A Probabilistic Method for Ranking Refinement in Geographic Information Retrieval*

Un Método Probabilístico para el Re-ordenamiento en Recuperación de Información Geográfica

Esaú Villatoro-Tello, R. Omar Chavéz-García, Manuel Montes-y-Gómez Luis Villaseñor-Pineda, L. Enrique Sucar

Laboratory of Language Technologies, Department of Computational Sciences,
National Institute of Astrophysics, Optics and Electronics (INAOE), Mexico.
E-mail:{villatoroe, romarcg ,mmontesg, villasen, esucar}@ccc.inaoep.mx

Resumen: Resultados recientes en la tarea de Recuperación de Información Geográfica (GIR) indican que los métodos de recuperación de información actuales son efectivos para recuperar documentos relevantes a las consultas geográficas, sin embargo tienen serias dificultades para generar un orden apropiado con los documentos recuperados. Motivado por estos resultados, este trabajo propone un método novedoso para re-ordenar la lista de documentos recuperados por un sistema GIR. El método propuesto está basado en un Campo Aleatorio de Markov (CAM), el cual combina el orden original obtenido por el sistema GIR, la similitud entre documentos, y un enfoque de retroalimentación de relevancia. La combinación de éstas características tiene el propósito de separar los documentos relevantes de los que no lo son y así obtener un orden más apropiado. Se realizaron experimentos con los recursos del foro GeoCLEF. Los resultados obtenidos muestran la viabilidad del método para re-ordenar documentos geográficos y también muestran una mejora en la medida MAP (Mean Average Precision) comparados con el modelo tradicional de espacio vectorial.

Palabras clave: Recuperación de Información, Recuperación de Información Geográfica, Re-rankeo, Modelos Probabilísticos, Campo Aleatorio de Markov

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 propose a novel method to re-order the list of documents returned by a GIR system. The proposed method is based on a Markov Random Field (MRF)model that combines the original order obtained by the GIR system, the similarity among documents and a relevance feedback approach, all of them with the purpose of separating relevant from irrelevant documents, and thus, obtaining a more appropriate order. Experiments were conducted with resources from the GeoCLEF forum. Obtained results show the feasibility of the method for re-ranking documents in GIR and also depict an improvement in mean average precision (MAP) compared to the traditional vector space model.

Keywords: Information Retrieval, Geographic Information Retrieval, Ranking Refinement, Probabilistic Models, Markov Random Fields

1 Introduction

Information Retrieval (IR) deals with the representation, storage, organization, and access to information items[1] (Baeza-Yates and Ribeiro-Neto, 1999). Given some query, formulated in natural language by some user, the IR system is suppose to retrieve and sort according to their relevance degree doc-

* This work was done under partial support of CONACyT(scholarships 165545 and 258311/224392)

[1]Depending on the context, items may refer to text documents, images, audio or video sequences.

uments satisfying user's information needs (Grossman and Frieder, 2004).

The word *relevant* means that retrieved documents should be semantically related to the user information need. Hence, one main 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 (Baeza-Yates and Ribeiro-Neto, 1999; Grossman and Frieder, 2004). In general, these models have shown to be quite effective over several tasks in different evaluation forums[2]. 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.

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 geographical terms (e.g., geographical references) (Jones and Prurves, 2008). 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 (Mandl et al., 2008) 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 (Larson, Gey, and Petras, 2006; Gillén, 2007; Ferrés and Rodríguez, 2008; Larson, 2008), or employing knowledge databases (Wang and Neumann, 2008; Cardoso, Sousa, and Silva, 2008). Although these strategies are effective improving precision values, is known that query expansion strategies are very sensitive 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 reranking strategy, which we apply 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, i.e., we focus on the *ranking refinement* problem. Our method combines the original order obtained by a GIR system, the similarity between documents obtained with textual features and a relevance feedback approach, all of them with the purpose of separating the relevant documents from does that are not relevant, and thus obtain a more appropriate order for the results generated by the base GIR system.

The proposed method is based on a Markov random field (MRF) model, in which each document in the list is represented as a random variable that could be *relevant* or *no relevant*. The relevance feedback is incorporated in the initialization of the model, making these documents *relevant*. The energy function of the MRF combines two factors: the similarity between the documents in the list (*internal* similarity); and external information obtained from the original order and the similarity of each document with the query (*external* similarity). Taking these factors into account and assigning a weight to each, the MRF is solved (obtaining the more probable configuration) so it separates the relevant documents from the rest. Based on this result, the list of documents is reordered (*re-ranked*) and given as final result to the user.

