GikiCLEF: Crosscultural issues in multilingual information access

Diana Santos*, Luis Miguel Cabral*, Corina Forascu**, Pamela Forner***, Fredric Gey****, Katrin Lamm*****, Thomas Mandl******, Petra Osenovat, Anselmo Peñas$, Álvaro Rodríguez$, Julia Schulz******, Yvonne Skalbnak$$, Erik Tjong Kim Sang$%

*STINTEC, **UAIC, ***CELCT, ****Univ. Berkeley, *****Univ. Hildesheim, BTB group, IPIO, BAS, SUNEDE, &Univ. Wolverhampton, %Univ. Groningen

http://www.linguateca.pt/GikiCLEF/ (CLEF 2009, http://www.clef-campaign.org/) 50 topics developed with non-English users in mind, to evaluate systems that should answer open natural language questions to Wikipedia, using its multilingual and crosslingual properties

10 collections, 8 participants, ca.18,000 different answers, 1,400 correct answers

GikiCLEF topics

Geographical distribution of topic scope

Expected answer type

Wikipedia solutions

Number of solutions per topic

Language coverage on topics

Language autonomy on topics

Conclusions

Systems answer complex questions in several different languages against Wikipedia

A sizable resource for further development and evaluation of such systems is available at http://www.linguateca.pt/GikiCLEF/GIRA/ together with an open source system, SIGA, to drive and evaluate GikiCLEF-like evaluation contests.

Acknowledgments: We thank the remaining GikiCLEF organizers, Soren Auer, Gosse Bouma, Iustin Demcecu, Danilo Gimpiccolo, Sven Hartmann, Roy Larson, Johannes Levinger, and Constantin Orsan; the other assessors, Asnabela Barreiro, Leda Cazanova, Luis Costa, Ana Frug$%, Laska Laskova, Cristina Mota, Rosário Silva, and Kiril Simov; Paula Carvalho and Christian-Emil Oye for help on Portuguese and Norwegian topics respectively, and GikiCLEF participants. The Linguateca project is funded by the Portuguese Government, the European Union (FEDER and FSE), under contract ref. POSC/339/1.3/C/NAC, by UMIC and FCCN. Nuno Cardoso is supported by FCT scholarship grant SFRH/BD/45480/2008. We acknowledge support of the Treble-CLEF Coordination Action. ICT-1-4-1 Digital libraries and technology-enhanced learning (Grant agreement: 215231) for GikiCLEF assessment. Álvaro Rodrigo has been partially supported by the Education Council of the Regional Government of Madrid and the European Social Fund.

NO: 26 PT: 117

Answer in the body text:

- EN: 156 PT: 43

Answer in the infobox:

- EN: 30 PT: 14

No answer: