Cross-Language Evaluation Forum

CLEF Workshop 2006

Poster Boaster Session

Cross Language search on the WEB and Digital Libraries

Luca Dini

Paolo Curtoni

dini/curtoni@celi.it

CELI-Torino

www.celi.it

Ad-hoc Clef – Ad-hoc ImageClef

Challenges

- Challenges
  - Absence of a consolidated corpus
  - Absence of a precise domain
  - Presence of very technical query/docs
  - Absence of style constraints and document formats (WEB only).
- Approach (IT-EN):
  - Query translation;
  - Huge dictionaries;
  - Query disambiguation mechanisms.

Results

CELIdescNOEXPANSIONboost	NUM_RELEV	MEAN_AVE	R_PRECISION	PRECISION_AT_10_	PRECISION_AT_20_R
CELItitleNOEXPANSION	773	0,2397	0,2381	0,2400	0,1890
CELIdescNOEXPANSION	764	0,2268	0,2381	0,2320	0,1720
CELItitleLisaExpansion	814	0,2238	0,2212	0,2160	0,1720
CELItitleCwnCascadeExpansion	673	0,2390	0,2400	0,2020	0,1650
CELItitleCwnExpansion	636	0,2110	0,2074	0,1980	0,1520
CELItitleNOEXPANSIONboost	680	0,2035	0,2036	0,1900	0,1430
CELIdescLisaExpansion	793	0,1941	0,2016	0,1840	0,1470
CELIdescCwnExpansion	594	0,1792	0,1900	0,1760	0,1460
CELIdescCwnCascadeExpansion	602	0,1957	0,1982	0,1720	0,1380
CELIdescLisaExpansionboost	767	0,1908	0,1966	0,1660	0,1370
CELIdescNOEXPANSIONboost	733	0,1812	0,1811	0,1600	0,1300
CELItitleLisaExpansionboost	712	0,1732	0,1795	0,1600	0,1370

System	**REL_RETR**	**MAP**	**P10**	**P20**
IT-EN-AUTO-NOFB-TXT-CELI-AND_CwnExpansion	1474	0,1392	0,2233	0,2000
IT-EN-AUTO-NOFB-TXT-CELI-AND_NOEXPANSION	1551	0,1486	0,2150	0,2092
IT-EN-AUTO-NOFB-TXT-CELI-OR_NOEXPANSION	1551	0,1441	0,2033	0,1975
IT-EN-AUTO-NOFB-TXT-CELI-AND_CwnCascadeExpansion	1590	0,1362	0,1917	0,1817
IT-EN-AUTO-NOFB-TXT-CELI-OR_CwnExpansion	1474	0,1284	0,1917	0,1775
IT-EN-AUTO-NOFB-TXT-CELI-AND_LisaExpansion	1655	0,1237	0,1867	0,1725
IT-EN-AUTO-NOFB-TXT-CELI-OR_LisaExpansion	1655	0,1189	0,1633	0,1558
IT-EN-AUTO-NOFB-TXT-CELI-OR_CwnCascadeExpansion	1590	0,1208	0,1583	0,1492

Amharic – English Information Retrieval

Atelach Alemu Argaw
Lars Asker

Stockholm University

<num> **C346** </num>
<AMH-title>
yegrand slam axenafiwoc
</AMH-title>
<AMH-desc>
be1995 `a.m betederegew yetEnis grand slam wddr man axenefe?
</AMH-desc>

Grand Slam Winners

Who won a tennis Grand Slam Tournament event in 1995?

<num> C346 </num>
<AMH-title>
yegrand slam axenafiwoc
</AMH-title>
<AMH-desc>
be1995 `a.m betederegew yetEnis grand slam wddr man axenefe?
</AMH-desc>

stemming

POS tagging

Dictionary lookup

Query expansion

Term weighting

Approximate string matching

COLESIR at CLEF 2006: Rapid Prototyping of an *N*-gram-Based CLIR System

Jesús VILARES

Departamento de Computación
Universidade da Coruña
Spain

Michael P. OAKES
John I. TAIT

School of Computing and Technology
University of Sunderland
United Kingdom

Tracks: ad-hoc (Portuguese, robust)

N-gram-Based Translation

- Character *n*-grams not only as indexing units, but also as translating units.

- Based on the previous work of the Johns Hopkins University Applied Physics Lab.

- Freely available resources: TERRIER, Europarl, GIZA++.

- Tries to speed up the training process: 2 steps
    1) Word-level alignment: statistical model.
    2) *N*-gram-level alignment: association measures.

Advantages

- No word normalization.
- Out-of-vocabulary words.
- No language-specific processing.
- Faster n-gram alignment process.

Esfinge – a Modular Question Answering System for Portuguese

Luís Costa

(Linguateca/SINTEF)

Track(s):Multiple Language Question Answering

Esfinge@CLEF2006

- 3rd participation in QA@CLEF. Esfinge exploits the redundancy in the Web where Portuguese is one of the most used languages.
- New in 2006:

  - Made a fuller use of the capabilities of a named entity recognizer.

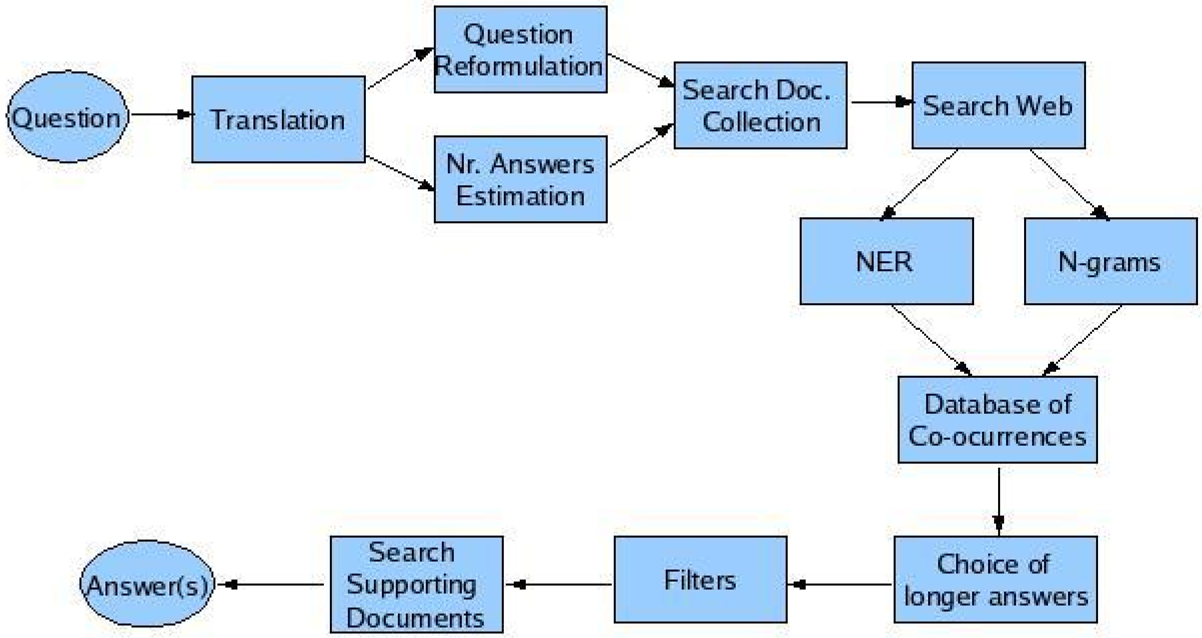
  - Used a database of word co-occurrences to re-rank the candidate answers.