The rest of the paper is organized as follows. Section 2 discusses some related work in the field of geographic information retrieval. Section 3 shows the proposed method. Section 4 describes the experimental platform used to evaluate our ranking strategy. Section 5 presents the experimental

[2]CLEF (http://www.clef-campaign.org/) and TREC (http://trec.nist.gov/) forums

results. Finally, section 6 depicts our conclusions and future work.

2 Related Work

Formally, a geographic query (geo-query) is defined by a tuple <*what, relation, where*>(Henrich and Luedecke, 2007). 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 since year 2005, under the name of the GeoCLEF task (Mandl et al., 2008). 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 (Wang and Neumann, 2008; Cardoso, Sousa, and Silva, 2008; García-Cumbreras et al., 2009). Here, they first recognize and disambiguate all geographical entities in the given geo-query by employing a GeoNER[3] system. Afterwards, they employ a geographical ontology or thesaurus 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*[4], therefore, the ranking function measures levels of *topological space proximity* among the geo-tags of retrieved documents and geo-queries (Martins et al., 2007). 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 (Larson, Gey, and Petras, 2006; Gillén, 2007; Ferrés and Rodríguez, 2008; Larson, 2008), 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 (Ferrés and Rodríguez, 2008), while some other works apply techniques based on information redundancy (e.g., CombMNZ or Round-Robin)(Larson, Gey, and Petras, 2006; Villatoro-Tello, Montes-y-Gómez, and Villaseñor-Pineda, 2008; Ortega et al., 2008).

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 (Villatoro-Tello, Montes-y-Gómez, and Villaseñor-Pineda, 2009). Motivated by these results, our proposed method does not make any special consideration for geographical elements, i.e., we consider for measuring similarity among documents all its textual components. Also, our method does not require accessing again the entire collection, it considers only the list provided by the GIR system.

Our main hypothesis is that by employing information obtained through a feedback strategies, is possible to perform an accurate ranking refinement process avoiding the

[3]Geographical Named Entity Recognizer.

[4]A *Geo-tag* indicates the geographical focus of certain item. As can be seen in (Borges et al., 2007), Geo-tagging and geo-disambiguating are both major problems in GIR.

drawbacks of query expansion techniques. In addition, based on the fact that geo-queries often contain implicit information, our intuition is that by considering full documents in the process of re-ranking, it is possible to make explicit some of the implicit information contained in the original geo-queries.

3 Proposed Method

A general outline of the proposed method is given in Figure 1. Given a query, the GIR system retrieves from a given collection of documents a list of files sorted according to a relevance criteria. From this list, some relevant documents are selected based on a relevance feedback approach. For each document in the list, the textual features are extracted. The text contained in each document in the list, the query given by the user, and a subset of documents selected via relevance feedback, are combined to produce a re-ordered list. This re-ranking is obtained based on a Markov random field (MRF) model that separates the relevant documents from irrelevant ones, generating a new list by positioning the relevant documents first, and the others after. Next we give a brief review of MRFs, and then we describe in detail each component of the proposed method.

3.1 Markov Random Fields

Markov Random Fields (Li, 1994) are probabilistic models which combine *a priori* knowledge given by some observations and knowledge given by the interaction with neighbors.

Let $F = \{F_1, F_2, \ldots, F_n\}$ be random variables on a set S, where each F_i can take a value f_i in a set of labels L. This F is called a random field, and the instantiation of each of these $F_i \in F$ as an f_i, is what is called a configuration of F, so, the probability that a random variable F_i takes the value f_i is denoted by $P(f_i)$, and the joint probability is denoted as $P(F_1 = f_1, F_2 = f_2 \ldots, F_n = f_n)$. A random field is said to be an MRF if it has the property of *locality*, i.e., if the field satisfies the following property:

$$P(f_i | f_{S-\{i\}}) = P(f_i | f_{N_i})$$

where $S - \{i\}$ represents the set S without the i^{th} element, $f_{N_i} = \{f'_i | i' \in N_i\}$, and N_i is the set of neighboring nodes of the node f_i. The joint probability can be expressed as:

