

# Ranking Refinement via Relevance Feedback in Geographic Information Retrieval

Esaú Villatoro-Tello, Luis Villaseñor-Pineda, and Manuel Montes-y-Gómez

Language Technologies Group, Computer Science Department,  
National Institute of Astrophysics, Optics and Electronics (INAOE),  
Mexico.{villatoroe, villasen, mmontesg}@inaoep.mx

**Abstract.** Recent evaluation results from Geographic Information Retrieval (GIR) indicate that current information retrieval methods are effective to retrieve relevant documents for geographic queries, but they have severe difficulties to generate a pertinent ranking of them. Motivated by these results in this paper we present a novel re-ranking method, which employs information obtained through a relevance feedback process to perform a *ranking refinement*. Performed experiments show that the proposed method allows to improve the generated ranking from a traditional IR machine, as well as results from traditional re-ranking strategies as query expansion via relevance feedback.

## 1 Introduction

Information Retrieval (IR) deals with the representation, storage, organization, and access to information items<sup>1</sup> [1]. Given some query, formulated in natural language by some user, the IR system is supposed to retrieve and sort according to its relevance degree documents satisfying user's information needs [2].

The word *relevant* means that retrieved documents should be semantically related to the user information need. Hence, one central problem of IR is determining which documents are, and which are not relevant. In practice this problem is usually regarded as a *ranking* problem, whose goal is to define an ordered list of documents such that documents similar to the query occur at the very first positions.

Over the past years, IR models, such as: Boolean, Vectorial, Probabilistic and Language models have represented a document as a set of representative keywords (i.e., index terms) and defined a *ranking* function (or retrieval function) to associate a relevance degree for each document with its respective query [1, 2]. In general, these models have shown to be quite effective over several tasks in different evaluation forums as can be seen in [3, 4]. However, the ability of these models to effectively *rank* relevant documents is still limited by the ability of the user to compose an appropriate query.

---

<sup>1</sup> Depending on the context, items may refer to text documents, images, audio or video sequences.

In relation to this fact, IR models tend to fail when desired results have implicit information requirements that are not specified in the keywords. Such is the case of **Geographic Information Retrieval** (GIR), which is a specialized IR branch, where search of documents is based not only in conceptual keywords, but also on spatial information (e.g., geographical references) [5]. For example, for the query: “Cities near active volcanoes”, expected documents should mention explicit city and volcanoes names. Therefore, GIR systems have to interpret implicit information contained in documents and queries to provide an appropriate response to geographical queries.

Recent development on GIR systems [6] evidence that: *i*) traditional IR systems are able to retrieve the majority of the relevant documents for most queries, but that, *ii*) they have severe difficulties to generate a pertinent ranking of them. To tackle this problem, recent works have explored the use of traditional re-ranking approaches based on query expansion via either relevance feedback [7–10], or employing knowledge databases [11, 12]. Although these strategies are very effective improving precision values, is known that query expansion strategies are very sensible to the quality of the added elements, and some times may result in degradation of the retrieval performance.

In this paper we propose a novel re-ranking strategy for a given set of retrieved documents in the context of Geographic Information Retrieval. Since retrieving relevant documents to geographic queries is not a problem for traditional IR systems, we focus on improving the order assigned to a set of retrieved documents by employing information obtained through a relevance feedback process, i.e., *ranking refinement via relevance feedback*. Furthermore, given that geographic queries tend to show a lot of implicit information, we propose the use of complete documents instead of isolated terms in the *ranking refinement* process.

The rest of the paper is organized as follows. Section 2 discusses some related work. Section 3 shows the proposed method. Section 4 describes the experimental platform used to evaluate our ranking strategy. Section 5 presents the experimental results. Finally, section 6 depicts our conclusions and future work.

## 2 GIR Related Work

Formally, a geographic query (geo-query) is defined by a tuple  $\langle \textit{what}, \textit{relation}, \textit{where} \rangle$  [5]. The *what* part represents generic terms (non-geographical terms) employed by the user to specify its information need, it is also known as the *thematic* part. The *where* term is used to specify the geographical areas of interest. Finally, the *relation* term specifies the “spatial relation”, which connects *what* and *where*.

