



GeoCLEF 2006



Cross-Language Geographical Information Retrieval

First Announcement and Call for Participation



Why cross-language geographical IR?

Geographical Information Retrieval (GIR) concerns the retrieval of information involving some kind of spatial awareness. Given that many documents contain some kind of spatial reference, there are examples where geographical references (geo-references) may be important for IR. For example, to retrieve, re-rank and visualise search results based on a spatial dimension (e.g. "find me news stories about riots *near* Dublin City"). In addition to this, many documents contain geo-references expressed in multiple languages which may or may not be the same as the query language. This would require an additional translation step to enable successful retrieval. Existing evaluation campaigns such as TREC and CLEF do not explicitly evaluate geographical IR relevance. The aim of GeoCLEF is to provide the necessary framework in which to evaluate GIR systems for search tasks involving both spatial and multilingual aspects. After its success as a pilot experiment in CLEF 2005, GeoCLEF will be a regular track in CLEF 2006

Proposed retrieval task



Goal: given a multilingual statement describing a spatial user need (topic), find as many relevant documents or other entities as possible from all target document collections.

Topics: 25 textual descriptions in a range of languages including English, Spanish, Portuguese and German. Topics will be structured in the form of: <concept> <spatial relation> <region> (e.g. "find stories about disasters in Geneva"). Spatial relations can include "near to", "within X miles of", "north of", "south of" etc. Topics will also include a short description (title, description) and a longer narrative describing relevance. In addition we may provide a spatial footprint for the query.

Spatial analysis: not required to participate in this task, but can augment text-based retrieval methods.

Collections: English and Portuguese (and possibly German or Spanish) newswire stories from CLEF. These include a variety of topics and geographical regions from news stories between 1994 and 1995. We may experiment with CLEF Web and Question Answering collections

Tasks/Challenges: translating locations, ambiguity of geo-references (e.g. "Jack London" the author not place; South Yorkshire and S. Yorks. refer to the same place), spatial ambiguity (e.g. Sheffield in UK or USA), finding/creating suitable multilingual gazetteer lists and combining both text and spatial retrieval methods. A geographic named entity extraction task may be devised.

Aims: to compare methods of query translation, query expansion, translation of geographical references, use of text and spatial retrieval methods separately or combined, retrieval models and indexing methods.

Organisers

Paul Clough and Mark Sanderson
p.d.clough, m.sanderson@sheffield.ac.uk
 University of Sheffield, UK
Hideo Joho
hideo@dcs.gla.ac.uk
 University of Glasgow, UK



Fred Gey
gey@uclink.berkeley.edu
 University of California, Berkeley, USA
Ray Larson
ray@simms.berkeley.edu
 University of California, Berkeley, USA

Registration

You can register for GeoCLEF 2006 by contacting Carol Peters (carol.peters@isti.cnr.it), the main co-ordinator for CLEF. For more specific information about any aspect of GeoCLEF or the tasks, please contact Hideo Joho (hideo@dcs.gla.ac.uk) or Fred Gey (gey@uclink.berkeley.edu).

The GeoCLEF 2006 website
<http://ir.shef.ac.uk/geoclef2006/index.html>

will be the main source of information about the tasks and through which the data, guidelines, resources and further information will be published. We also have a GeoCLEF mailing list (geoclef@sheffield.ac.uk) to provide participants with track information. Contact Hideo Joho to be added to this list. More information about CLEF can be found at <http://www.clef-campaign.org>

Schedule for CLEF 2006

Registration opens	15 January 2006
Data release	from February 2006
Topics release	from 15 March 2006
Runs submitted	15 May 2006
Release of results	from 15 July 2006
Workshop papers deadline	15 August 2006
Workshop	20-22 September 2006