



GikiCLEF: Crosscultural issues in multilingual information access

Diana Santos*, Luis Miguel Cabral*, Corina Forascu**, Pamela Forner***, Fredric Gey****, Katrin Lamm*****, Thomas Mandl*****, Petya Osenova\$, Anselmo Peñas\$, Alvaro Rodrigo\$, Julia Schulz*****, Yvonne Skalban&, Erik Tjong Kim Sang%

Nuno Cardoso
Faculty of Sciences, University of Lisbon, LaSIGE, and SINTEF ICT
ncardoso@xldb.di.fc.ul.pt



*SINTEF ICT, **UAIC, ***CELCT, ****Univ. Berkeley, *****Univ. Hildesheim, £BTB group, IPOI, BAS, \$UNED, &Univ. Wolverhampton, % Univ. Groningen

Diana.Santos@sintef.no, Luis.M.Cabral@sintef.no, corinfor@info.uaic.ro, forner@celct.it, gey@berkeley.edu, katrin.lamm@uni-hildesheim.de, mandl@uni-hildesheim.de, petya@bulreebank.org, anselmo@lsi.uned.es, alvaror@lsi.uned.es, julia-maria.schulz@uni-hildesheim.de, Yvonne.Skalban@wlv.ac.uk, erikt@xs4all.nl

<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

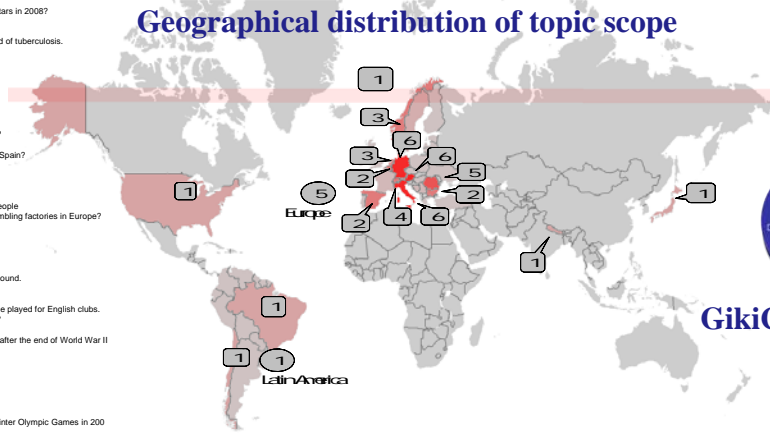
GikiCLEF topics

- [IT] 01. List the Italian places where Ernest Hemingway visited during his life.
- [BG] 02. In which countries outside Bulgaria are there published opinions on Petar Dunov's (Beinsa Duno's) ideas?
- [RO] 03. Name Romanian poets who published volumes with ballads until 1941.
- [RO] 04. Name Romanian poets who published volumes with ballads until 1941.
- [NL] 05. Which written fictional works of non-Romanian authors have as subject the Carpathians mountains?
- [NL] 06. Which Dutch violinists held the post of concertmaster at the Royal Concertgebouw Orchestra in the twentieth century?
- [NL] 07. What capitals of Dutch provinces received their town privileges before the fourteenth century?
- [DE] 08. Which authors were born in and write about the Bohemian Forest?
- [DE] 09. Name places where Goethe fell in love.
- [NL] 10. Which Finnish towns hosted a restaurant with two or three Michelin stars in 2008?
- [NL] 11. What Belgians won the Ronde van Vlaanderen exactly twice?
- [IT] 12. Present monarchies in Europe headed by a woman.
- [IT] 13. Romantic and realist European novelists of the XIXth century who died of tuberculosis.
- [IT] 14. Name rare diseases with dedicated research centers in Europe.
- [IT] 15. List the basic elements of the cassata.
- [IT] 16. In which European countries is the bidet commonly used?
- [IT] 17. List the 5 Italian regions with a special statute.
- [IT] 18. In which Tuscan provinces is Chianti produced?
- [ES] 19. Name mountains in Chile with permanent snow.
- [IT] 20. List the name of the sections of the North-Western Alps.
- [IT] 21. List the left side tributaries of the Po river.
- [ES,PT] 22. Which South American national football teams use the yellow color?
- [EN] 23. Name American museums which have any Picasso painting.
- [ES] 24. Which countries have won a total European championship played in Spain?
- [ES] 25. Name Spanish drivers who have driven in Minardi.
- [NL] 27. Which Dutch bands are named after a Bulgarian footballer?
- [PT] 28. Find coastal states with Petrobras refineries.
- [IT] 29. Places above the Arctic circle with a population larger than 100,000 people.
- [IT] 30. Which Japanese automakers companies have manufacturing or assembling factories in Europe?
- [RO] 31. Which countries have Italian as an official language?
- [RO] 32. Name Romanian writers who were living in USA in 2003.
- [IT] 33. What European Union countries have national parks in the Alps?
- [IT] 34. What eight-thousanders are at least partially in Nepal?
- [RO] 35. Which Romanian mountains are declared biosphere reserves?
- [RO] 36. Name Romanian caves where Paleolithic human fossil remains were found.
- [NO,NB] 37. Which Norwegian musicians were convicted for burning churches?
- [NO,NB] 38. Which Norwegian waterfalls are higher than 200m?
- [NO,NB] 39. National team football players from Scandinavia with sons who have played for English clubs.
- [DE] 40. Which rivers in North Rhine Westphalia are approximately 10km long?
- [DE] 41. Chefs born in Austria who have received a Michelin Star.
- [DE] 42. Political parties in the National Council of Austria which were founded after the end of World War II.
- [DE] 43. Austrian ski resorts with a total ski trail length of at least 100 km.
- [DE] 44. Find Austrian grape varieties with a vineyard area below 100 ha.
- [DE] 45. Find Swiss skiing show winners.
- [DE] 46. German writers who are Honorary Citizens in Switzerland.
- [DE] 47. Which cities in Germany have more than one university?
- [DE] 48. Which German-speaking movies have been nominated for an Oscar?
- [DE] 49. Formula One drivers who moved to Switzerland.
- [IT] 50. Which Swiss people were Olympic medalists in snowboarding at the Winter Olympic Games in 200

