A Ranking Approach based on Example Texts for Geographic Information Retrieval[*]

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

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

Abstract. This paper focuses on the problem of ranking documents for Geographic Information Retrieval. It aims to demonstrate that by using some query-related *example texts* it is possible to improve the final ranking of the retrieved documents. Experimental results indicated that our approach could improve the MAP of some sets of retrieved documents using only two *example texts*.

1 Introduction

Geographic Information Retrieval (GIR) considers the search of documents based not only on conceptual keywords, but also on spatial information (i.e. geographical references) [1, 2]. Recent development of GIR systems [3] evidence that: *i)* traditional IR machines 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. Based on these facts, we designed a new GIR method that aims to improve the ranking of retrieved documents by considering information from some query-related *example texts*.

The proposed method was evaluated in the Monolingual English exercise of the 2008 GeoCLEF task [4]. In particular, the purpose of our experiments was two-fold: first, to confirm that traditional IR machines can achieve high recall levels, and second, to probe that using some query-related *example texts* allow improving the original ranking of the retrieved documents.

2 Proposed Method

Our method is divided in two main stages: the *retrieval stage* and the *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 ranking of the retrieved documents.

2.1 The Retrieval Stage

The core module of our method is the information retrieval (IR) machine. It is used two times: in a *first iteration*, it retrieves a set of relevant documents using the original query; then, in a *second iteration*, it retrieves a larger set of relevant documents considering an expanded query. The IR machine was implemented using LEMUR[1].

2.2 The Ranking Stage

Feedback Module. This module selects some "presumably relevant" items from the set of retrieved documents generated in the first iteration of the IR process. We call these items *example texts*, and use them for two different purposes: *i)* to modify the original query and perform the second iteration of the IR process, and *ii)* to re-rank the set of retrieved documents. The implementation of this module was based on the blind relevance feedback (BRF) technique.

Query Expansion Module. This module takes as input the set of *example texts* and extracts from them a set of relevant terms. Then, it uses these terms to expand the original query. The expanded query is sent to the IR machine, and a new set of documents is retrieved. Finally, this new set of documents is analyzed by the re-ranking module and the output of the system is generated.

Re-ranking Module. This module is the main contribution of our system. Its goal is to re-rank the set of retrieved documents using the information contained in the query-related *example texts*. This process considers the following steps:

1. **Geo-Expansion.** Expands all geographical terms contained in the *example texts*. It adds to each term its two nearest ancestors (e.g., *Madrid* → *Spain,Europe*). For this process we employed the Geonames database[2].

2. **Similarity Calculation.** Compares the retrieved documents against each *example text*, generating this way several different ranking proposals (one for each *example text*). The comparison of documents considers their *thematic* and *geographic* information. In particular, the similarity is computed as follows.

$$SQ(s,r) = (\lambda \times SQ_{thematic}(s,r)) + ((1-\lambda) \times SQ_{geographic}(s,r)) \quad (1)$$

where s represents an *example text*, r represents a document from the set of retrieved documents, and λ is a weighting value.

3. **Information Fusion.** Combines, into one single result list, all the information from the different ranking proposals. For this process we employed the well-known Round Robin technique.

---

[1] http://www.lemurproject.org/
[2] http://www.geonames.org/

3 Experiments and Results

This section describes the results from a subset of our experiments evaluated at GeoCLEF 2008. Table 1 shows the results from our four baseline runs. The first two rows correspond to the results of the first IR iteration. In this case, the run inaoe-BASELINE1 employed the title and description fields, whereas, the run inaoe-BASELINE2 used all available information: title, description, and narrative. As it can be noticed, the inclusion of the narrative did not improve the IR performance.

The third and fourth rows of the table show the results achieved in the second IR iteration, after the query expansion process. For these two experiments, we expanded the original query using the K most frequent terms from the top N documents retrieved by the inaoe-BASELINE1 run. We named these experiments as inaoe-BRF-N-K. As expected, the query expansion process allowed both configurations to obtain better results than the BASELINE1, especially for the case of the recall rate.

