

# Comparing Several Textual Information Retrieval Systems for the Geographical Information Retrieval Task

José M. Perea-Ortega, Miguel A. García-Cumbreras, Manuel García-Vega,  
and L.A. Ureña-López

SINAI Research Group, Computer Science Department, University of Jaén, Spain  
{jmperea, magc, mgarcia, laurena}@ujaen.es

**Abstract.** This paper presents a comparison between three different Information Retrieval (IR) systems employed in a particular Geographical Information Retrieval (GIR) system, the GeoUJA IR, a GIR architecture developed by the SINAI research group. It could be interesting and useful for determining which of the most used IR systems works better in GIR task. In the experiments, we have used the Lemur, Terrier and Lucene search engines using mono and bilingual queries. We present baseline cases, without applying any external processes, such as query expansion or filtering. In addition, we have used the default settings of each IR system. Results show that Lemur works better using monolingual queries and Terrier works better using the bilingual ones.

## 1 Introduction

Geographic Information Retrieval (GIR) is related to a specialized branch of traditional Information Retrieval (IR). GIR concerns the retrieval of information involving some kind of spatial awareness. Existing evaluation campaigns as GeoCLEF<sup>1</sup> whose aim is to provide the necessary framework in which to evaluate GIR systems for search tasks, involving both spatial and multilingual aspects. GeoCLEF is a cross-language geographic retrieval track included in the Cross-Language Evaluation Forum<sup>2</sup> (CLEF) campaign. The selection of a good IR system is essential in this task. The main contribution of this paper is to compare three different textual IR systems for the GIR task.

### 1.1 The Geographical Information Retrieval Task

We can define GIR as the retrieval of geographically and thematically relevant documents in response to a query of the form <theme, location>, where the spatial relationship may either implicitly imply containment, or explicitly be selected from a set of possible topological, proximity and directional options

---

<sup>1</sup> <http://ir.shef.ac.uk/geoclef/>

<sup>2</sup> <http://www.clef-campaign.org/>

## 4 Conclusions

After the analysis of the overall results, and using the default configuration of each IR system, Lemur works better with the English monolingual queries, but the difference is not important. With the multilingual queries, Terrier works better than Lucene and Lemur. Specifically, Terrier obtains around a 38% and 65% of improvement with respect to Lemur and Lucene using German queries. Using Portuguese and Spanish queries, Terrier also improves the results.

Another conclusion is that the simple Okapi weighting function works better than simple TF-IDF. In addition, the use of PRF combined with Okapi or TF-IDF in Lemur improves the results obtained with simple Okapi or TF-IDF.

These conclusions have been obtained using the default configuration of each IR System. As future work, it would be interesting to test several weighting models and the simple TF-IDF, Okapi or PRF schemas in Lucene and Terrier for comparison them with the results obtained using the same weighting models in Lemur.

## Acknowledgments

The authors would like to thank CLEF in general, and Carol Peters in particular. This work has been supported by a grant from the Spanish Government, project TIMOM (TIN2006-15265-C06-03).

## References

1. Bucher, B., Clough, P., Joho, H., Purves, R., Syed, A.K.: Geographic IR Systems: Requirements and Evaluation. In: Proceedings of the 22nd International Cartographic Conference (2005)
2. Perea-Ortega, J.M., García-Cumbreras, M.A., García-Vega, M., Montejo-Ráez, A.: GEOUJA System. University of Jaén at GEOCLEF 2007. In: Working Notes of the Cross Language Evaluation Forum (CLEF 2007), p. 52 (2007)
3. García-Cumbreras, M.A., Ureña-López, L.A., Martínez-Santiago, F., Perea-Ortega, J.M.: BRUJA System. The University of Jaén at the Spanish task of QA@CLEF 2006. LNCS. Springer, Heidelberg (2007)
4. Mandl, T., Gey, F., Nunzio, G.D., Ferro, N., Larson, R., Sanderson, M., Santos, D., Womser-Hacker, C., Xie, X.: Geoclef 2007: the clef 2007 cross-language geographic information retrieval track overview. In: Proceedings of the Cross Language Evaluation Forum (CLEF 2007) (2007)
5. Buckley, C., Salton, G., Allan, J., Singhal, A.: Automatic query expansion using smart: Trec 3. In: Proceedings of TREC3, pp. 69–80. NIST, Gaithersburg (1995)