$$P(f) = \frac{e^{-U_p(f)}}{Z}$$

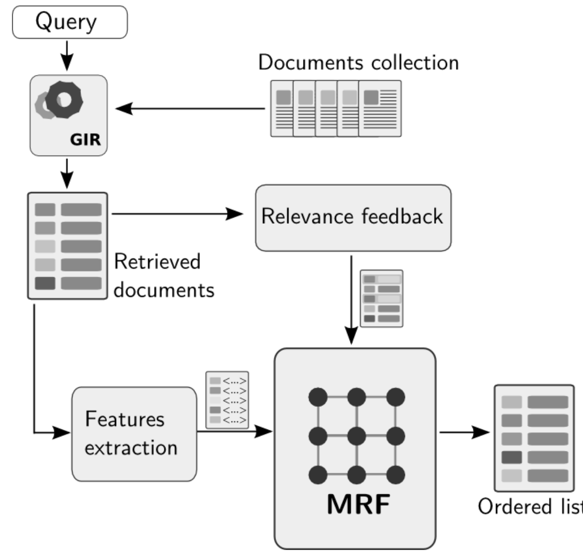

Figure 1: Block diagram of the proposed method. As input, it takes the original list obtained by an GIR system. Then, it considers a subset of relevant documents obtained via relevance feedback, and the textual features of the documents in the list. These elements, together with the original order, are integrated through the use of a MRF that divides the relevant documents from the rest to build a new, re-ordered list.

where Z is called the partition function or normalizing constant, and $U_p(f)$ is called the energy function. The optimal configuration is found by minimizing the energy function $U_p(f)$, obtaining a value for every random variable in F.

3.2 Model

In our case we consider a MRF in which each node corresponds to a document in the list. Each document is represented as a random variable with 2 possible values: *relevant* and *irrelevant*. We consider a fully connected graph, such that each node (document) is connected to all other nodes in the field; that is, we defined a neighborhood scheme in which each variable is adjacent to all the others. Given that the number of documents in the list is relatively low (1000 in the experiments), to consider a complete graph is not a problem computationally, and allows us to consider the relations between all documents in the list.

For representing the documents, and evaluating the internal and external similarities, we consider all the words contained in each document (except stopwords), without any special consideration for geographic elements. To describe the documents we used a binary bag of words representation, in which each vector element represents a word from the collection vocabulary; and the query is represented in the same manner. The internal and external similarities are considered via the energy function described next.

3.3 Energy Function

The energy function of the MRF combines two factors: the similarity between the documents in the list (*internal* similarity); and external information obtained from the original order and the similarity of each document with the query (*external* similarity). The internal similarities correspond to the interaction potentials and the external similarities to the observation potentials. The proposed energy function takes into account both aspects and is defined as follows:

$$U(f) = V_c(f) + \lambda V_a(f)$$

Where V_c is the interaction potential and it considers the similarity between random variable f and its neighbors, representing the support that neighboring variables give to f. V_a is the observation potential and represents the influence of external information on variable f. The weight factor λ favors V_c ($\lambda < 1$), V_a ($\lambda > 1$), or both ($\lambda = 1$).

V_c is defined as:

$$V_c(f) = \begin{cases} \bar{Y} + (1 - \bar{X}) & if\ f = irrelevant \\ \bar{X} + (1 - \bar{Y}) & if\ f = relevant \end{cases}$$

Where $\bar{Y}$ represents the average distance between variable f and its neighbors with irrelevant value. $\bar{X}$ represents the average distance between variable f and its neighbors with relevant value. The distance metric used to measure the similarity between variables is defined as: $1 - dice(f, g)$, where $dice(f, g)$ represents the *Dice* coefficient (Mani, 2001), and is defined as: $dice(f, g) = \frac{2|f \cap g|}{|f \cup g|}$. V_a is defined as follows:

$$V_a(f) = \begin{cases} (1 - dist(f, q)) \times g(posinv(f)) \\ iff = irrelevant \\ \\ dist(f, q) \times g(pos(f)) \\ iff = relevant \end{cases}$$

The V_a potential is obtained by combing two factors. The first indicates how similar, $dist(f, q)$, or different, $1 - dist(f, q)$ is the f variable with the query q. Where $dist(f, q)$ is defined as: $dist(f, q) = |f \cap q|/|q|$. The second is a function that converts the position in the list given by a base IR machine to a real value. The function used $g(x) = exp(x/20)/exp(5)$ (Chávez, Sucar, and Montes, 2010)[5]. The function $pos(f)$ returns the position of the document f in the original list, $posinv(f)$ returns the inverse position of the f variable in this list.

