DELOS
Network of Excellence on Digital Libraries

CLEF

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

Thomas Mandl[3], Fredric Gey[1], Giorgio Di Nunzio[6], Nicola Ferro[6],
Ray Larson[1], Mark Sanderson[2], Diana Santos[4], Christa Womser-Hacker[3], Xing Xie[5]

[1]University of California, Berkeley, CA, USA gey@berkeley.edu, ray@sims.berkeley.edu
[2]Department of Information Studies, University of Sheffield, Sheffield, UK, m.sanderson@sheffield.ac.uk
[3]Information Science, University of Hildesheim, GERMANY, {mandl|womser}@uni-hildesheim.de
[4]Linguateca, SINTEF ICT, NORWAY, Diana.Santos@sintef.no
[5]Microsoft Research Asia, Beijing, China, Xingx@microsoft.com
[6]Department of Information Engineering, University of Padua, Italy, {dinunzio|ferro}@dei.unipd.it

GeoCLEF Administration

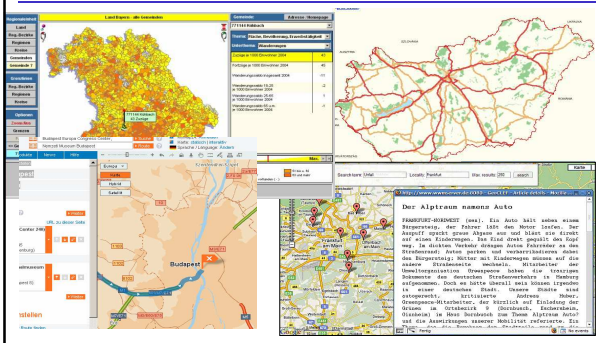
- Joint effort of
  - Fredric Gey (U. California at Berkeley)
  - Diana Santos (Linguateca, SINTEF ICT, Norway)
  - Mark Sanderson (U. Sheffield)
  - Nicola Ferro, Giorgio Di Nunzio (U. Padua)
  - Xing Xie (Microsoft Research, Asia)
  - Thomas Mandl, Christa Womser-Hacker (U. Hildesheim)
  - and many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

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*

Geographic Information Systems



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*"

(Fred Gey @ CLEF Workshop 2005)

Participation

CLEF Year	2005	2006	2007
Nr. of Participants	11	17	13
Nr. of submitted Experiments	117	149	108

New: Query Classification Task
6 participants

Search Task 2007

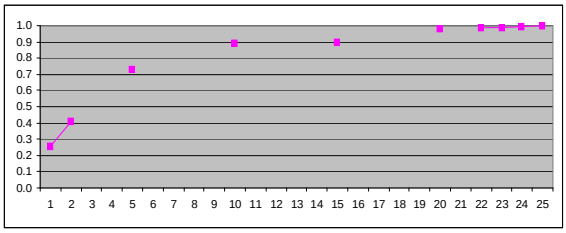
- Three languages
- 600,000 + docs
- 25 topics (75 in three years now)

- 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

Task

- Find queries with a geographic scope
  - Extract where component
  - Extract geo-relation-type
  - Extract what component
  - Classify what type {information, yellow page, map}

Example:

<local>YES

Lottery in Florida

< what>lottery
<what-type>information

<where>Florida, US

< geo-relation>in

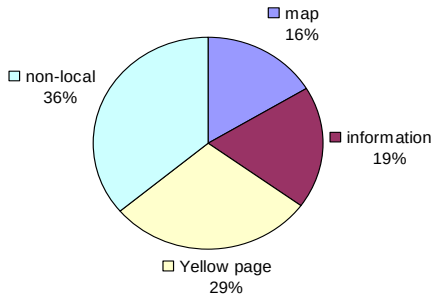
Geo-relation-types

- 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-relations 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



Evaluation Metrics

- Three assessors
  - individually assessed all system answers
  - reached an agreement
- Fully Correct classified query instances
- Recall, precision and combined F1-Score

Results

Team	Precision	Recall	F1
Ask	**0.625**	0.258	0.365
Csusm	0.201	0.197	0.199
Linguit	0.112	0.038	0.057
Miracle	0.428	**0.566**	**0.488**
Talp	0.222	0.249	0.235
Xldb	0.096	0.08	0.088

Approaches

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

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

geoCLEF topic creation

Mark Sanderson

Topics

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

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"

Surrounding…

- "Find information about social problems afflicting places in greater Lisbon."
- "Find documents mentioning airplane crashes close to Russian cities"

Local knowledge

- "To be relevant, a document must describe a whisky made, or a whisky distillery located, on a Scottish island."