GIR has been evaluated at the CLEF forum [3] since year 2005, under the name of the GeoCLEF task [6]. Their results evidence that traditional IR methods are able to retrieve the majority of the relevant documents for most geo-queries, but, they have severe difficulties to generate a pertinent ranking of them. Due to this situation, recent GIR methods have focused on the ranking subtask.

Common employed strategies are: *i*) query expansion through some feedback strategy, *ii*) re-ranking retrieved elements through some adapted similarity measure, and *iii*) re-ranking through some information fusion technique. These strategies have been implemented following two main approaches: first, techniques that had paid attention on constructing and including robust geographical resources in the process of retrieving and/or ranking documents. And second, techniques that ensure that geo-queries can be treated and answered employing very little geographical knowledge.

As an example of those on the first category, some works employ geographical resources in the query expansion process [11, 12]. Here, they first recognize and disambiguate all geographical entities in the given geo-query by employing a GeoNER<sup>2</sup> system. Afterwards, they employ a geographical ontology to search for these geo-terms, and retrieve some other related geo-terms. Then, retrieved geo-terms are given as feedback elements to the GIR machine. Some others approaches that focus on the ranking refinement problem, propose algorithms that consider the existence of *Geo-tags*<sup>3</sup>, therefore, the ranking function measures levels of *topological space proximity* among the geo-tags of retrieved documents and geo-queries [14]. In order to achieve this, geographical resources (e.g., geographical databases) are needed.

In contrast, approaches that do not depend on any robust geographical resource have proposed and applied variations of the query expansion process via relevance feedback, where no special consideration for geographic elements is made [7–10], and they have achieved good performance results. There are also works focusing on the ranking refinement problem; they consider the existence of several lists of retrieved documents (from one or many IR machines). Therefore, the ranking problem is seen as a information fusion problem, without any special processing for geo-terms contained in the retrieved documents. Some simple strategies only apply logical operators to the lists (e.g., AND) in order to generate one final re-ranked list [9], while some other works apply techniques based on information redundancy (e.g., CombMNZ or Round-Robin)[15–17].

Recent evaluation results indicate that there is not a notable advantage of knowledge-based strategies over methods that do not depend on any geographic resource. Motivated by these results, our proposed method do not make any special consideration for geographical terms. Our main hypothesis is that by employing information obtained through traditional relevance feedback strategies, is possible to perform an accurate ranking refinement process avoiding the drawbacks of query expansion techniques.

In addition, based on the fact that geo-queries often contain implicit information, we performed a set of experiments considering full documents (called *example documents*) in the process of re-ranking, showing that it is possible to become explicit some of the implicit information contained in the original geo-queries.

---

<sup>2</sup> Geographical Named Entity Recognizer.

<sup>3</sup> A *Geo-tag* indicates the geographical focus of certain item. As can be seen in [13], Geo-tagging and geo-disambiguating are both major problems in GIR.

### 3 Proposed Method

The proposed method consists of two main stages: the *retrieval stage* and the *re-ranking stage*. The goal of the first is to retrieve as many as possible relevant documents for a given query, whereas, the function of the second is to improve the final ranking of the retrieved documents by applying *ranking refinement* via *relevance feedback*. Figure 1 shows a general overview of the proposed method.

