GeoCLEF 2007: the CLEF 2007 Cross-Language Geographic Information Retrieval Track Overview

Thomas Mandl, Fredric Gey, Giorgio Di Nunzio, Nina Hau1.
Ray Lewer, Mark Sanderson1, Diana Santos, Cristina Womser-Hacker, Xing Xie1

1University of California Berkeley, CA, USA: gey@ics.berkeley.edu, ray@ics.berkeley.edu
2Department of Information Studies, University of Sheffield, UK, c.womser@sheffield.ac.uk
3Institute of Informatics, University of Padua, Italy: gdiunzio@info.unipd.it
4Institute for Information Systems, University of Hildesheim, Germany: diana.santos@sinif.ict.ntnu.no

CLEF 2007, Budapest, September 19-21, 2007

Overview

More on the geographic search task ...

- Thomas Mandl: Overview
- Query Classification Subtask
- Mark Sanderson: Topic Creation and Relevance Assessment
- Giorgio Di Nunzio: Results
- Diana Santos: Approaches and Interpretation

Initial Aim of GeoCLEF

- Aim: to evaluate retrieval of multilingual documents with an emphasis on geographic search (GIR)
  - "find me news stories about riots near Dublin"

(Fredric Gey @ CLEF Workshop 2005)

GeoCLEF Administration

- Joint effort of
  - Fredric Gey (U. California at Berkeley)
  - Diana Santos (Linguateca, SINTEF ICT, Norway)
  - Mark Sanderson (U. Sheffield)
  - Nicola Ferri, Giorgio Di Nunzio (U. Padua)
  - Xing Xie (Microsoft Research, Asia)
  - Thomas Mandl, Christa Womser-Hacker (U. Hildesheim)
  - and many others ...

Participation

<table>
<thead>
<tr>
<th>CLEF Year</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nr. of Participants</td>
<td>11</td>
<td>17</td>
<td>13</td>
</tr>
<tr>
<td>Nr. of submitted Experiments</td>
<td>117</td>
<td>149</td>
<td>108</td>
</tr>
</tbody>
</table>
Search Task 2007

- Three languages
- 600,000 docs
- 25 topics (75 in three years now)

- Intention behind topics
  - geographically challenging

Reliability?

- 25 topics are sufficient under most circumstances to reliably order systems (Sanderson & Zobel 2005)

Partial Swap Rate Analysis

Average Correlation between system rankings of full and partial topic set for German Mono-lingual task 2006

Search Task

- How much and which geo knowledge and reasoning is necessary?
- Each year, keyword based systems do well on the task

Query Classification Task

- Goal: find geo queries in a log of real queries
- New in 2007
- Organized by Xing Xie (Microsoft Research Asia, Beijing, China)

Data

- Query log from the MSN search engine
  - In English
  - 800,000 queries (collected August 2006)
  - 500 queries were labelled and used for evaluation
    - 100 queries for training
    - 400 for testing

Intention behind topics

- Geographically challenging
Task

- Find queries with a geographic scope
  - Find what component
  - Find geolocation-type
  - Find where component
  - Classify what type (information, yellow page, map)

Example:

```
local:YES
Lottery in Florida
WHERE:Florida, US
```

Geo-relation-types

- 27 classes
- Examples:
  - In
  - On
  - Near
  - Along
  - Distance
  - North_of
  - North_west_of
  - North_to
  - ...

Evaluation Set

1. Choose 800 queries randomly from the query set.
2. Remove the typos and the ambiguous queries from the 800 ones manually.
3. Select the queries with special geo-relation from the remainder queries in the query set manually and add them to the evaluation set.
4. Select 500 queries for the final evaluation set.

Query Types in Evaluation Set

- Map 16%
- Non-local 36%
- Information 19%
- Yellow page 29%

Evaluation Metrics

- Three assessors
  - Individually assessed all system answers
  - Reached an agreement
  - Fully correct classified query instances
  - Recall, precision, and combined F1-Score