Table 1. Baseline results

	Experiment ID	MAP	R-Prec	P@5	Recall
1st	inaoe-BASELINE1	0.234	0.261	0.384	0.835
iteration	inaoe-BASELINE2	0.201	0.226	0.272	0.815
2nd	inaoe-BRF-5-2	**0.258**	**0.267**	**0.344**	**0.863**
iteration	inaoe-BRF-5-5	0.246	0.264	0.328	**0.863**

Table 2, under the column "submitted runs", shows the results achieved by the proposed method. In these experiments, we used the same N *example texts* for query expansion and for re-ranking the output of the *2nd iteration*. In particular, we considered $N = 5$ (i.e., inaoe-BRF-5-5) because we wanted to provide the greatest information to the re-ranking method. The results correspond to the following configurations:

1. RRBF: retrieved documents in the *2nd iteration* were re-ranked making **no** distinction between the *thematic* and *geographic* parts, i.e., similarity was computed using entire documents.
2. RRGeo: retrieved documents in the *2nd iteration* were re-ranked considering both *thematic* and *geographic* parts separately, i.e., applying Formula 1.
3. RRGeoExp: retrieved documents in the *2nd iteration* were re-ranked making distinction between *thematic* and *geographic* parts (applying Formula 1), and considering the Geo-Expansion process.

From Table 2, we can observe that the distinction of the thematic and geographic parts allowed obtaining the best performance. It is also possible to notice that the MAP difference between the experiments RRGeo-5-5 and RRGeoExp-5-5 was not very significant. We believe this performance was consequence of the noise introduced by our naïve geo-expansion process, which does not consider the disambiguation of geographical terms. That is, it can not distinguish between *Cordoba-Spain* and *Cordoba-Mexico*.

Table 2. Results of the proposed approach

Submitted Runs				Additional Experiments			
Experiment ID	**MAP**	**R-Prec**	**P@5**	**Experiment ID**	**MAP**	**R-Prec**	**P@5**
inaoe-RRBF-5-5	0.241	0.268	**0.384**	inaoe-RRBF‡	0.306	0.304	0.496
inaoe-RRGeo-5-5	0.244	0.266	**0.384**	inaoe-RRGeo‡	0.315	0.307	0.520
inaoe-RRGeoExp-5-5	**0.246**	**0.270**	**0.384**	inaoe-RRGeoExp‡	**0.318**	**0.310**	**0.536**

On the other hand, Table 2, under the column "additional experiments", shows the results of our method when there were used only truly relevant *example texts*. These experiments considered the manual selection of the *example texts*, and, somehow, they aimed to determine the best-possible result of our method. In all cases we used, at most, two *example texts*. Therefore, these results can be interpreted as: "By determining only two relevant *example texts*, we could reach a MAP of 0.318". These results show that the proposed method works well, but also indicate that it is very sensitive to the presence of incorrect *example texts*.

Finally, in order to support this conclusion, we made some significant tests. In particular, we employed the well-known Wilcoxon test. As noticed in Table 2(‡), only when we used manually-selected *example texts* we could obtain a significant improvement over the baseline result.

4   Conclusions

The results from our participation at GeoCLEF 2008 showed that the use of query-related *example texts* allows improving the original ranking of the retrieved documents. Nevertheless, they also showed that the proposed method is very sensitive to the presence of incorrect *example texts*, and that it is also affected by the incorrect expansion of the geographical terms. Our current work is mainly focused on tackling these drawbacks. In particular, due to our interest for having a fully automatic GIR process, we are working in a new example-text selection method based on machine learning techniques. On the other hand, we are also working in a better strategy for geographic query expansion.

References

1. R. Purves, and C. Jones editors : SIGIR 2004 Workshop on Geographic Information Retrieval, Sheffield, UK, 2004.
2. A. Henrich, and V. Lüdecke. Characteristics of Geographic Information needs. In *Proceedings of Workshop on Geographic Information Retrieval, GIR'07*, Lisbon, Portugal, 2007. ACM Press.
3. T. Mandl, F. Gey, G. Di Nunzio, N. Ferro, R. Larson, M. Sanderson, D. Santos, C. Womser-Hacker, and X. Xie. GeoCLEF 2007: the CLEF 2007 Cross-Language Geographic Information Retrieval Track Overview. In *the 8th Workshop of the Cross-Language Evaluation Forum*, CLEF 2007, Budapest, Hungary, 2007.
4. 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, 2008.