

# How geographic was GikiCLEF? A GIR-critical review

Diana Santos  
Linguatca, SINTEF ICT  
Oslo, Norway  
Diana.Santos@sintef.no

Nuno Cardoso  
Univ. of Lisbon, Faculty of  
Sciences, LaSIGE  
ncardoso@xldb.di.fc.ul.pt

Luís Miguel Cabral  
Linguatca, SINTEF ICT  
Oslo, Norway  
Luis.M.Cabral@sintef.no

## ABSTRACT

In this paper we draw a balance of GikiCLEF as far as its appropriateness for the evaluation of GIR systems is concerned. We measure its degree of dealing with geographic matter, and offer GIRA, the final resource, for GIR evaluation purposes.

## Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.3 Information Search and Retrieval

## General Terms

Algorithms, Design, Evaluation

## Keywords

Evaluation, Question Answering, Geographical IR, Wikipedia, Crosslinguality, Multilinguality

## 1. INTRODUCTION

As proposed in [5], a new track in QA@CLEF was organized in 2009 as a follow-up of the GikiP pilot track in GeoCLEF 2008 [6] with the purpose of continuing the evaluation of geographically complex queries (either stated as ad hoc IR topics or open list questions). This goal, however, was met with mixed success [3, 4, 1].

We provide a very short description of GikiCLEF and its resources, with a look at the “geographicity” of the task and a discussion of the relationship of GikiCLEF and the GIR area, as well as measure up what was actually done compared to what the intentions of the proposers were.

## 2. TRACK DESCRIPTION

GikiCLEF participant systems were offered the following task against a Wikipedia collection in ten languages: they should answer difficult open questions which all imply, to a smaller or larger degree, a measure of geographical reasoning. Each answer was given by a Wikipedia page, along

with a list of justification pages in case the justification was not included in the answer page. In order to make it a truly multilingual venture, answers in any language were considered correct as long as the full justification had been found in at least one language.

The full topic set (with a couple of example topics) is publicly available from GikiCLEF’s site, <http://www.linguatca.pt/GikiCLEF/>, from where one can also download the pool(s) and SIGA, the GikiCLEF management system [4]. SIGA was deployed to help multilingual topic creation and assessment, as well as process submissions and provide final scores, and was offered to the community for the organization of further similar evaluation contests. The set of GikiCLEF resources can be downloaded as the GIRA package, released November 2009.

There were 17 runs submitted by eight participants, although about thirty groups showed interest when the initial task draft was published. There was exactly one participant each from Brazil, Germany, the Netherlands, Portugal, Romania, Spain, UK and the US (some of them were distributed teams so we are simplifying matters slightly here), so we have to say that the participation was pretty well distributed as to geographical distribution by country/language.

Evaluation measures were kept simple (and not geographic), designed to obey two constraints only: i) the more languages the participant systems provided answers in, the better, and ii) systems should not be penalized if there were no answers in a particular Wikipedia language. The score of a submitted run was therefore the sum, for each language, of precision times the number of correct ( $C$ ) answers. For each language, the score was  $C * C / N$  (so that one had a score for de, pt, etc, as  $C_{de} * C_{de} / N_{de}$ ,  $C_{pt} * C_{pt} / N_{pt}$ , etc.). The winner run had a score of 182, the two following systems (96 and 91) were only semi-automatic, and the fourth run scored 25, while the worst three runs had a zero score.

As explained in the GikiCLEF overview papers already mentioned, the multilinguality bias of the GikiCLEF track was seriously flawed by the way the topic group chose their topics, which in fact resulted in a preference for English as a pivot language. But what concerns us here is not the crosslingual or multilingual aspects of GikiCLEF but how it fared as an evaluation testbed for geographical reasoning.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GIR’10, 18-19th Feb. 2010, Zurich, Switzerland