  - Used a parser to estimate the number of answers to be returned.
- Some Empirical Results:

  - Compare the importance of using the Web as resource for the different types of questions.

  - Show why the Web helped more this year than in previous editions of CLEF.

System Architecture



Source code freely available, interface on the Web:
http://www.linguateca.pt/Esfinge/

The UPV at QA@CLEF 2006

Davide Buscaldi, José Manuel Gómez, Paolo Rosso, Emilio Sanchis

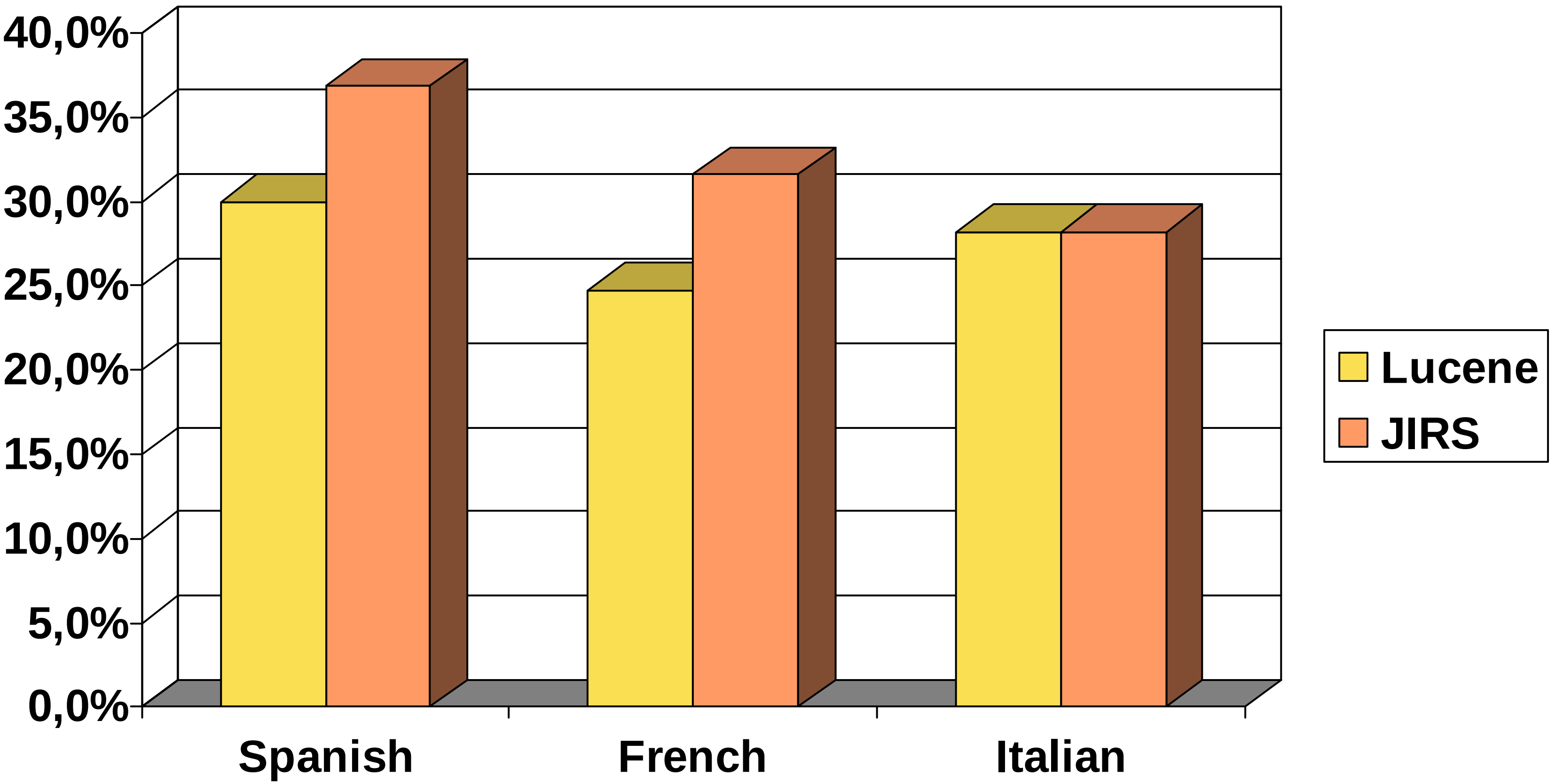
(***DSIC, Universidad Politécnica de Valencia***)

Track:QA

Comparison between search engines

- The same QA system, with two different search engines
  - JIRS and Lucene
- JIRS (Java Information Retrieval System) is a Passage Retrieval engine based on n-grams
- JIRS performed much better than Lucene (in Spanish and French)

JIRS vs. Lucene



Using Syntactic Knowledge for QA

Gosse Bouma, Ismail Fahmi, Jori Mur,
Gertjan van Noord, Lonneke van der Plas, Jörg Tiedemann
(University of Groningen)

Mono (nlnl) and multilingual (ennl) QA

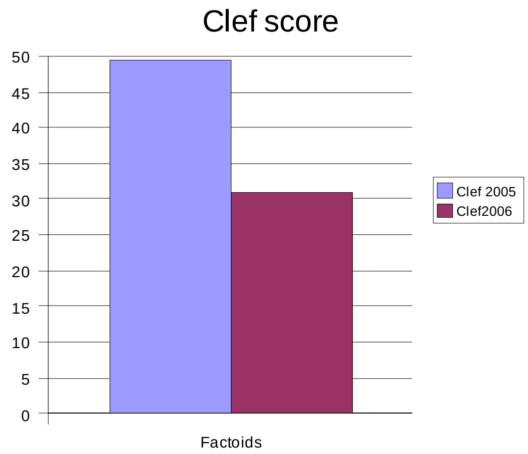
Joost

- Full syntactic analysis of Questions and Text collection
- Linguistically Informed IR (Lucene)
- Co-reference resolution for off-line extraction
- Lexical Equivalences (spelling, synonyms, adjectival, genitive forms, compounds)
- Definition questions use Wikipedia information
- Multilingual QA using Babelfish and Wikipedia cross-language links (for NEs)

Dutch as Target

"Annotators were explicitly asked to think of harder questions, involving paraphrases and some limited general reasoning"

(Clef 2006 WorkingNotes)



Hunting answers with RAPOSA (FOX)

Luís Sarmento

Universidade do Porto & Linguateca (Porto)

Q A - Monolingual Portuguese

RAPOSA: what is it?…

- Early Prototype of a QA system for Portuguese
  - First attempt at CLEF
- Strongly based on our NER system: SIEMÊS
  - Shallow parsing techniques
- it assumes that answers are “name- entities”. BUT:
  - 100 different types of NE are considered
  - NE related concepts are also included (“ergonyms”)
- Currently it only addresses factoid questions!!