Having described each potential, the proposed energy function is defined as:

$$U(f) = \begin{cases} \bar{Y} + (1 - \bar{X}) + \lambda[1 - dist(f, q)) \\ \times g(posinv(f))] \\ iff = irrelevant \\ \\ \bar{X} + (1 - \bar{Y}) + \lambda dist(f, q) \times g(pos(f)) \\ iff = relevant \end{cases}$$

The initial configuration of the MRF is obtained by relevance feedback. That is, the subset of documents selected via relevance feedback are initialized as relevant, and all other documents as irrelevant. Then, the MRF configuration of minimum energy (MAP) is obtained via stochastic simulation using the ICM algorithm (we experimented using also Simulated Annealing with similar results). At the end of this optimization process, each variable (document) has a value of relevant or irrelevant. Based on these values, a new re-ordered list is produced, by positioning first the relevant documents according to the MRF, and then the not-relevant ones.

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.

4.2 Topics

We worked with the topics from GeoCLEF 2005 to GeoCLEF 2008. A total of 25 topics or queries were emitted for each year to total at the last conference in 2008 a set of 100 queries. Table 1 shows the structure of each

[5] The intuitive idea of this function is such that it first increases slowly so that the top documents have a small potential, and then it increases exponentially to amplify the potential for those documents in the bottom of the list.

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.

