TextMESS at GeoCLEF 2008: Result Merging with Fuzzy Borda Ranking

José M. Perea Ortega and L. Alfonso Ureña López
Dpto. de Informática,
Universidad de Jaén, Spain
{jmperea, laurena}@ujaen.es

Davide Buscaldi and Paolo Rosso
Natural Language Engineering Lab,
Dpto. de Sistemas Informáticos y Computación (DSIC),
Universidad Politécnica de Valencia, Spain
{dbuscaldi, prosso}@dsic.upv.es

Abstract

This paper describe the joint participation by the Universidad Politécnica de Valencia and the Universidad of Jaén to the GeoCLEF task. This activity has been carried out within the framework of the Spanish TextMESS project (Intelligent, Interactive and Multilingual Text Mining based on Human Language Technologies). The method employed for the participation is a result merging algorithm based on the fuzzy Borda voting scheme. This method takes as input the two document lists returned by the two systems developed by the participating groups and creates a document list where the documents are ranked according to the fuzzy Borda voting scheme. The results obtained are better than the individual systems, and also ones of the best ones of the task (second as group). However, the best result was obtained with a run which combined the baseline systems. The analysis of the results showed that the best runs were those in which only title and description were used, and unfortunately we chose to submit only a run of this type, with the base systems. The results confirm the effectiveness of the fuzzy Borda scheme for the combination of different systems.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software; I.2 [**Artificial Intelligence**]: I.2.3 Uncertainty, "fuzzy," and probabilistic reasoning; I.2.7 Natural Language Processing

General Terms

Measurement, Performance, Experimentation

Keywords

Geographical Information Retrieval, Merging techniques, Fuzzy Borda

Table 3: Results obtained for the submitted runs.

run ID	MAP	UPV run MAP	Jaen run MAP	system diff.	improvement
TMESS01	0.226	0.201	0.225	0.024	0.1%
TMESS02	0.227	0.201	0.226	0.025	0.1%
TMESS03	0.219	0.201	0.212	0.011	0.8%
TMESS04	0.234	0.216	0.225	0.009	0.9%
TMESS05	0.235	0.216	0.226	0.01	0.9%
TMESS06	0.226	0.216	0.212	0.004	1.4%
TMESS07A	0.286	0.224	0.284	0.06	0.2%
TMESS08	0.216	0.203	0.212	0.009	0.4%
TMESS09	0.213	0.202	0.212	0.01	0.1%

Acknowledgements

We would like to thank the TIMOM (TIN2006-15265-C06-03) and TIN2006-15265-C06-04 research projects for partially supporting this work.

References

[1] Eric Brill. A simple rule-based part-of-speech tagger. In *Proceedings of the third Conference on Applied Natural Language Processing (ANLP'92)*, pages 152–155, Trento, Italy, 1992.

[2] Davide Buscaldi and Paolo Rosso. Upv-wsd : Combining different wsd methods by means of fuzzy borda voting. In *Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 434–437. ACL, 2007.

[3] Davide Buscaldi and Paolo Rosso. A conceptual density-based approach for the disambiguation of toponyms. *International Journal of Geographical Information Systems*, 22(3):301–313, 2008.

[4] Jenny Rose Finkel, Trond Grenager, and Christopher Manning. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics (ACL 2005)*, pages 363–370, U. of Michigan - Ann Arbor, 2005. ACL.

[5] Miguel A. García-Cumbreras, L. Alfonso Ureña López, Fernando Martínez Santiago, and José M. Perea Ortega. BRUJA System. The University of Jaén at the Spanish task of QA@CLEF 2006. In *Lecture Notes in Computer Science*, volume 4730 of LNCS Series, pages 328–338. Springer-Verlag, 2007.

[6] José Luis García Lapresta and Miguel Martínez Panero. Borda Count Versus Approval Voting: A Fuzzy Approach. *Public Choice*, 112(1-2):167–184, 2002.

[7] Ronald L. Graham. An efficient algorith for determining the convex hull of a finite planar set. *Information Processing Letters*, 1(4):132–133, 1972.

[8] Zhisheng Li, Chong Wanga, Xing Xie, and Wei-Ying Ma. Query Parsing Task for GeoCLEF 2007 Report. In *Proceedings of the Cross Language Evaluation Forum (CLEF 2007)*, 2007.

[9] Hannu Nurmi. Resolving Group Choice Paradoxes Using Probabilistic and Fuzzy Concepts. *Group Decision and Negotiation*, 10(2):177–199, 2001.

[10] M.F. Porter. An algorithm for suffix stripping. In *Program 14*, pages 130–137, 1980.