Results

<table>
<thead>
<tr>
<th>Team</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ask</td>
<td>0.825</td>
<td>0.258</td>
<td>0.365</td>
</tr>
<tr>
<td>Clear</td>
<td>0.201</td>
<td>0.157</td>
<td>0.198</td>
</tr>
<tr>
<td>Lingu</td>
<td>0.112</td>
<td>0.038</td>
<td>0.077</td>
</tr>
<tr>
<td>Mirac</td>
<td>0.428</td>
<td>0.546</td>
<td>0.468</td>
</tr>
<tr>
<td>Tapj</td>
<td>0.222</td>
<td>0.249</td>
<td>0.235</td>
</tr>
<tr>
<td>Tids</td>
<td>0.096</td>
<td>0.158</td>
<td>0.168</td>
</tr>
</tbody>
</table>
Approaches

- Gazeteers for location identification
  - Large base of geo names
- Pre-defined Rules

- Issues
  - Low Performance
  - Few training classes for many geo-types

Topics

- 25 adhoc topics
- Developed
  - One third in English
  - One third in German
  - One third in Portuguese

geoCLEF topic creation

Mark Sanderson

Classic CLEF topic creation

- Developed topics in local language
- Other geoCLEF partners translated topics and checked for relevance in their collection.

Motivation behind topic design

- Text only often wins
- How to tackle that
  - Imprecise regions
  - Regions surrounding, but not including a point
  - Regions where local knowledge is important

Imprecise regions

- "Documents describing the damage caused by acid rain in the countries of northern Europe"
Participation (2/2)

<table>
<thead>
<tr>
<th>Track</th>
<th>Source Language</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilingual X2DE</td>
<td>DE 1</td>
<td>6 1 1</td>
</tr>
<tr>
<td>Bilingual X2EN</td>
<td>EN 1 5 6 1</td>
<td>13</td>
</tr>
<tr>
<td>Bilingual X2PT</td>
<td>PT 1 1 5</td>
<td>7</td>
</tr>
<tr>
<td>TOTAL</td>
<td>10 7 11 6</td>
<td>28</td>
</tr>
</tbody>
</table>

Participation (1/2)

- **Problems**
  - Always hard to find topics that work well across languages

- **Assessment - DIRECT**
  - Mostly volunteers
    - Hildesheim
    - SINTEF
    - Fred in Berkeley
  - Some funding
    - Sheffield - Tripod project

- **Participation**
  - **Participation (1/2)**
    - Cristie
    - Checkmate
    - CLEF
    - Hildesheim
    - L2
    - LINGUIST
    - M2L
    - NLP
    - ES
    - TOTAL

  - **Participation (2/2)**
    - | Track | Source Language | TOTAL |
    - | Bilingual X2DE | DE 1 | 6 1 1 | 8 |
    - | Bilingual X2EN | EN 1 5 6 1 | 13 |
    - | Bilingual X2PT | PT 1 1 5 | 7 |
    - | TOTAL | 10 7 11 6 | 28 |
### Monolingual Tasks

- **English**
- **German**
- **Portuguese**

### Monolingual English

<table>
<thead>
<tr>
<th>Track</th>
<th>Rank</th>
<th>Part</th>
<th>Experiment DOI</th>
<th>MAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monolingual English</td>
<td>1</td>
<td>English</td>
<td>txt/light Monterrey-SAMT/Mono-English-Short-1006006-05-02-2006</td>
<td>25.15%</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>English</td>
<td>txt/light Monterrey-SAMT/Mono-English-Short-1006006-05-02-2006</td>
<td>24.10%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>English</td>
<td>txt/light Monterrey-SAMT/Mono-English-Short-1006006-05-02-2006</td>
<td>23.42%</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>English</td>
<td>txt/light Monterrey-SAMT/Mono-English-Short-1006006-05-02-2006</td>
<td>22.49%</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>English</td>
<td>txt/light Monterrey-SAMT/Mono-English-Short-1006006-05-02-2006</td>
<td>21.56%</td>
</tr>
</tbody>
</table>

### Monolingual German

<table>
<thead>
<tr>
<th>Track</th>
<th>Rank</th>
<th>Part</th>
<th>Experiment DOI</th>
<th>MAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monolingual German</td>
<td>1</td>
<td>German</td>
<td>txt/light Monterrey-SAMT/Mono-German-Short-1006006-05-02-2006</td>
<td>26.42%</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>German</td>
<td>txt/light Monterrey-SAMT/Mono-German-Short-1006006-05-02-2006</td>
<td>25.15%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>German</td>
<td>txt/light Monterrey-SAMT/Mono-German-Short-1006006-05-02-2006</td>
<td>24.10%</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>German</td>
<td>txt/light Monterrey-SAMT/Mono-German-Short-1006006-05-02-2006</td>
<td>23.42%</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>German</td>
<td>txt/light Monterrey-SAMT/Mono-German-Short-1006006-05-02-2006</td>
<td>22.49%</td>
</tr>
</tbody>
</table>