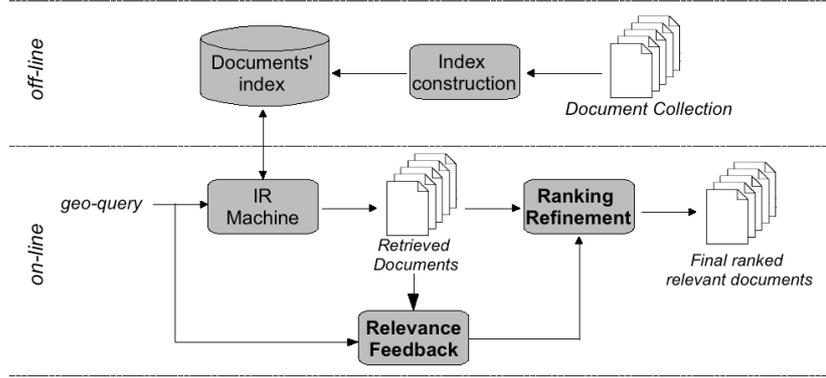


Fig. 1. Ranking refinement method

**Retrieval.** For this stage, we employed the vectorial space model (VSM), which is one of the most accurate and stable IR methods. In the *VSM* each document  $d$  is represented as a vector  $(\mathbf{d}_i)$  of length equal to the vocabulary size  $|V|$ . Each element  $j$  from the vector  $\mathbf{d}_i$  indicates how important is the word  $j$  inside the document  $\mathbf{d}_i$ . The set of vectors representing all the documents contained in the collection generate a vectorial space where documents can be compared through their representations. This vectorial space is represented by a matrix  $(M^{TD})$ , usually called term-document matrix (TD), of size  $N \times M$ , where  $N$  is the vocabulary size in the collection,  $N = |V|$ , and  $M$  is the number of documents in the collection. Each entry  $M_{i,j}^{TD}$  indicates the weight or contribution of term  $t_j$  in the document  $d_i$ .

We employed the *tf-idf* (*term-frequency inverse-document-frequency*) weighting scheme for each entry  $M_{i,j}^{TD}$ , computed as follows:

$$M_{ij}^{TD} = tf_{ij} \times \log\left(\frac{|D|}{df_j}\right) \quad (1)$$

where  $tf_{ij}$  is the number of occurrences of term  $j$  inside document  $d_i$ ,  $|D|$  is the total number of documents in the collection and  $df_j$  is the number of documents containing the term  $j$ .

When a query arrives to the VSM, it is transformed to the same representation (i.e., a vector). Once both query and documents are in the same representation it is possible to compare the query against all the documents in the collection. For this, we employed the cosine measure, which is computed as follows:

$$sim(q, d_i) = \frac{\sum_{j=1}^{|q|} w_{qj} d_{ij}}{\sqrt{\sum_{j=1}^{|q|} (d_{ij})^2 \sum_{j=1}^{|q|} (w_{qj})^2}} \quad (2)$$

where  $sim(q, d_i)$  represents the degree of similarity between the query ( $q$ ) and the document  $d_i$ .  $w_q$  is the query vector while  $d_i$  is the document vector. The cosine formula measures the angle between two vectors in a space of dimension  $|V|$  considering a normalization process to avoid that the vector's magnitude affects the retrieval process. Finally, the VSM method sort retrieved documents considering as its ranking score the result obtained with the cosine formula.

**Re-ranking.** Once a set of documents had been retrieved, we collect relevance judgments for the  $n$  top documents. Provided judgments could be either user provided (*manually*) or automatically obtained (*blindly*). Afterwards, the  $k$  most representative terms are extracted and along with the original geo-query are transformed to the VSM representation.

Our ranking refinement process considers the previous constructed vector to re-rank retrieved documents by employing the cosine formula (see formula 2), and hence, generate a final list of re-ranked documents. Notice that given this configuration, we preserve in every case the original recall levels obtained by the IR machine since we do not perform any further retrieval process.

At this point is where we also propose using complete documents (i.e., *example documents*) instead of  $k$  isolated terms in order to become explicit all implicit information contained in geo-queries. As will be seen in the experiments section, the  $n$  top documents along with the original geo-query are transformed to the VSM representation. Finally, retrieved documents are re-ranked according to its similarity degree against the *example documents* vector.

It is worth mentioning that we do not consider information contained in non-relevant documents for the relevance feedback process given that for geographical queries happens that: *i*) usually there are few relevant documents for geo-queries, i.e., there are many non-relevant documents in the retrieved set, and *ii*) non-relevant documents are not homogeneous. These reasons avoid the possibility of correctly represent non-relevant information.

## 4 Experimental Setup

### 4.1 Datasets

For our experiments we employed the GeoCLEF document collection composed from news articles from years 1994 and 1995. Articles cover as national as international events and, as a consequence, documents contain several geographic references. Table 1 shows some statistics about the collection.

**Table 1.** GeoCLEF Document Collection

Name	Origin	Num of Documents	Language
GH95	<i>The Glasgow Herald</i>	56,472	English
LAT94	<i>The Los Angeles Times</i>	113,005	English
		Total: 169,477	

## 4.2 Topics

We worked with the GeoCLEF 2008 queries. Table 2 shows the structure of each topic. The main query or title is between labels `<EN-title>` and `</EN-title>`. Also a brief description (`<EN-desc>`, `</EN-desc>`) and a narrative (`<EN-narr>`, `</EN-narr>`) are given. These last two fields usually contain more information about the requirements of the original query.