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 (1/2)

Participant	Monolingual Tasks			Bilingual Tasks			TOTAL
	DE	EN	PT	X2DE	X2EN	X2PT	
catalunya		5					5
cheshire	1	1	1	3	3	3	12
csusm	6	6	5		4	4	25
depok*					6		6
groningen		5					5
hagen	5			5			10
hildesheim	4	4					8
icl		4					4
linguit*		4					4
moscow*		2					2
msasia		5					5
valencia		12					12
xldb		5	5				10
TOTAL	**16**	**53**	**11**	**8**	**13**	**7**	**108**

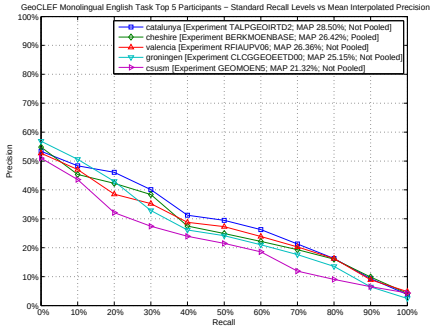
Participation (2/2)

Track	Source Language					TOTAL
	DE	EN	ES	ID	PT	
Bilingual X2DE		6	1		1	**8**
Bilingual X2EN	1		5	6	1	**13**
Bilingual X2PT	1	1	5			**7**
TOTAL	**2**	**7**	**11**	**6**	**2**	**28**

Monolingual Tasks

- English
- German
- Portuguese

Monolingual English

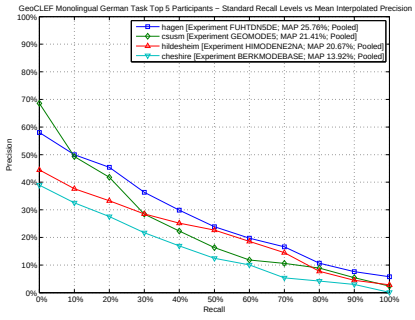


Monolingual English

Track	Rank	Part.	Experiment DOI	MAP
Monolingual English	1st	catalunya	10.2415/GC-MONO-EN-CLEF2007.CATALUNYA.TALPGEOIRTD2	28.50%
	2nd	cheshire	10.2415/GC-MONO-EN-CLEF2007.CHESHIRE.BERKMOENBASE	26.42%
	3rd	valencia	10.2415/GC-MONO-EN-CLEF2007.VALENCIA.RFIAUPV06	26.36%
	4th	groningen	10.2415/GC-MONO-EN-CLEF2007.GRONINGEN.CLCGGEOEETD00	25.15%
	5th	csusm	10.2415/GC-MONO-EN-CLEF2007.CSUSM.GEOMOEN5	21.32%
	Diff.			33,68%

POS, lemmas, named entities
WordNet
LingPipe
World Gazetteer, GeoNET, Wikipedia, WordNet
Ling Pipe
mixed tfxidf (Lucene)

Monolingual German

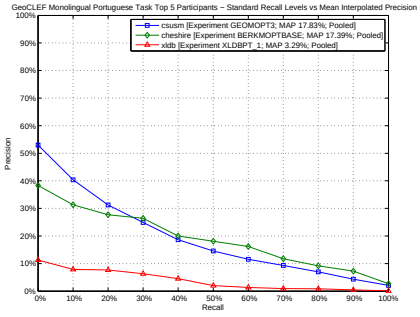


Monolingual German

Track	Rank	Part.	Experiment DOI	MAP
Monolingual German	1st	hagen	10.2415/GC-MONO-DE-CLEF2007.HAGEN.FUHTDN5DE	25.76%
	2nd	csusm	10.2415/GC-MONO-DE-CLEF2007.CSUSM.GEOMODE4	21.41%
	3rd	hildesheim	10.2415/GC-MONO-DE-CLEF2007.HILDESHEIM.HIMODENE2NA	20.67%
	4th	cheshire	10.2415/GC-MONO-DE-CLEF2007.CHESHIRE.BERKMODEBASE	13.92%
	5th			
	Diff.			85.06%