But… why should you even look at the poster?

- Yes:
  - RAPOSA only answers a few types of questions
  - results ARE below average!
  - the system IS very simple and inefficient!
- BUT you will see
  - how a simple semantic type checking strategy gives better results than a more complex context-checking one
  - many ideas (and little exciting experiments!) about how RAPOSA (FOX) could be taught to be a better answer hunter!
  - how we were able to improve our NER because of QA
- SO:
  - Don’t miss RAPOSA!

Overview of QA 2006 with Spanish as Target

Valentín Sama, Álvaro Rodrigo, Anselmo Peñas and Felisa Verdejo

(UNED)

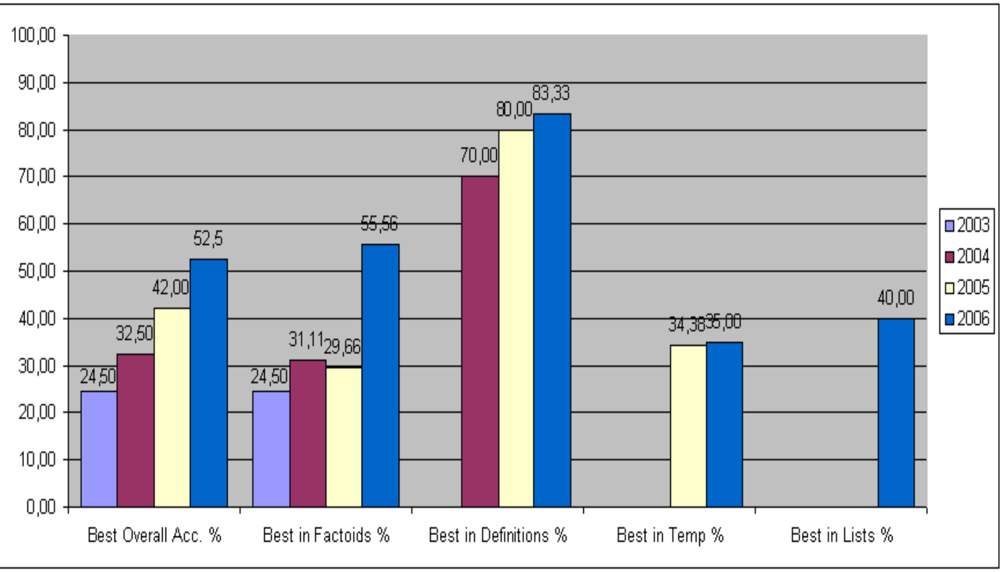
Track(s): Question Answering main task

Overview of QA 2006 with Spanish as Target

- 9 groups
- 17 runs: 12 monolingual + 5 bilingual
- Best bilingual comparable to monolingual
- 81% of the questions were answered by at least one system
- They do it better and better