**Table 2.** Topic GC030: Car bombings near Madrid

```

<top>
<num>GC030</num>
<EN-title>Car bombings near Madrid</EN-title>
<EN-des>Documents about car bombings occurring
near Madrid</EN-desc>
<EN-narr>Relevant documents treat cases of car bombings occurring
in the capital of Spain and its outskirts</EN-narr>
</top>

```

## 4.3 Evaluation

The evaluation of results was carried out using two measures that have demonstrated their pertinence to compare IR systems, namely, the Mean Average Precision (*MAP*) and the *R-prec*. The *MAP* is defined as the norm of the average precisions (*AveP*) obtained for each query [1]. The *AveP* for a given query *q* is calculated as follows:

$$AveP = \frac{\sum_{r=1}^{Ret_{docs}} P(r) \times rel(r)}{Rel_{docs}} \quad (3)$$

where  $P(r)$  is the precision of the system at the  $r$  considered documents;  $rel(r)$  is a binary function that indicates if document  $r$  is relevant to the query, or not;  $Ret_{docs}$  is the number of retrieved documents, while  $Rel_{docs}$  is the number of relevant documents for the given query. Intuitively, this measure indicates how well the system puts into the firsts position relevant documents. It is worth pointing out that since our IR machine was configured to retrieve 1000 documents for each query, *AveP* values are measured at 1000 documents.

On the other hand,  $R$ -*prec* is defined as the precision reached after  $R$  documents have been retrieved, where  $R$  indicates the number of relevant documents for query  $q$  that exist in the entire document collection.

#### 4.4 Experiments definition

In order to achieve our goals, two experiments were performed. First, the Experiment 1 has as main goal to compare the proposed *ranking refinement* strategy against to the traditional re-ranking technique *query expansion via relevance feedback*. For this experiment, different number of documents ( $n$ ) as well different number of terms ( $k$ ) were considered. Additionally, we are also interested on evaluating the impact of both re-ranking strategies when selected documents are provided either by an active user (*manually selected*) or via an automatic selection process (*blindly selected*).

In the other hand, Experiment 2 has as main goal to evaluate the impact of considering full documents, which we call *example documents*, for both re-ranking strategies (*ranking refinement* and *query expansion*). Same as before, we are also interested on evaluate the performance obtained when *example documents* are manually or blindly selected.

## 5 Results

Experimental results are reported in Tables 3-8. Results are reported in terms of  $R$ -*prec* and  $MAP$ . Underlined results indicate the cases where the *baseline* was improved, while results marked in bold indicate the *best* results obtained over the different configurations.

For all our experiments, we consider as *baseline* the ranking generated by the IR machine, i.e., the rank assigned to documents employing the VSM method.

### 5.1 Experiment 1

Tables 3 and 4 show results obtained when documents are blindly selected, i.e., there are no user intervention. Table 3 shows results obtained after applying a query expansion process via relevance feedback (QEviaRF). As we can observe, the  $R$ -*prec* value obtained by the baseline method is never improved. This means that in any configuration the query expansion process was able to put more relevant documents among the first  $R$  retrieved documents (see section 4.3).

On the contrary, notice that when 2 documents are considered for the relevance feedback process the  $MAP$  values are better than the baseline, which means that the QEviaRF was able to shift some relevant documents to higher positions. However, remember that  $MAP$  values are measured at 1000 documents, so it is possible that the improvement implies that a document was shifted from position 1000 to position 900.

Table 4 shows results obtained after applying a *ranking refinement via relevance feedback* (RRviaRF). As we can observe, for this case it was possible to

**Table 3.** QEviaRF when documents are blindly selected

# selected terms	2 docs		5 docs		10 docs	
	<i>R-Prec</i>	<i>MAP</i>	<i>R-Prec</i>	<i>MAP</i>	<i>R-Prec</i>	<i>MAP</i>
5	0.2451	<u>0.2435</u>	0.2211	0.2214	0.2132	0.2064
10	<b>0.2545</b>	<b>0.2492</b>	0.2278	0.2152	0.2291	0.2025
15	0.2499	<u>0.2405</u>	0.2270	0.2017	0.2364	0.2209

*baseline: R-Prec = 0.2610; MAP= 0.2347*

improve *R-prec* value, particularly for the case when 2 documents with 10 and 15 documents are considered for the relevance feedback process, which means that more relevant documents are been collocated among the first *R* retrieved documents. Notice, that results obtained with RRviaRF are slightly better than those obtained with QEviaRF (Table 3), indicating that our ranking refinement method is less sensible to the noise contained in documents considered for the relevance feedback process.