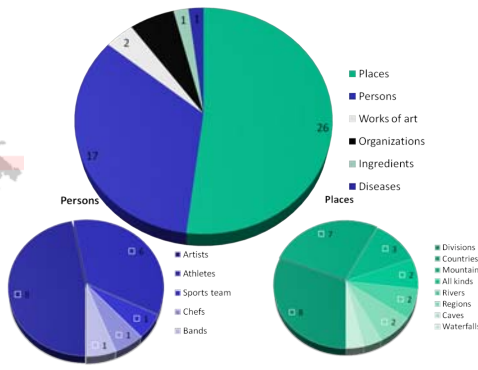
Language bias of topics

| none | de | it | nl | ro | es | no | bg | pt | en |
|------|----|----|----|----|----|----|----|----|----|
| 13 | 10 | 6 | 5 | 5 | 4 | 3 | 2 | 2 | 1 |

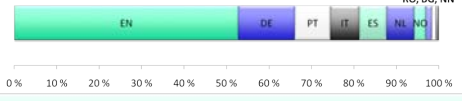
Geographical distribution of topic scope



Expected answer type



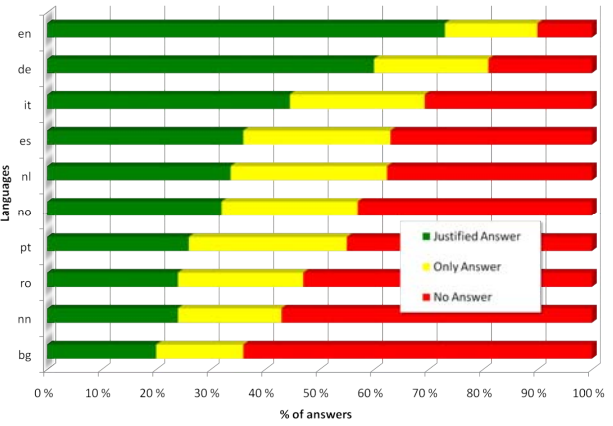
GikiCLEF collection size (no. of pages)



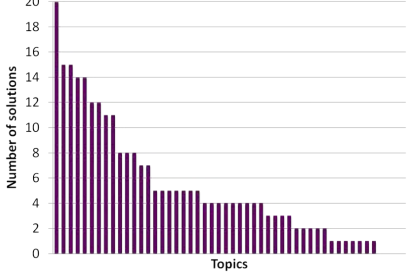
Wikipedia solutions

Solutions: entities found in the pool of all GikiCLEF participant's answers, merged to a single identifier
Example: Italy is one solution for a topic, regardless of the different answers given by Wikipedia articles from different languages

Language coverage on topics



Number of solutions per topic



Best language(s) per topic bias

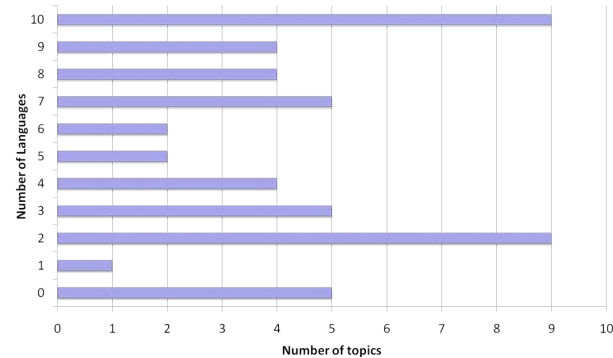
| | bg | de | en | es | it | nl | nn | no | pt | ro |
|------|----|----|----|----|----|----|----|----|----|----|
| none | 13 | 7 | 10 | 11 | 8 | 9 | 7 | 5 | 7 | 7 |
| bg | 2 | 2 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| de | 10 | 3 | 9 | 9 | 4 | 5 | 5 | 3 | 4 | 4 |
| en | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| es | 4 | 1 | 3 | 3 | 2 | 1 | 1 | 0 | 1 | 2 |
| it | 6 | 3 | 6 | 6 | 5 | 6 | 4 | 3 | 5 | 5 |
| nl | 5 | 2 | 3 | 4 | 2 | 3 | 3 | 1 | 3 | 2 |
| nn | 3 | 0 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 1 |
| no | 2 | 0 | 2 | 2 | 1 | 0 | 0 | 0 | 0 | 2 |
| pt | 2 | 0 | 2 | 2 | 1 | 0 | 0 | 0 | 0 | 2 |
| ro | 5 | 1 | 3 | 5 | 3 | 3 | 2 | 1 | 2 | 2 |

Location of solutions in Wikipedia

In a non-answer page in the same language:
EN: 45 PT: 76

No answer:
EN: 26 PT: 117

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.

Acknowledgements: We thank the remaining GikiCLEF organizers, Soren Auer, Gosse Bouma, Justin Dornescu, Danilo Giampiccolo, Sven Harttrumpf, Ray Larson, Johannes Leveling, and Constantin Orasan; the other assessors, Anabela Barreiro, Leda Casanova, Luis Costa, Ana Eng, Laska Laskova, Cristina Mota, Rosário Silva, and Kiril Simov; Paula Carvalho and Christian-Emil Ore 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. A lvaro Rodrigo has been partially supported by the Education Council of the Regional Government of Madrid and the European Social Fund.