Table 1: 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>
```

For our experiments we employed the <EN-title> and the <EN-desc> fields.

4.3 Evaluation

The evaluation of results was carried out using measures that have demonstrated their pertinence to compare IR systems, namely, the Mean Average Precision (*MAP*) and the precision at *N (P@N)*. MAP is defined as follows:

$$MAP = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\left(\frac{\sum_{r=1}^{m} P_i(r) \times rel_i(r)}{n}\right)$$

Where $P_i(r)$ is the precision at the first r documents, $rel_i(r)$ is a binary function which indicates if document at position r is relevant or not for the query i; n is the total number of relevant documents for the query i, m is the number of relevant documents retrieved and Q is the set of all queries.

Intuitively, this measure indicates how well the system puts into the first positions relevant documents. It is worth pointing out that since our IR machine was configured to retrieve 1000 documents for each query, MAP values are measured at 1000 documents.

On the other hand, *P@N* is defined as the percentage of retrieved relevant items at the first N positions of the result list.

4.4 Experiments definition

We conducted a series of experiments with the following objectives: *i)* to test the results of the proposed method compared with the original list in the context of GIR, *ii)* to evaluate the sensitivity of the method to the model parameters.

Several experiments were conducted varying λ. Each experiments were made taking into account 1, 5 and 10 documents as relevance feedback. Simulated user feedback technique was used to perform the experiments. The collection used contains, in addition to the queries and documents, relevance judgments indicating which documents are relevant to each of the proposed queries, given that it is known beforehand which documents are relevant in the retrieved list, hence this documents are taken as feedback. This feedback type is known as simulated user feedback.

5 *Results*

Experimental results are reported in Table 3. Results are reported in terms of *P@5, P@10, P@20* and *MAP*. Results marked in **bold** indicate the *best* results obtained over the different configurations.

For our experiments we employed the results produced by the vectorial space model (VSM) configured in Lemur[6] using a TF-IDF weighting scheme as our baseline ranking. For comparison purposes with the rest of the GeoCLEF participants, Table 2 shows the best *MAP* results obtained among all the sumitted runs over the different years of the GeoCLEF, as well as the median and the worst results.

Table 2: Results obtained in the GeoCLEF

	MAP		
Year	*worst*	*median*	*best*
GeoCLEF 2005	0.1022	0.2600	0.3936
GeoCLEF 2006	0.0732	0.2700	0.3034
GeoCLEF 2007	0.1519	0.2097	0.2850
GeoCLEF 2008	0.1610	0.2370	0.3037

As can be seen in Table 3 our *baseline* results are very close to the median MAP obtained among participants in each of the GeoCLEF tracks (Table 2), except for the year 2007. However, remember that the majority of the GeoCLEF participants employ one or several geographical resources, or even more

[6] An open-source system designed to facilitate research in information retrieval (http://www.lemurproject.org/)

robust IR machines in order to retrieve relevant documents. Given this fact, we consider that our Lemur IR configuration is yielding acceptable baseline results.

Table 3 show a comparision between the results of the original list retrived by the Lemur IR machine and the results obtained with the proposed method for some different configurations of parameters. Notice that the values shown are an average of the values obtained for the 25 queries for each year. Also notice that for each of the considered measures, all variants of the proposed method improve the values of its corresponding baseline.

Table 3: A comparison between the results obtained by the VSM base ranker and the proposed method with some of its variants. The number after the letter F indicates the number of documents taken for relevance feedback, the number following the letter L indicates the value of λ

Year	Experiment	P@5	P@10	P@20	MAP
	Baseline	*0.5200*	*0.4360*	*0.3380*	*0.3191*
GeoCLEF	F1-L0.3	0.5440	0.4520	0.3440	0.3486
2005	F5-L0.3	0.9840	0.5760	0.3800	0.4627
	F10-L0.0	**0.9840**	**0.9320**	**0.5040**	**0.5910**
	Baseline	*0.3200*	*0.2560*	*0.1960*	*0.2618*
GeoCLEF	F1-L0.3	0.3680	0.2760	0.2060	0.3658
2006	F5-L0.0	0.8160	0.4600	0.2800	0.5881
	F10-L0.3	**0.8160**	**0.6520**	**0.3580**	**0.6942**
	Baseline	*0.2400*	*0.2160*	*0.1620*	*0.1612*
GeoCLEF	F1-L0.5	0.3040	0.2400	0.1720	0.1970
2007	F5-L0.3	0.7920	0.4360	0.2580	0.3909
	F10-L0.3	**0.7920**	**0.6600**	**0.3560**	**0.4960**
	Baseline	*0.3840*	*0.2960*	*0.2440*	*0.2347*
GeoCLEF	F1-L0.0	0.3840	0.2840	0.2360	0.2911
2008	F5-L0.0	0.8160	0.4720	0.3120	0.4068
	F10-L0.5	**0.8160**	**0.7320**	**0.3960**	**0.4959**

Results show that an improvement of 9%, 39%, 22% and 24% for years 2005, 2006, 2007 and 2008 respectively, is reached when only one document is selected as feedback; and as expected, as more documents are given as feedback, better performance is obtained. It is also important to notice, that when selecting one document as feedback element, reached results improve the median values from Table 2, except for year 2007. Aditionally, observe that when more elements are given as feedback (5, 10), MAP values are even better than the best result obtained for each year of the GeoCLEF track (Table 2).