**Table 4.** RRviaRF when documents are blindly selected

# selected terms	2 docs		5 docs		10 docs	
	<i>R-Prec</i>	<i>MAP</i>	<i>R-Prec</i>	<i>MAP</i>	<i>R-Prec</i>	<i>MAP</i>
5	0.2448	0.2281	0.2405	0.2313	0.2102	0.2013
10	<b>0.2675</b>	<u>0.2402</u>	0.2408	0.2194	0.1863	0.1914
15	<u>0.2619</u>	<b>0.2475</b>	0.2435	0.2257	0.2332	0.2265

*baseline: R-Prec = 0.2610; MAP= 0.2347*

As general conclusions, when documents for the re-ranking process are blindly selected, better results are achieved when only two documents are considered. Also, it is important to notice that RRviaRF strategy (Table 4) is less sensible to noise introduced by the selection of more terms from more documents than the QEviaRF (Table 3) technique. Besides, even when both strategies achieved a *MAP* close to 0.25, over a 0.23 from the *baseline*, RRviaRF is able to put more relevant documents among first positions according to the *R-prec* values.

Tables 5 and 6 show results obtained when documents are manually selected, i.e., an user intervention is considered. As expected, results get higher values under this schema. We notice in both tables that adding more terms from more documents to both QEviaRF and RRviaRF, allows to obtain better performance results than the baseline.

An interesting fact for these manual experiments, is that *R-prec* values are better for the case of QEviaRF (Table 5) than those obtained with RRviaRF (Table 6). However, observe that for QEviaRF results adding 15 or 10 terms has no notorious impact on the *MAP*. Same phenomena occurs when applying RRviaRF, except for the case when 5 documents are considered for the rele-

vance feedback process. Hence, one important question that emerge from these results is: *Why there is no improvement if added terms come from true relevant documents?*

**Table 5.** QEviaRF when documents are manually selected

# selected terms	2 docs		5 docs		10 docs	
	R-Prec	MAP	R-Prec	MAP	R-Prec	MAP
5	<u>0.3173</u>	<u>0.3260</u>	<u>0.3257</u>	<u>0.3367</u>	<u>0.3257</u>	<u>0.3378</u>
10	<u>0.3522</u>	<u>0.3494</u>	<u>0.3500</u>	<b>0.3601</b>	<u>0.3491</u>	<u>0.3584</u>
15	<u>0.3490</u>	<u>0.3471</u>	<u>0.3537</u>	<u>0.3590</u>	<b>0.3538</b>	<u>0.3593</u>

*baseline: R-Prec = 0.2610; MAP= 0.2347*

As an explanation to the previous question, first we must remember that given the configuration of both strategies, as a previous step to the re-ranking phase, an expanded keyword query is constructed from selecting  $k$  terms from  $n$  documents. Second, as mentioned in Section 1, IR models are limited by the ability of the user (or an automatic process) to compose an effective keyword query. These facts, plus obtained results, make us think that isolated keyword terms are not sufficient to achieve the best performance in a GIR system since they do not effectively describe implicit information needs contained in geographical queries.

**Table 6.** RRviaRF when documents are manually selected

# selected terms	2 docs		5 docs		10 docs	
	R-Prec	MAP	R-Prec	MAP	R-Prec	MAP
5	<u>0.2857</u>	<u>0.2760</u>	<u>0.2998</u>	<u>0.2924</u>	<u>0.2998</u>	<u>0.2932</u>
10	<u>0.3237</u>	<u>0.3178</u>	<u>0.3283</u>	<u>0.3274</u>	<u>0.3349</u>	<u>0.3313</u>
15	<u>0.3267</u>	<u>0.3236</u>	<b>0.3433</b>	<b>0.3378</b>	<u>0.3348</u>	<u>0.3348</u>

*baseline: R-Prec = 0.2610; MAP= 0.2347*

The general conclusion of these experiments is that, considering the user intervention allows to obtain better performance results; however adding more true relevant elements to the re-ranking strategies seems to have no impact at all, i.e., apparently a maximum level is achieved. Additionally, we confirmed that without any special treatment for geo-terms it is possible to achieve high recall levels ( $\simeq 90\%$ ).

