for these queries. Our manual evaluation of the relevance of DDs on original topics proves that our assumption is true.

5 Conclusion and Future Work

In this paper we have introduced the incomplete annotation problem in image retrieval. As a solution to this, an external knowledge resource was introduced in the relevance feedback process. Comparing PRF on the target annotation collection to PRF on external resource, the external method achieves better results in our experiments. Furthermore, we presented a DD based relevance feedback method for QE from external resources. The key idea of the DRF method is to use the DDs identified from Wikipedia as an indicator to judge the quality of the feedback documents. The assumption is that the DDs provide more useful external knowledge in the process of feedback term selection. Thus combining the new weights from different rating scores, the DRF method can help to ensure that selected expansion terms from these documents with a high probability of being useful are used to expand the query knowledge, with the objective of solving the IAP problem in image retrieval. Our results show that the DRF method outperforms the PRF using the same external resource significantly.

We conclude that using the DDs as an evidence to help in QE is a good direction for utilizing Wikipedia related resources in text-based image retrieval research. For future work, the DRF method will be explore for other information retrieval tasks, including those which do not suffer so obviously from incomplete annotation.

6 Acknowledgments

This research is supported by the Science Foundation Ireland (Grant 07/CE/I1142) as part of the Centre for Next Generation Localisation (CNGL) project.

References

1. Tsikrika, T., Kludas, J.: Overview of the wikipediamm task at ImageCLEF 2008. In: Proceedings of the 9th Cross-language evaluation forum conference on Evaluat-

ing systems for multilingual and multimodal information access. CLEF'08, Berlin, Heidelberg, Springer-Verlag (2009) 539–550
2. Elsas, J.L., Arguello, J., Callan, J., Carbonell, J.G.: Retrieval and feedback models for blog feed search. In: SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA, ACM (2008) 347–354
3. Yang, X., Jones, G.J.F., Wang, B.: Query dependent pseudo-relevance feedback based on Wikipedia. In: SIGIR '09: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA, ACM (2009) 59–66
4. Yin, Z., Shokouhi, M., Craswell, N.: Query expansion using external evidence. In: ECIR '09: Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval, Berlin, Heidelberg, Springer (2009) 362–374
5. Kwok, K.L.: Improving English and Chinese ad-hoc retrieval: A Tipster text phase 3 project report. Inf. Retr. **3**(4) (2000) 313–338
6. Xu, Y., Ding, F., Wang, B.: Entity-based query reformulation using Wikipedia. In: CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge management, New York, NY, USA, ACM (2008) 1441–1442
7. Weerkamp, W., de Rijke, M.: External query expansion in the blogosphere. In: Seventeenth Text REtrieval Conference (TREC 2008), NIST, NIST (February 2009)
8. Custis, T., Al-Kofahi, K.: Investigating external corpus and clickthrough statistics for query expansion in the legal domain. In: CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge management, New York, NY, USA, ACM (2008) 1363–1364
9. Weerkamp, W., Balog, K., de Rijke, M.: A generative blog post retrieval model that uses query expansion based on external collections. In: ACL-IJCNLP '09: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2, Morristown, NJ, USA, Association for Computational Linguistics (2009) 1057–1065
10. Hersh, W.R., Bhupatiraju, R.T., Price, S.: Phrases, boosting, and query expansion using external knowledge resources for genomic information retrieval. In: TREC. (2003) 503–509
11. Min, J., Wilkins, P., Leveling, J., Jones, G.J.: DCU at WikipediaMM 2009: Document expansion from Wikipedia abstracts. In: Working Notes for the CLEF 2009 Workshop, Corfu, Greece (2009)
12. Robertson, S., Spärck Jones, K.: Simple, proven approaches to text retrieval. Technical Report UCAM-CL-TR-356, University of Cambridge, Computer Laboratory (December 1994)
13. Jaccard, P.: Etude comparative de la distribution florale dans une portion des alpes et des jura. Bulletin de la Socit Vaudoise des Sciences Naturelles **37** (1901) 547C–579
14. Westerveld, T., van Zwol, R.: The INEX 2006 multimedia track. In: N. Fuhr, M. Lalmas, and A. Trotman, editors, Advances in XML Information Retrieval:Fifth International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2006, Lecture Notes in Computer Science/Lecture Notes in Artificial Intelligence (LNCS/LNAI), Springer-Verlag, ACM (2007)