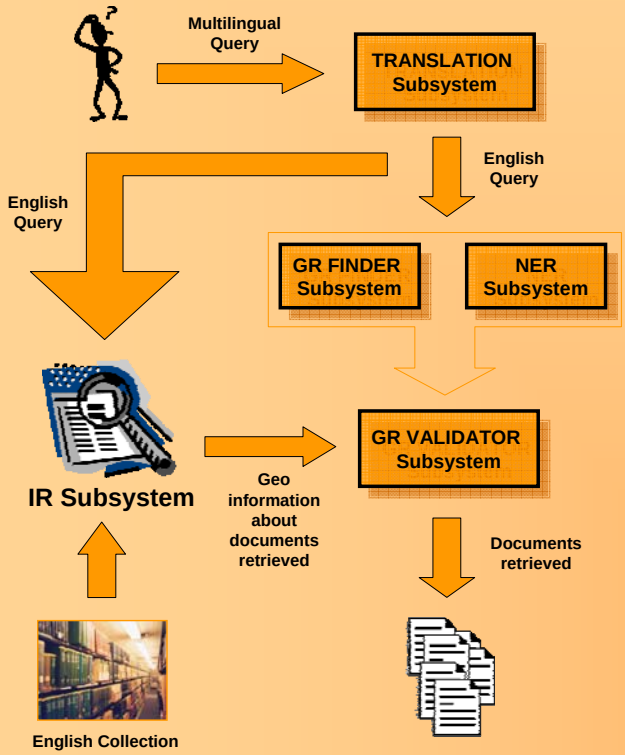
GeoUJA System. The University of Jaén at GeoCLEF 2007

José Manuel Perea-Ortega, Miguel Ángel García-Cumbreras, Manuel García-Vega, Arturo Montejo-Ráez

GeoUJA System



GeoUJA: A Multilingual Geographical IR System

- Second participation: monolingual (EN-EN) and bilingual (DE-EN, PT-EN, SP-EN).
- Different system from the presented in 2006: Query expansion vs. Filtering of relevant documents
- English collections.
- Pre-processing (stop words, Named Entity Recognizer and the Porter stemmer)
- Queries labelled (entities and geo-relation Information).
- The heuristics applied are quite restrictive.
- Some rules have to be improved and new rules must be created.

Subsystems Description

1. **Translation**
   - SINTRAM (SINai TRAnslation Module), an own translation module that works with several machine online translations, dictionaries and implements some heuristics.
2. **Information Retrieval**
   - Lemur (http://www.lemurproject.org).
3. **Geo-Relation Finder**
   - Detection of spatial relations in queries.
4. **Name Entity Recognizer**
   - Detection of the geographical places.
   - NER and gazetteer included in GATE (http://gate.ac.uk).
5. **Geo-Relation Validator**
   - Discriminates what documents among the recovered ones by the IR Subsystem are valid.
   - Uses the Geonames Gazetteer (http://www.geonames.org).

Best results

EXPERIMENT	MEAN AVERAGE PRECISION	R-PRECISION
Monolingual Okapi with FB	0.2605	0.2636
Monolingual Okapi without FB	0.2486	0.2624
Bilingual Okapi with FB	0.2362	0.2238
Bilingual Okapi without FB	0.2310	0.2476

Conclusions and Future work

- GR Validator Subsystem has to be improved.
- To add more heuristics and the rules applied must be more precise.
- The IR Subsystem would have to return more documents so that the GR Validator Subsystem has more variety of documents to verify.

SINAI
sistemas inteligentes de acceso a la información