Notice that the proposed method yields to better results when the value of *lambda* is small (e.g 0.3). So it seems that, at least for this collection, the information from the neighbors is more valuable than the information from the original order and the similarity with the query.

6 Conclusions

This paper proposed a method for improving the ranking of a list of retrieved documents by a GIR system. Based on a relevance feedback approach, the proposed method integrates the similarity between the documents in the list (*internal* similarity); and external information obtained from the original order and the query (*external* similarty), via a MRF to separate the relevant and irrelevant images in the original list.

Experiments were conducted using the resources of the forum GeoCLEF from years 2005 to 2008. For our experiments we avoid using any specialized geographical resource, since our main goal was to prove the pertinece of the method employing olny textual (document's words) features. Results showed that considering only one document as feedback, the proposed method improved the MAP up to 9%, 39%, 22% and 24% for years 2005, 2006, 2007 and 2008 respectively. An initial analysis idicates that for this collection, greater importance is given to the information from neighbors, obtained from the textual similarity between documents.

As future work we are considering including instead of textual features, geographical features, and we also intend to include a combination of both (textual and geographical) features to exploit the advantages of both.

References

Baeza-Yates, R. and B. Ribeiro-Neto. 1999. *Modern Information Retrival.* Addison Wesley.

Borges, K. A., A. H. F. Laender, C. B. Medeiros, and C. A. Davis Jr. 2007. Discovering geographic locations in web pages using urban addresses. In *Proceedings of Workshop on Geographic Information Retrieval GIR*, Lisbon, Portugal. ACM Press.

Cardoso, N., P. Sousa, and M. J. Silva. 2008. The university of lisbon at geoclef 2008. In *Working notes for the CLEF 2008 Workshop*, Aarhus, Denmark, September.

Chávez, O., E. Sucar, M. Montes. 2010. Image re-ranking based on relevance feedback combining internal and external similarities. In *The 23rd International FLAIRS Conference*, Daytona Beach, Florida, USA. In press.

Ferrés, D. and H. Rodríguez. 2008. Talp at geoclef 2007: Results of a geographical knowledge filtering approach with terrier. In *Advances in Multilingual and Multimodal Information Retrieval: 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007*, volume 5152 of *Lecture Notes in Computer Science*. Springer.

García-Cumbreras, M. A., J. M. Perea-Ortega, M. García-Vega, and L. A. Ureña-López. 2009. Information retrieval with geographical references. relevant documents filtering vs. query expansion. *Information Processing and Management*.

Gillén, Rocio. 2007. Monolingual and bilingual experiments in geoclef2006. In *Evaluation of Multilingual and Multi-modal Information Retrieval: 7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006*, volume 4730 of *Lecture Notes in Computer Science*. Springer.

Grossman, D. A. and O. Frieder. 2004. *Information Retrieval, Algorithms and Heuristics*. Springer, second edition.

Henrich, A. and V. Luedecke. 2007. Characteristics of geographic information needs. In *Proceedings of Workshop on Geographic Information Retrieval, GIR'07*, Lisbon, Portugal. ACM Press.

Jones, C. B. and R. S. Prurves. 2008. Geographical information retrieval. *International Journal of Geographic Information Science*, 22(3):219–228.

Larson, Ray R. 2008. Cheshire at geoclef 2008: Text and fusion approaches for gir. In *Working notes for the CLEF 2008 Workshop*, Aarhus, Denmark, September.

Larson, R. R., F. Gey, and V. Petras. 2006. Berkeley at geoclef: Logistic regression and fusion for geographic information retrieval. In *Accessing Multilingual Information Repositories: 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005*, volume 4022 of *Lecture Notes in Computer Science*. Springer.

Li, S. 1994. Markov random field models in computer vision. *Computer Vision ECCV '94*, pages 361–370.

Mandl, T., P. Carvalho, G. M. Di Nunzio, F. Gey, R. R. Larson, D. Santos, and C. Womser-Hacker. 2008. Geoclef 2008: The clef 2008 cross-language geographic information retrieval track overview. In *Evaluating Systems for Multilingual and Multimodal Information Access*, volume 5706 of *Lecture Notes in Computer Science*. Springer.

Mani, Inderjeet. 2001. *Automatic Summarization (Natural Language Processing, 3 (Paper))*. John Benjamins Publishing Co, June.

Martins, B., N. Cardoso, M. S. Chaves, L. Andrade, and M. J. Silva. 2007. The university of lisbon at geoclef 2006. In *Evaluation of Multilingual and Multi-modal Information Retrieval: 7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006*, volume 4730 of *Lecture Notes in Computer Science*. Springer.

Perea-Ortega, J. M., L. A. Ureña, D. Buscaldi, and P. Rosso. 2008. Textmess at geoclef 2008: Result merging with fuzzy borda ranking. In *Working notes for the CLEF 2008 Workshop*, Aarhus, Denmark, September.

Villatoro-Tello, E., M. Montes-y-Gómez, and L. Villaseñor-Pineda. Ranking refinement via relevance feedback in geographic information retrieval. In *Mexican International Conference on Artificial Intelligence (MICAI 2009)*, Guanajuato, Mexico, November 2009. Lecture Notes in Artificial Intelligence 5845, Springer, 2009.

Villatoro-Tello, E., M. Montes-y-Gómez, and L. Villaseñor-Pineda. 2008. Inaoe at geoclef 2008: A ranking approach based on sample documents. In *Working notes for the CLEF 2008 Workshop*, Aarhus, Denmark, September.

Wang, Riu and Gunter Neumann. 2008. Ontology-based query construction for geoclef. In *Working notes for the CLEF 2008 Workshop*, Aarhus, Denmark, September.