Copyright 2010 ACM ISBN 978-1-60558-826-1/10/02 ...\$10.00

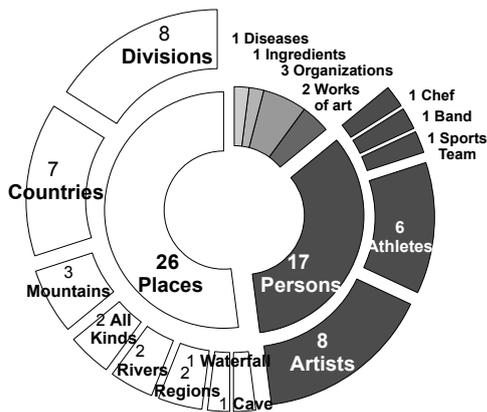


Figure 1: The expected answer type distribution of the topics

### 3. GEOGRAPHIC IN WHAT SENSE?

We take here the broad view that every question that is directly or indirectly associated with geographical knowledge (even if not requesting geographic places as answer) should be tackled by geographically-aware and geographically-conscious systems, and so GikiCLEF as an evaluation task should be interesting for the GIR community irrespective of the techniques used or the particular goals. But we have to understand more specifically in which sense(s) GikiCLEF was geographical and in which senses it is lacking for the purposes of the GIR community at large. The first approach to characterize the “geographicity” of the task is to actually look at the expected answer types in GikiCLEF. From Figure 1, we observed that the most frequent expected answer types (26) were places after all. Although we have to remember that countries for example are not necessarily strictly places, that is: they are or can be conceived equally easily as organizations, societies, ideas/cultures, or groups of people. As to granularity, one can see that countries are the largest “places”, while caves, ski resorts and rivers are found at the other end of the scale.

We also surveyed where in the earth were the topics located (again, assuming that this question makes sense, which is not always the case as argued for by [2]). Making the rough approximation that cultural bias corresponds to geographical bias, Figure 2 provides an idea of the places under GikiCLEF’s spotlight. Intersection areas in the figure correspond to either broader areas such as Latin America, specific presence of a culture icon in another culture (such as Picasso in American museums or in Hemingway in Italy), or to language-complex places such as Switzerland. However, this is a very gross distribution, since it mingles human (political) geography with natural geography, and in some cases the topics can be considered, though mixing two cultures, not geographical at all, such as Dutch bands named after a Bulgarian fighter.

What we intend to do next is study the need and kind of geographical reasoning implied by each topic, so that we can arrive at a sense of geographically challenging topics. Whether one should consider questions about nationality or culturally geographical is a matter of free choice, but there

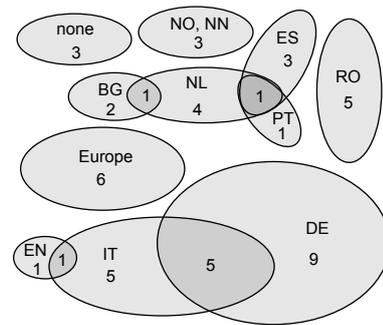


Figure 2: Language-bias of GikiCLEF topics

is no denying that languages (and people) mix both, so that in the end systems have to deal with it.

### 4. ACKNOWLEDGEMENTS

Thanks to fellow organizers and topic choosers C. Forascu, P. Forner, D. Giampiccolo, K. Lamm, T. Mandl, P. Osenova, A. Peñas, E. Tjong Kim Sang, J. Schulz, Y. Skalban and A. Rodrigo Yuste. This work was supported by FCT for its LASIGE Multi-annual support, GREASE-II project (grant PTDC/EIA/73614/2006) and a PhD scholarship grant SFRH/BD/45480/2008, and by the the Linguateca project, financed by UMIC and FCCN.

### 5. REFERENCES

- [1] N. Cardoso. GikiCLEF topics and Wikipedia articles: did they blend? In C. Peters et al., editors, *Multilingual Information Access Evaluation Vol. I Text Retrieval Experiments*. Springer, 2010.
- [2] F. Gey, R. Larson, M. Sanderson, K. Bishoff, T. Mandl, C. Womser-Hacker, D. Santos, P. Rocha, G. D. Nunzio, and N. Ferro. GeoCLEF 2006: the CLEF 2006 Cross-Language Geographic Information Retrieval Track Overview. In C. Peters et al., editors, *Evaluation of Multilingual and Multi-modal Information Retrieval: 7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006. Revised selected papers*, pages 852–876. Springer, 2007.
- [3] D. Santos and L. M. Cabral. GikiCLEF: Crosscultural issues in an international setting: asking non-English-centered questions to Wikipedia. In F. Borri, A. Nardi, and C. Peters, editors, *Cross Language Evaluation Forum: Working Notes of CLEF 2009*, Corfu, Greece, 2009.
- [4] D. Santos and L. M. Cabral. GikiCLEF: Expectations and lessons learned. In C. Peters et al., editors, *Multilingual Information Access Evaluation Vol. I Text Retrieval Experiments*. Springer, 2010.
- [5] D. Santos and N. Cardoso. GikiP: Evaluating geographical answers from Wikipedia. In *Proceedings of the 5th Workshop on Geographic Information Retrieval, GIR’08 (CIKM’2008 Workshop)*, Napa Valley, CA, USA, 29–30th October 2008. ACM.
- [6] D. Santos, N. Cardoso, P. Carvalho, I. Dornescu, S. Hartrumpf, J. Leveling, and Y. Skalban. GikiP at GeoCLEF 2008: Joining GIR and QA forces for querying Wikipedia. In C. Peters et al., editors, *Evaluating Systems for Multilingual and Multimodal Information Access: 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008, Revised Selected Papers*, pages 894–905. Springer, 2009.