new

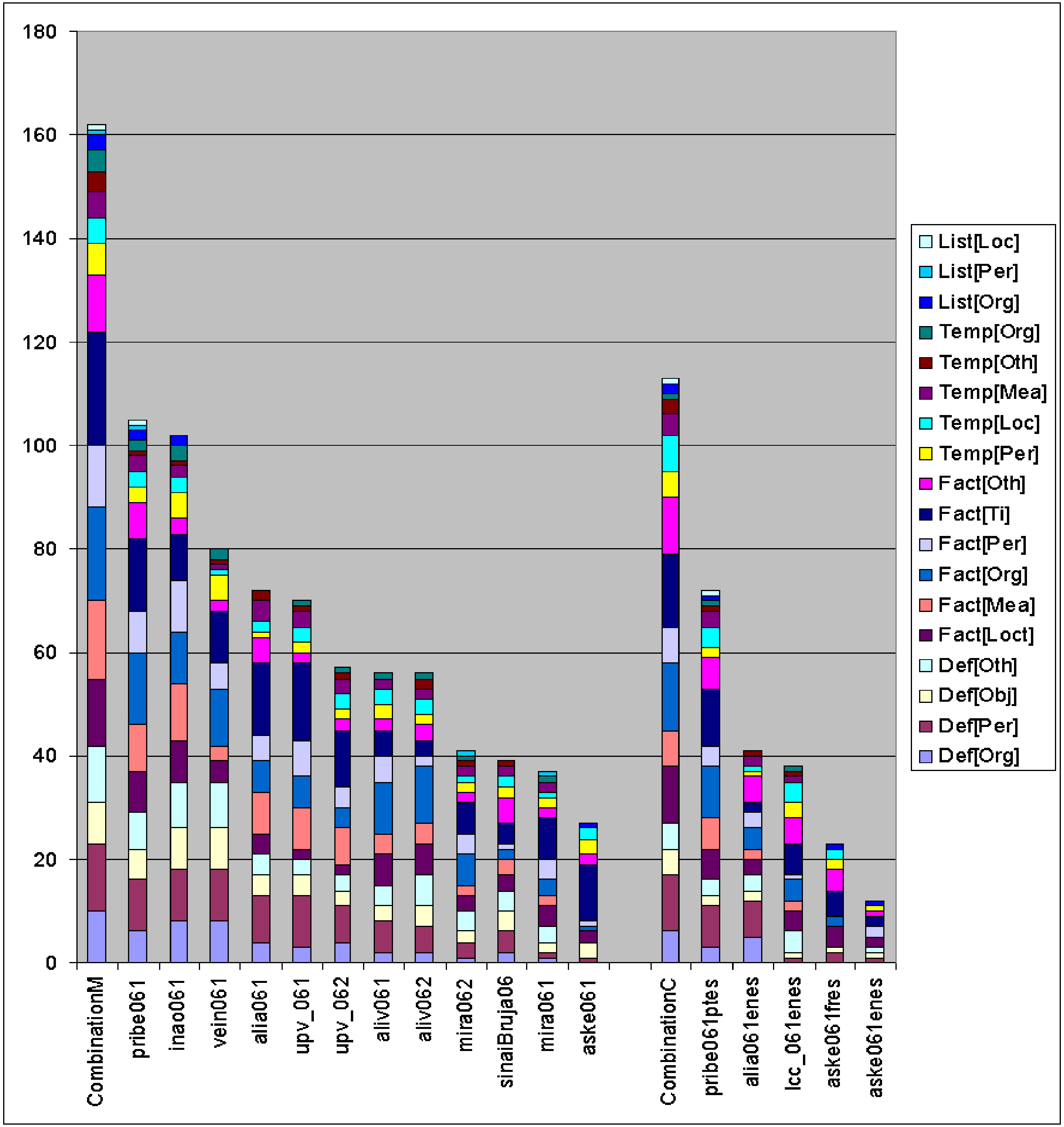
Answer Extraction



run	Right #	Right %	W #	X #	U #	Right % F [108]	Right % T [40]	Right % D [42]	Right % L [10]	NIL [20] F	NIL [20] P	NIL [20] R	r	Answer Extraction %
pribe061	**105**	**52,50**	86	4	5	**55,56**	30,00	69,05	**40,00**	0,44	0,34	0,60	-	**84,68**
inao061	**102**	51,00	86	3	9	47,22	**35,00**	**83,33**	20,00	**0,46**	0,38	0,60	0,216	**86,44**
vein061	**80**	40,00	112	3	5	32,41	25,00	**83,33**	-	0,34	0,21	0,80	0,133	**86,02**
alia061	**72**	36,00	105	15	8	38,89	22,50	50,00	-	0,34	0,22	0,75	**0,322**	69,23
upv_061	**70**	35,00	119	5	6	37,04	25,00	47,62	-	0,43	0,33	0,65	0,194	70,71
upv_062	57	28,50	123	6	14	27,78	25,00	40,48	-	0,41	0,32	0,60	0,163	66,28
aliv061	56	28,00	123	8	13	29,63	22,50	35,71	-	0,34	0,33	0,35	0,190	65,12
aliv062	56	28,00	132	6	6	26,85	25,00	40,48	-	0,33	0,26	0,45	0,153	72,73
mira062	41	20,50	148	4	7	21,30	17,50	23,81	10,00	0,35	0,35	0,35	0,145	43,62
sinaiBruja06	39	19,50	145	6	9	16,67	17,50	33,33	-	0,23	0,13	0,90	-0,119	**79,59**
mira061	37	18,50	154	3	6	21,30	15,00	16,67	10,00	0,34	0,26	0,50	0,136	51,39
aske061	27	13,50	143	1	29	15,74	12,50	9,52	10,00	0,08	0,20	0,05	0,199	62,79
pribe061ptes	**72**	**36,00**	123	3	2	**39,81**	**27,50**	38,10	20	0,29	0,29	0,30	-	**78,26**
alia061enes	41	20,50	134	9	16	17,59	12,50	**40,48**	-	0,31	0,19	0,80	0,142	65,08
lcc_061enes	38	19,00	141	14	7	20,37	25,00	14,29	-	**0,35**	0,35	0,35	0,067	55,07
aske061fres	23	11,50	162	-	15	13,89	10,00	7,14	10	0,08	0,17	0,05	**0,302**	53,49
aske061enes	12	6,00	178	-	10	6,48	2,50	7,14	10	0,10	1,00	0,05	0,091	40,00

Genomics

Meeting point



Experimenting a “general purpose” textual entailment learner in AVE

Fabio Massimo Zanzotto

Alessandro Moschitti

University of Rome “Tor Vergata”

Italy

Track: Question Answering, Answer Validation Exercise

Motivating the approach (1)

$T_1 \Rightarrow H_1$

T_1	*"At the end of the year, all solid companies pay dividends."*
H_1	*"At the end of the year, all solid <u>insurance</u> <u>companies</u> pay dividends."*

$T_1 \not\Rightarrow H_2$

T_1	*"At the end of the year, all solid companies pay dividends."*
H_2	*"At the end of the year, all solid companies pay <u>cash</u> <u>dividends</u>."*

Similarity Models would ask:

$$sim(T_1,H_1) > sim(T_1,H_2) \ ?$$

Motivating the approach (2)

$T_1 \Rightarrow H_1$

T_1	*"At the end of the year, all solid companies pay dividends."*
H_1	*"At the end of the year, all solid insurance companies pay dividends."*

$T_1 \not\Rightarrow H_2$

T_1	*"At the end of the year, all solid companies pay dividends."*
H_2	*"At the end of the year, all solid companies pay cash dividends."*

$S_2 < S_1$

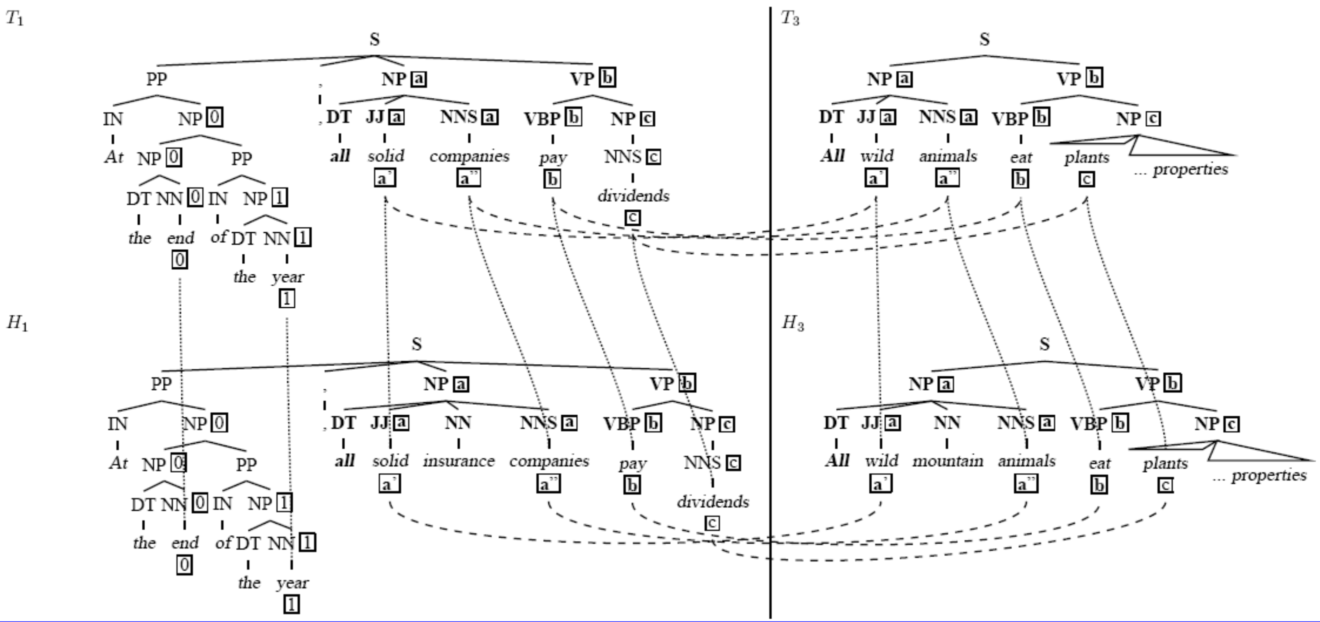
$T_3 \Rightarrow H_3$

T_3	*"All wild animals eat plants that have scientifically proven medicinal properties."*
H_3	*"All wild mountain animals eat plants that have scientifically proven medicinal properties."*

Our Model

Learn *entailment rules* using a similarity between pairs based on: $K((T',H'),(T'',H''))=K_I((T',H'),(T'',H''))+K_S((T',H'),(T'',H''))$

- Intra-pair similarity: $K_I((T',H'),(T'',H''))=s(T',H')\times s(T'',H'')$
- Cross-pair similarity: $K_S((T',H'),(T'',H''))\approx K_T(T',T'')+K_T(H',H'')$