### Monolingual Portuguese

<table>
<thead>
<tr>
<th>Track</th>
<th>Rank</th>
<th>Part</th>
<th>Experiment DOI</th>
<th>MAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monolingual Portuguese</td>
<td>1</td>
<td>Portuguese</td>
<td>txt/light Monterrey-SAMT/Mono-Portuguese-Short-1006006-05-02-2006</td>
<td>25.76%</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Portuguese</td>
<td>txt/light Monterrey-SAMT/Mono-Portuguese-Short-1006006-05-02-2006</td>
<td>24.10%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Portuguese</td>
<td>txt/light Monterrey-SAMT/Mono-Portuguese-Short-1006006-05-02-2006</td>
<td>23.42%</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Portuguese</td>
<td>txt/light Monterrey-SAMT/Mono-Portuguese-Short-1006006-05-02-2006</td>
<td>22.49%</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Portuguese</td>
<td>txt/light Monterrey-SAMT/Mono-Portuguese-Short-1006006-05-02-2006</td>
<td>21.56%</td>
</tr>
</tbody>
</table>
Bilingual Tasks

- X -> English
- X -> German
- X -> Portuguese

Monolingual Portuguese

<table>
<thead>
<tr>
<th>Track</th>
<th>Rank</th>
<th>Part</th>
<th>Experiment ID</th>
<th>MAP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bilingual X -> English

Bilingual X -> German

Bilingual X -> EN

Bilingual X -> German

System with different modules
Geographic Ontology + Text Mining
MOAD + mixed Okapi BM25

T. Mandl et al.
Bilingual X-> Portuguese

More Analyses

Monolingual vs Bilingual Analyses
- Mean of average precision for each topic of a task
- Median of average precision for each topic of a task

Monolingual Mean VS Bilingual Mean German
Overview of GeoCLEF 2007

- IR techniques
- IE/NLP techniques
- GIR techniques

- Systems
- Resources
- Experiments
- Translation
- General comments

IR techniques and systems

- automatic and manual query expansion
- blind relevance feedback
- INL2 term weighting model
- divergence from randomness framework
- latent Dirichlet allocation model
- logistic regression
- query decomposition
- vector space model
- stemming
- systems: Lemur (2), Lucene (4), MG4J, Terrier (2), Zebra DBMS

NLP techniques and systems

- named entity recognition
- WordNet-based expansion
- semantic analysis
- part-of-speech tagging
- use of a QA system for subqueries
- decompounding (for German)
- WordNet lemmatizer
- systems: LingPipe, Annie (GATE)

GIR/GIE techniques

- location disambiguation
- geographic unique strings
- query expansion based on geographical terms
- query expansion based on a geographic ontology
- heuristic geographically informed filtering
  - removing candidates
  - using shape files for close or near geographical relations
- separate geographical indexes
- geographic cooccurrence model (based on Wikipedia)
- geographic relation finder

GIR/GIE resources

- geonames gazetteer
- World Gazetteer
- Getty TGN
- GNIS
- GeoWorldMap World gazetteer
- ADL Feature Type thesaurus
- GKB 2.0
- Spanish toponymy list
- plus Wikipedia and WordNet
Translation (only queries)

- Systran (PT-EN)
- Prompt (ES-EN)
- Promt (EN-DE)
- LEC Power Translator (all possible pairs)
- transfer MT system (ES-EN, ES-PT)
- Togletext (ID-EN)

Experiments

- In addition to the particular experiments of each group concerning their original approaches
- Separate indexes for geographic and non-geographic information
- Apply different techniques to different parts of the query
- Use T, TD or TDN

Comments

- Very few papers mention distinguished treatment for different kinds of topics
  - although some discuss differences between topics
  - only one provides some per-topic analysis
- Most participants have best results with text only!
- Most participants use NER,
  - but no NER evaluation in itself has been reported: does it really help? or does it also diminish performance?
  - they don’t use the output of NER the same way: wide difference in behaviour after NER has been invoked