Semantic analysis for GIR
tfxidf (Zebra DBMS)
GeoNet name server
BM25, LM, InL2 (Terrier)
LingPipe
tfxidf (Lucene)

Monolingual Portuguese



Monolingual Portuguese

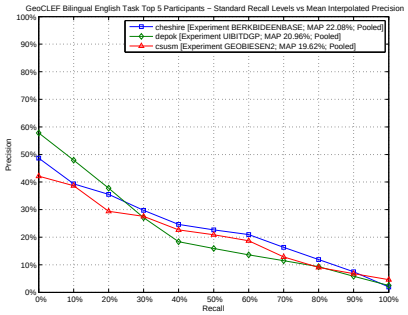
Track	Rank	Part.	Experiment DOI	MAP
Monolingual Portuguese	1st	csusm	10.2415/GC-MONO-PT-CLEF2007.CSUSM.GEOMOPT3	17.83%
	2nd	cheshire	10.2415/GC-MONO-PT-CLEF2007.CHESHIRE.BERKMOPTBASE	17.39%
	3rd	xldb	10.2415/GC-MONO-PT-CLEF2007.XLDB.XLDBPT_1	3.29%
	4th			
	5th			
	Diff.			441.95%

System with different modules
Geographic Ontology + Text Mining
MG4J + mixed Okapi BM25

Bilingual Tasks

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

Bilingual X->English

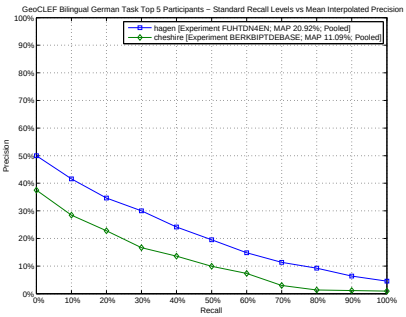


Bilingual X->EN

Track	Rank	Part.	Experiment DOI	MAP
Bilingual English	1st	cheshire	10.2415/GC-BILI-X2EN-CLEF2007.CHESHIRE.BERKBIDEENBASE	22.08%
	2nd	depok*	10.2415/GC-BILI-X2EN-CLEF2007.DEPOK.UIBITDGP	20.96%
	3rd	csusm	10.2415/GC-BILI-X2EN-CLEF2007.CSUSM.GEOBIESEN2	19.62%
	4th			
	5th			
	Diff.			12.54%

LEC Power Translator
World Gazetteer
LR
Free MT tool
Geonames + Wikipedia
LM (Lemur)
Spanish Toponomy from Euro Parl.
GeoNet name server
BM25, LM, InL2 (Terrier)

Bilingual X-> German

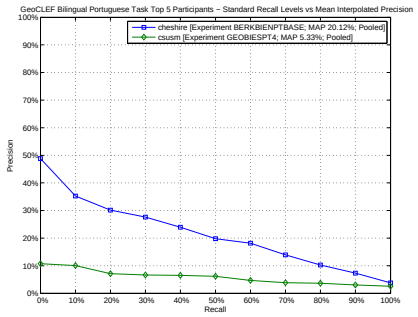


Bilingual X-> German

Track	Rank	Part.	Experiment DOI	MAP
Bilingual German	1st	hagen	10.2415/GC-BILI-X2DE-CLEF2007.HAGEN.FUHTDN4EN	20.92%
	2nd	cheshire	10.2415/GC-BILI-X2DE-CLEF2007.CHESHIRE.BERKBIPTDEBASE	11.09%
	3rd			
	4th			
	5th			
	Diff.			88,64%