Paraphrase Substitution for Recognizing Textual Entailment

Wauter Bosma

University of Twente

Chris Callison-Burch

University of Edinburgh

QA Answer Validation Exercise

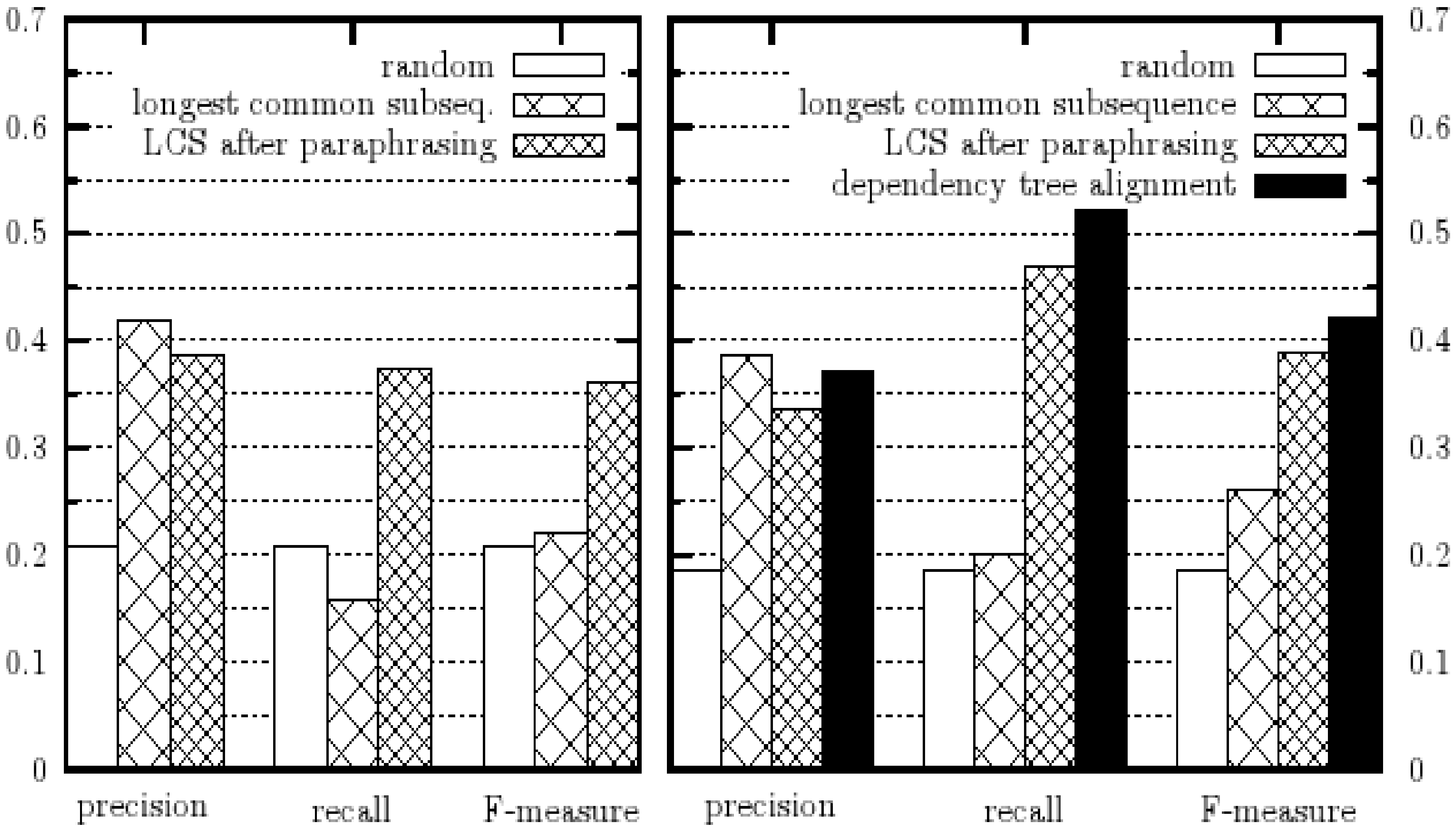
Paraphrasing

- AVE: **detect entailment** between two text passages (*Text* and *Hypothesis*)
- **Paraphrase extraction** from multilingual corpora
- Paraphrasing the *Hypothesis by **substituting word sequences***
- Measuring the **Longest Common Subsequence** to detect entailment relations between *Text* and paraphrases of *Hypothesis*

Example

- **Pair:** 8597 (entailment: yes)
- **Text:** Anthony Busuttil, Professor of Forensic Medicine at Edinburgh University, examined the boy.
- **Hypothesis:** *Anthony Busuttil* is *professor of Forensic Medicine at* the University of *Edinburgh*. (negative judgment; entailment score = 0.67)
- **Hypothesis paraphrase:** *Anthony Busuttil professor of Forensic Medicine at Edinburgh University.* (positive judgment; entailment score = 1.00)

Results


0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
random
longest common subseq.
LCS after paraphrasing
precision
recall
F-measure
random
longest common subsequence
LCS after paraphrasing
dependency tree alignment
precision
recall
F-measure

The bilingual system MUSCLEF

Brigitte Grau, Anne-Laure Ligozat,
Isabelle Robba, Anne Vilnat,
Michael Bagur and Kevin Sejourné

(LIR-LIMSI, CNRS)

Track(s): Question Answering, AVE

Evaluation of translations

- **3 systems :**
  - Monolingual (English) on Web or on CLEF collection
  - Bilingual (French to English)
  - MUSCLEF : fusion of different results
- **Evaluation of system results on:**
  - Initial well-formed questions in English
  - Automatic translation of French version of this set
  - From French questions to English answers by translating terms
- **Evaluation of answer extraction**
  - Named Entity + proximity of terms
  - Application of a syntactic pattern

AVE on French language

- **Hypothesis = question Q + answer R1**
- **Application of FRASQUES on French questions Qs**
  - A snippet = 1 document
  - FRASQUES results:
    - Retrieved terms
    - NE recognition
    - R2 = answer or NIL
- **Elimination of incorrect hypothesis**
- **Evaluation of answer justification**
  - Equivalence of R2 and R1
  - Term scoring: at least 1 term
  - R2 hyponym of the type found in Q -> Wikipedia
- **Questions Qs + known answers**

➢ **A system that says NO, NO, NO : why ?**

MIRACLE at CLEF@QA: QA for Spanish

César de Pablo Sánchez[1], José Luis Martínez[1,3], Ana González-Ledesma[2], Paloma Martínez[1], Antonio Moreno Sandoval[2]

MIRACLE group:

[1]Universidad Carlos III de Madrid

[2]Universidad Automoma de Madrid

[3]Daedalus S.A.

[4]Universidad Politécnica de Madrid

CLEF@QA – wiQA – Real Time QA

Do you have questions?

We have answers!

Try our demo!

We have (some) answers
(in Spanish):

Try our demo (anyway) !