## 5.2 Experiment 2

The main propose of following experiment is to employ more information than only a set of isolated keywords in the re-ranking process. For this, we based on the

ideas proposed in the Image Retrieval field [18], where in order to retrieve images of some particular type, queries are usually formulated through examples (i.e., example images<sup>4</sup>), which is easier than formulating an effective keyword query.

Same as for image retrieval, geographic retrieval also contain many implicit information that is hard to describe with a small query. Hence, the following tables show the results obtained when *example documents* are given to both re-ranking strategies, i.e., traditional query expansion, and the proposed method based on a ranking refinement strategy.

Table 7 shows results obtained when *example documents* are blindly selected, whereas in Table 8 example documents are manually selected. Tables compare both QEviaRF considering a query-by-example approach (QEviaRF-QBE) and RRviaRF under the same circumstances (RRviaRF-QBE).

**Table 7.** QEviaRF-QBE VS RRviaRF-QBE (documents are blindly selected)

# selected docs	QEviaRF-QBE		RRviaRF-QBE	
	<i>R-Prec</i>	<i>MAP</i>	<i>R-Prec</i>	<i>MAP</i>
2 docs	<b>0.2509</b>	<b>0.2393</b>	<b>0.2678</b>	<b>0.2498</b>
5 docs	0.2232	0.2254	<u>0.2656</u>	<u>0.2436</u>
10 docs	0.2118	0.2057	0.2148	0.2177
<i>baseline: R-Prec = 0.2610; MAP= 0.2347</i>				

Notice that when documents are blindly selected (Table 7), traditional query expansion technique is not able to improve the baseline, however, our proposed ranking refinement strategy obtains better results in two out of three cases (i.e., employing 2 and 5 example documents). Generally speaking our ranking refinement strategy allows better results than the query expansion strategy. However, when documents are selected manually (Table 8), query expansion strategy obtains better results than our ranking refinement method.

**Table 8.** QEviaRF-QBE VS RRviaRF-QBE (documents are manually selected)

# selected docs	QEviaRF-QBE		RRviaRF-QBE	
	<i>R-Prec</i>	<i>MAP</i>	<i>R-Prec</i>	<i>MAP</i>
2 docs	<u>0.3793</u>	<u>0.3795</u>	<u>0.3549</u>	<u>0.3528</u>
5 docs	<u>0.3850</u>	<u>0.3983</u>	<u>0.3651</u>	<u>0.3720</u>
10 docs	<b>0.3866</b>	<b>0.3995</b>	<b>0.3712</b>	<b>0.3747</b>
<i>baseline: R-Prec = 0.2610; MAP= 0.2347</i>				

As our general conclusion for these experiments, we consider that employing *example documents* allows generating a more pertinent ranking, since implicit

<sup>4</sup> This approach is usually known as query by example (QBE)

information contained in the original geo-query, is better represented by complete example documents. Obtained results also indicate that if the intervention of some user is considered, only providing 2 *example documents* is enough to reach acceptable results.

## 6 Conclusions

In this paper we have presented a *ranking refinement strategy via relevance feedback*. Obtained results showed that: *i)* our ranking refinement strategy considering a small set of keywords is able to improve the VSM method and also improves the traditional query expansion via relevance feedback technique, and *ii)* while more information is added to the ranking refinement strategy, a better ordering is provided.

Additionally, our experiments showed that employing full documents as relevance feedback elements (i.e., *example documents*), our ranking refinement strategy is able to obtain better performance results than those obtained when using a small set of keywords. This fact confirmed that it is possible to become explicit some of the implicit information contained in *geographical queries*.

Finally, performed experiments considering the user intervention, showed that is possible to reach a high performance results by providing only two *example documents*. This means that the user is not being overwhelmed in the process of selecting the *example documents*, since he has to mark very few documents as feedback elements from a small set of top retrieved documents.