Promt Web service MT
Semantic analysis for GIR
tfxidf (Zebra DBMS)

Bilingual X-> Portuguese



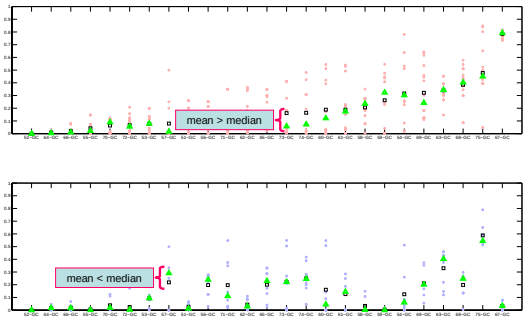
Bilingual X-> Portuguese

Track	Rank	Part.	Experiment DOI	MAP
Bilingual Portuguese	1st	cheshire	10.2415/GC-BILI-X2PT-CLEF2007.CHESHIRE.BERKBIENPTBASE	20.12%
	2nd	csusm	10.2415/GC-BILI-X2PT-CLEF2007.CSUSM.GEOBIESPT4	5.33%
	3rd			
	4th			
	5th			
	Diff.			277.49%

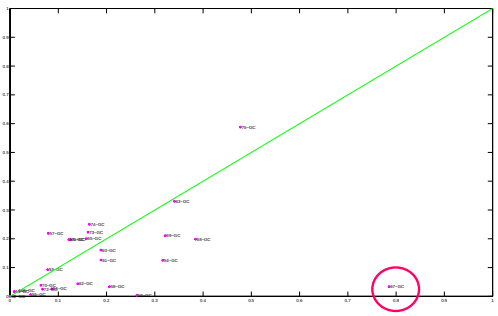
More Analyses

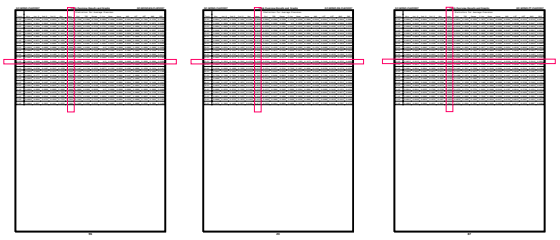
- Monolingual vs Bilingual Analyses
  - Mean of average precision for each topic of a task
  - Median of average precision for each topic of a task

Mean vs Median (mono, bili German)



Monolingual Mean VS Bilingual Mean German





Topic 10.2452/GC-67		
English (17 rel docs)	German (61 rel docs)	Portuguese (75 rel docs)
28.36	78.54	22.70

<top lang="en">
<num>10.2452/67-GC</num>
<title>F1 circuits where Ayrton Senna competed in 1994</title>
<desc>Find documents that mention circuits where the Brazilian driver Ayrton Senna participated in 1994. The name and location of the circuit is required</desc>
<narr>Documents should indicate that Ayrton Senna participated in a race in a particular stadion, and the location of the race track.</narr>
</top>

- <top lang="de">
<num>10.2452/67-GC</num>
<title>Formel-1-Rennstrecken, auf denen Ayrton Senna 1994 gefahren ist</title>
<desc>Dokumente, die Rennstrecken erwähnen, auf denen der Brasilianische Fahrer Ayrton Senna 1994 gefahren ist. Name und Ort müssen erwähnt sein.</desc>
<narr>Dokumente sollten angeben, dass Ayrton Senna an einem Rennen auf einer bestimmten Strecke teilgenommen hat. Der Ort der Strecke sollte genannt sein.</narr>
</top>

- <top lang="pt">
<num>10.2452/67-GC</num>
<title>Pistas em que Ayrton Senna correu em 1994</title>
<desc>Em que circuitos de fórmula 1 Ayrton Senna competiuem 1994?</desc>
<narr>Documentos que indiquem que Ayrton Senna participou em corridas num determinado autódromo, em que a localização deste é também mencionada.</narr>
</top>

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
- location normalization
- 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)
- Toggletext (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