Question Answering in Spanish

- **Main QA Task**
  - Based on NE type filtering
  - DAEDALUS analysis tools + ad-hoc NE processing
  - Fast! ~10sec-20sec → **Real-Time QA**

- **WiQA**
  - NE detection + cos similarity to detect novel content
  - Candidate selection: “inlink” sentences - paragraphs

Visual Micro-clustering Pre-processing for Cross-Language Ad hoc Image Retrieval

INOUE, Masashi

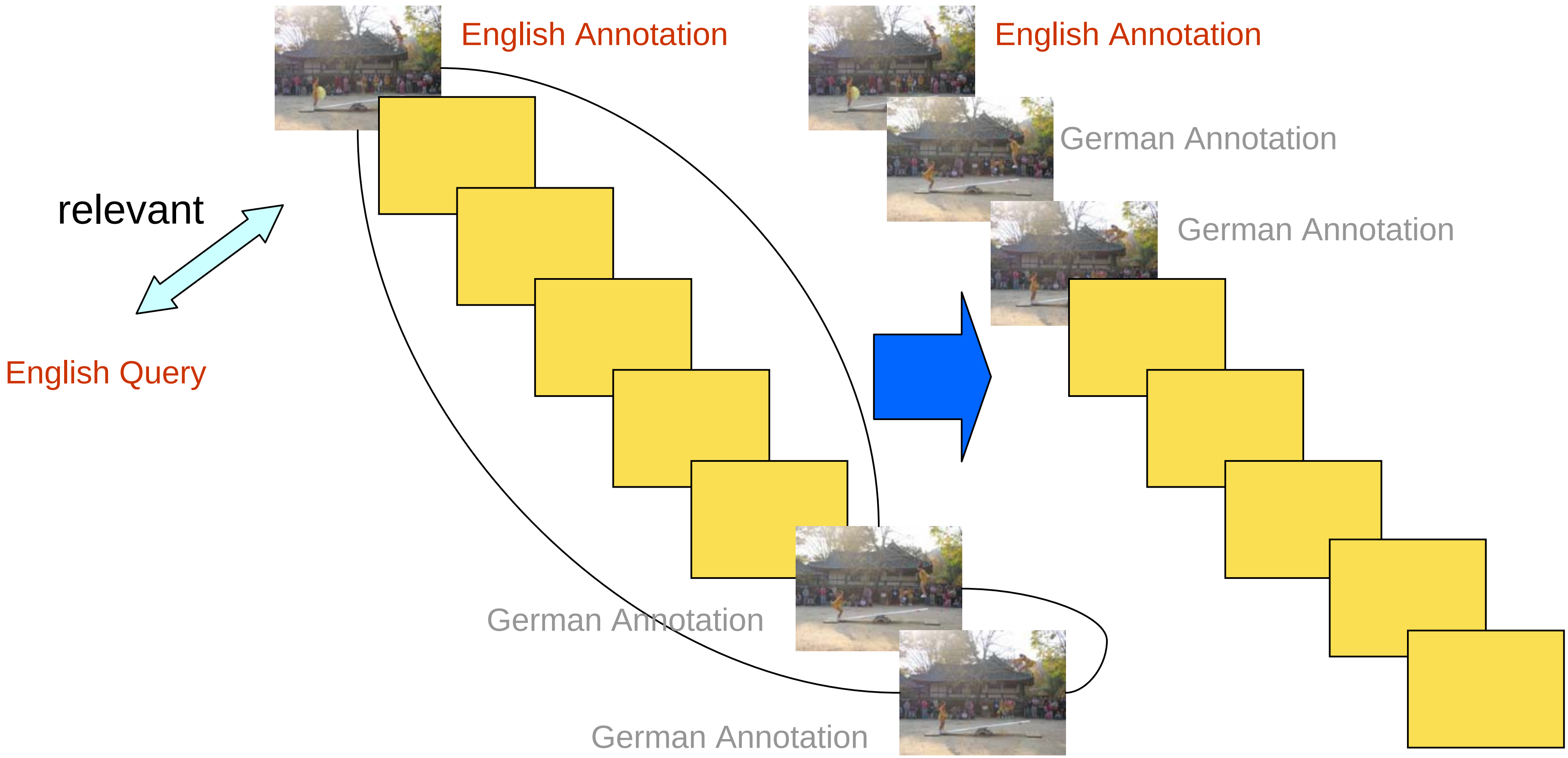
(National Institute of Informatics, Tokyo)

ImageCLEFphoto

Image linkage by visual near-identity

- Task
  - Find relevant images in a multilingual document collection.
- Motivation
  - Images are language independent.
- Assumption
  - Visually identical images are semantically the same.
- Method
  - Connect near-identical images prior to querying.
  - Regard clustered images as relevant when one of the member images is relevant.
  - Re-ordering ranked lists.
- Related Approaches
  - Not ‘find visually similar images’ at querying time
  - Not global partitioning of entire data space

Micro-cluster based re-ranking



OHSU at ImageCLEF 2006

William Hersh

Jayashree Kalpathy-Cramer

Jeffery Jensen

Department of Medical Informatics & Clinical Epidemiology
Oregon Health & Science University
Portland, OR, USA

ImageCLEF med (Medical Retrieval, Automatic)

OHSU participation at ImageCLEF medical retrieval task

- Goals of the OHSU experiments in the medical image retrieval task of ImageCLEF were to assess manual modification of topics
  - With and without visual retrieval techniques
- Submitted three general categories of runs:
  - Automatic textual, manual textual, interactive mixed
- Text based retrieval system based on the open-source search engine, Lucene (tf-idf based system)
- Visual system used medGIFT and neural network-based system
- Automatic textual runs were lowest scoring of our runs
- Best run was English-only manual run with MAP of 0.2132
- Manual modification of topic statements resulted in large performance improvement over automatic topic statements
- Addition of visual retrieval techniques resulted in higher precision at top of output but lower MAP
  - Might be better suited for real users of image retrieval systems

OHSU participation at ImageCLEF medical automatic annotation

- Used a combination of low-level image features and a neural network based classifier
  - Normalized texture, histogram and spatial features
  - Multilayer perceptron architecture to create the multi-class classifier
    - Two layer structure, with a hidden layer of approximately 200-400 nodes
- Error rate of 26.3% for our best run was in the middle of the range of results obtained for all groups
- Most misclassifications were between classes 108/111 and 2/56 (visually similar classes)

Knowledge-Based Medical Image Retrieval

Caroline LACOSTE, Joo Hwee LIM, Jean-Pierre CHEVALLET, Xiong Wei, Daniel RACOCEANU

IPAL, I2R Inst. for Infocomm Research, Singapore

Track(s) : IMAGE-CLEF Medical

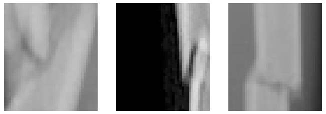
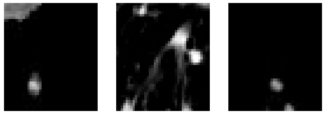
Institute for Infocomm Research

A*STAR

CLEF

Toward a Conceptual Inter-Media Indexing

- Using UMLS Concepts for **Text** and **Image** Indexing
  - Building Visual Terms, associated to concepts
- **Visual**: learning by visual examples
- **Text**: sematic dimension UMLS structure

		UMLS concepts		
Index	Instances	Modality	Anatomy	Pathology
VMT016		Plain X-ray C1306645	Bone C0016658	Fracture C0016658
VMT009		X-ray CT C0040405	Lung C0024109	Nodule C0028259
VMT131		Microscopy C0026018	Blood C0005767	-

How to make a Conceptual Indexing Works ??