As future work, we are planning to evaluate our method on a different data set as well as with a major number of queries, which will allow us to perform some statistical significance tests, in order to confirm the pertinence of the proposed method. Furthermore, we are interested in employing some strategy for selecting from the example documents the essential information, generating with this a more accurate ranking.

**Acknowledgments.** This work was done under partial support of CONACyT (scholarship 165545). We also thank SNI-Mexico and INAOE for their assistance.

## References

1. Ricardo Baeza-Yates and Berthier Ribeiro-Neto. *Modern Information Retrieval*. Addison Wesley, 1999.
2. David A. Grossman and Ophir Frieder. *Information Retrieval, Algorithms and Heuristics*. Springer, second edition edition, 2004.
3. Cross-lingual evaluation forum. <http://www.clef-campaign.org/>, May 2009.
4. Text retrieval conference (trec). <http://trec.nist.gov/>, May 2009.
5. Andreas Henrich and Volker Ldecke. Characteristics of geographic information needs. In *Proceedings of Workshop on Geographic Information Retrieval, GIR'07*, Lisbon, Portugal, 2007. ACM Press.

6. Thomas Mandl, Paula Carvalho, Fredric Gey, Ray Larson, Diana Santos, Christa Womser-Hacker, Giorgio Di Nunzio, and Nicola Ferro. Geoclef 2008: the clef 2008 cross language geographic information retrieval track overview. In *Working notes for the CLEF 2008 Workshop*, Aarhus, Denmark, September 2008.
7. Fredric Gey and Vivien Petras. Berkley2 at geoclef: Cross-language geographic information retrieval of german and english documents. In *Working notes for the CLEF 2005 Workshop*, Wien, Austria, September 2005.
8. Rocio Guillén. Monolingual and bilingual experiments in geoclef 2006. In *Working notes for the CLEF 2006 Workshop*, Alicante, Spain, September 2006.
9. Daniel Ferrés and Horacio Rodríguez. Talp at geoclef 2007: Using terrier with geographical knowledge filtering. In *Working notes for the CLEF 2007 Workshop*, Budapest, Hungary, September 2007.
10. Ray R. Larson. Cheshire at geoclef 2008: Text and fusion approaches for gir. In *Working notes for the CLEF 2008 Workshop*, Aarhus, Denmark, September 2008.
11. Riu Wang and Gunter Neumann. Ontology-based query construction for geoclef. In *Working notes for the CLEF 2008 Workshop*, Aarhus, Denmark, September 2008.
12. Nuno Cardoso, Patricia Sousa, and Mario J. Silva. The university of lisbon at geoclef 2008. In *Working notes for the CLEF 2008 Workshop*, Aarhus, Denmark, September 2008.
13. Karla A. Borges, Alberto H. F. Laender, Claudia B. Medeiros, and Clodoveu A. Davis Jr. Discovering geographic locations in web pages using urban addresses. In *Proceedings of Workshop on Geographic Information Retrieval GIR*, Lisbon, Portugal, 2007. ACM Press.
14. Bruno Martinis, Nuno Cardoso, Marcirio Silveira Chavez, Leonardo Andrade, and Mrio J. Silva. The university of lisbon at geoclef 2006. In *Working notes for the CLEF 2006 Workshop*, Alicante, Spain, September 2006.
15. Ray R. Larson. Cheshire ii at geoclef 2005: Fusion and query expansin for gir. In *Working notes for the CLEF 2005 Workshop*, Wien, Austria, September 2005.
16. Esaú Villatoro-Tello, Manuel Montes-y-Gómez, and Luis Villaseñor-Pineda. Inaoe at geoclef 2008: A ranking approach based on sample documents. In *Working notes for the CLEF 2008 Workshop*, Aarhus, Denmark, September 2008.
17. José M. Perea Ortega, L. Alfonso Ureña, Davide Buscaldi, and Paolo Rosso. Textmess at geoclef 2008: Result merging with fuzzy borda ranking. In *Working notes for the CLEF 2008 Workshop*, Aarhus, Denmark, September 2008.
18. M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: The qbic system. In *IEEE Computer*, pages 23–32, 1995.