- Ingredients of the recipe:
  - UMLS, MetaMap, Concepts, semantic dimension filtering and weighting, visual terms, modality classifier, UMLS concepts associated to Visual Terms, ….
- See Our poster ….

Inter-Media Pseudo-Relevance Feedback Application to ImageCLEF Photo 2006

Nicolas MAILLOT
Jean-Pierre CHEVALLET
Joo Hwee LIM

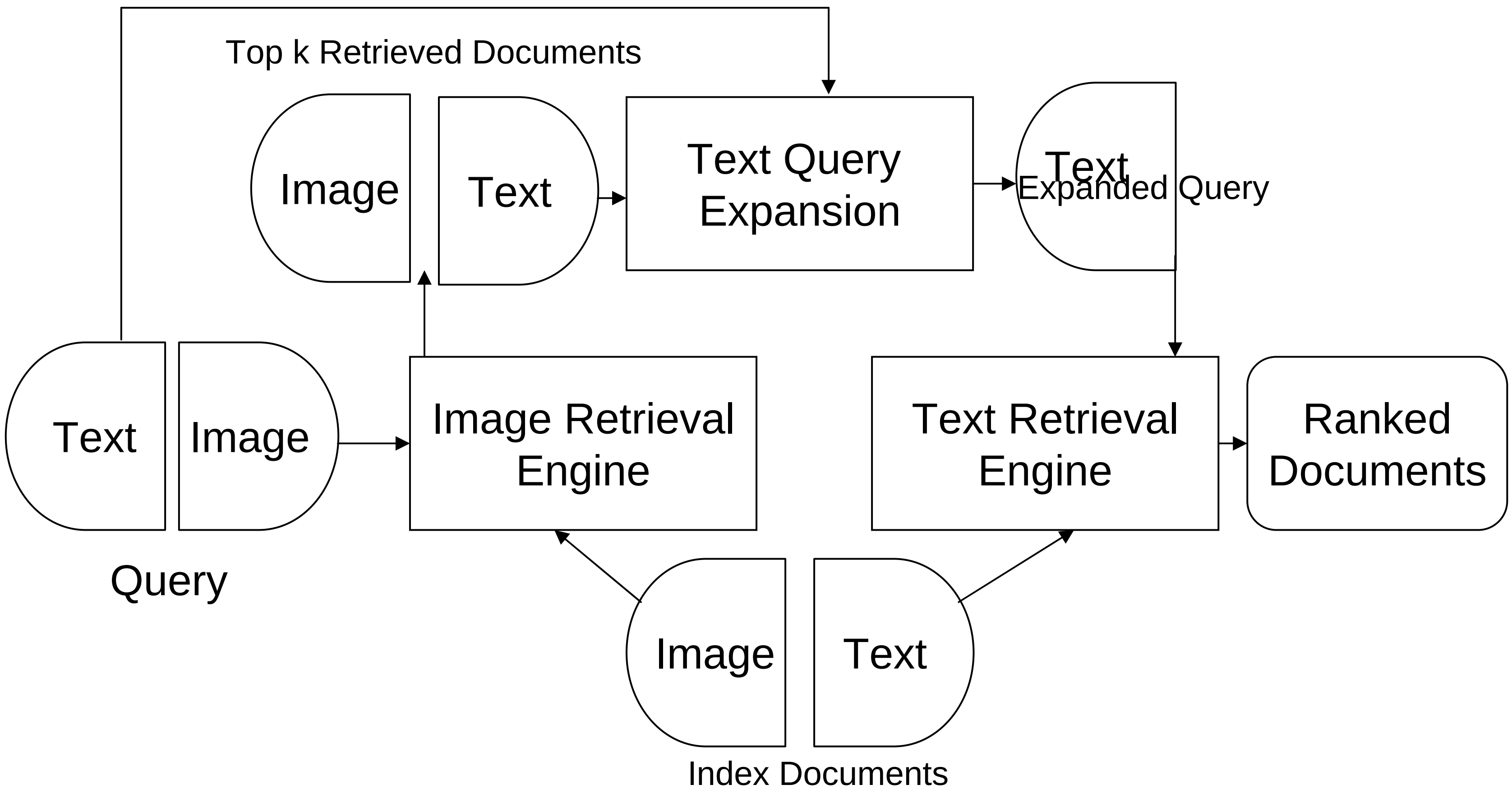
Image Perception, Access & Language (IPAL)
French-Singaporean Joint Lab (CNRS,I2R,NUS,UJF)

Inter-Media Pseudo-Relevance Feedback

- Problem:
  - Multi-modal (text+image) Information Indexing and Retrieval
    - Multi-modal documents and queries
  - Application to the ImageCLEF Photo Task

- Objectives:
  - Inter-media enrichment
  - Why ?
    - for dealing with synonymy
    - Image and text are not at the same semantic level
  - Re-using existing mono-media image and text retrieval systems
- Approach: inter-media pseudo-relevance feedback
  - Using the top k images retrieved by the image retrieval engine for text query expansion

Overview


Top k Retrieved Documents
Image
Text
Text Query Expansion
Text
Expanded Query
Text
Image
Image Retrieval Engine
Text Retrieval Engine
Ranked Documents
Query
Image
Text
Index Documents

Experiments for the CL-SR task at CLEF 2006

Muath Alzghool and Diana Inkpen

University of Ottawa

Canada

Track: Cross Language Spoken Retrieval (CL-SR)

Experiments

- Results for sumbitted runs - English collection
- Results for sumbitted runs - Czech collection
  - Segmentation issues, evaluation score
- Results for different systems: Smart, Terrier
- Query expansion
  - Log likelihood collocations scores
  - Terrier: divergence from randomness
  - Small improvements

Experiments

- Various ASR transcripts (2003, 2004, 2006)
  - New ASR 2006 transcripts do not help
  - Combinations do not help
  - Automatic keywords help

- Cross-language
  - Results good for French to English topic translations
  - Not for Spanish, German, Czech

- Manual summaries and manual keywords
  - Best results

WordNet-based Index Terms Expansion for Geographical Information Retrieval

Davide Buscaldi, Paolo Rosso, Emilio Sanchis

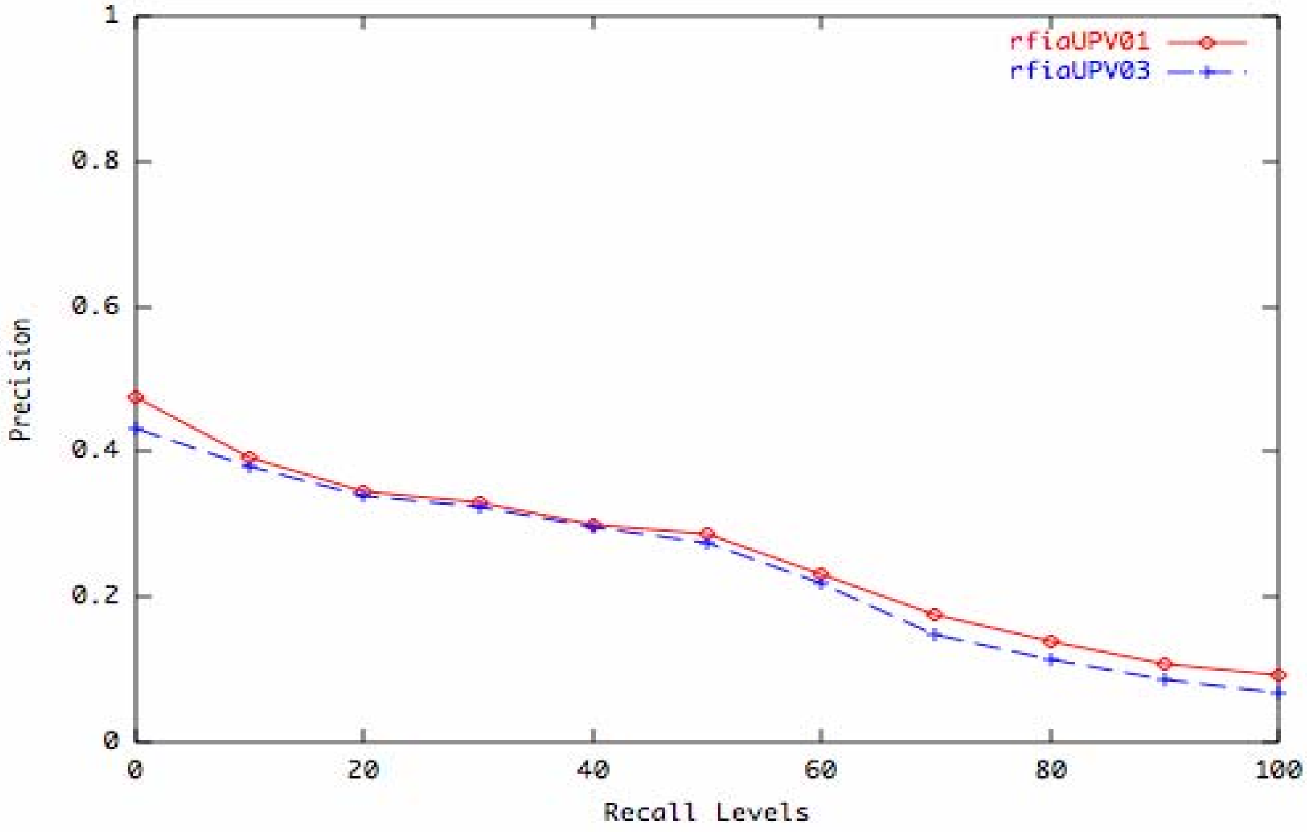
(***DSIC, Universidad Politécnica de Valencia***)

Track:GeoCLEF

Index Terms Expansion

- Helps in finding implicit geographical information in documents by means of WordNet holonyms
  - Ex: Detroit => Michigan => United States => North America => America
- Slight improvement in recall
- We still need to define a good ranking function
- WordNet is limited as a geographical information resource

Comparison between the normal system and the one using index term expansion



MSRA Columbus at GeoCLEF 2006

Zhisheng Li, Chong Wang, Xing Xie, Wei-Ying Ma

Web Search and Mining Group

Microsoft Research Asia

Track(s): GeoCLEF

Runs - Monolingual (EN-EN)

- MSRAWhitelist
  - Expanded the unrecognized locations (such as former Yugoslavia) to several countries manually.
- MSRAManual
  - Based on the MSRAWhitelist, manually modified several queries since these queries are too “natural language” and the keywords of the queries seldom appear in the corpus.
- MSRAExpansion
  - Extract locations from the returned documents and calculate the times each location appears in the documents. Then get the top 10 most frequent location names and combine them with the original geo-terms in the queries.
- MSRALocal
  - Do not use white list or query expansion to expand the query locations. Only utilize a location extraction module to extract the locations from the queries.
- MSRAText
  - Only utilize a pure text search engine “IREngine” to process the queries.

Results

- MSRAManual is the best run among the five ones and then the MSRAWhitelist approach.
- MSRALocal and MSRAText performed similarly.
- The MSRAExpansion performed worst due to the introduced unrelated locations.

- If we only extract the locations from the topics automatically, the retrieval performance does not improve significantly.
- Automatic query expansion will weaken the performance. This is because the topics are too difficult to be handled and the corpus may not be large enough.
- If the queries are formed manually, the performance will be improved significantly.

Blind Relevance Feedback and Named Entity based Query Expansion for Geographic Retrieval at GeoCLEF 2006

Kerstin Bischoff, Thomas Mandl, Christa Womser-Hacker

Information Science
University of Hildesheim, Germany

Track: GeoCLEF

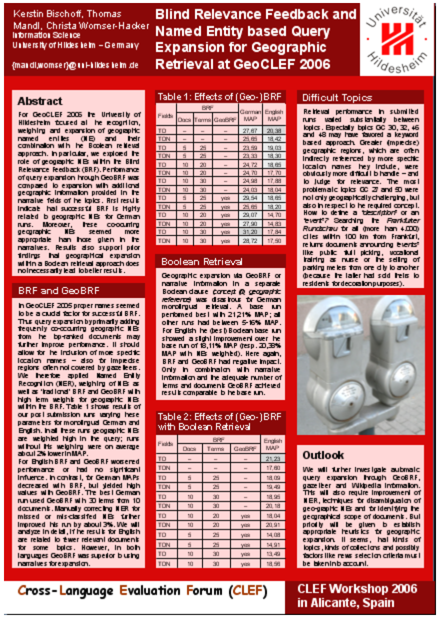
Parking meters, nurse training, fruit picking @ GeoCLEF

or Difficult Topics in Geographic Information Retrieval

Geographic Named Entities and Blind Relevance Feedback

We have more to say on:

- (Geographic) Named Entities
  in Geographic Information Retrieval
- Blind Relevance Feedback
  in Geographic Information Retrieval



Find out at our poster!

FACULDADE · DE · CIÊNCIAS UNIVERSIDADE · DE · LISBOA

The University of Lisbon at CLEF 2006

Nuno Cardoso, Bruno Martins, Marcirio Chaves, Leonardo Andrade and Mário J. Silva

`{ncardoso, bmartins, mchaves, leonardo, mjs} @ xldb.fc.ul.pt`

XLDB Group - Department of Informatics

Faculdade de Ciências da Universidade de Lisboa

http://xldb.fc.ul.pt

The University of Lisbon at CLEF 2006

Poster #1: ad-hoc participation

- **Goal:** Develop a stable testbed to test GIR approaches
- **Improvements since 2005:** QE , and Retrieval & Ranking module.

Poster #2: GeoCLEF participation

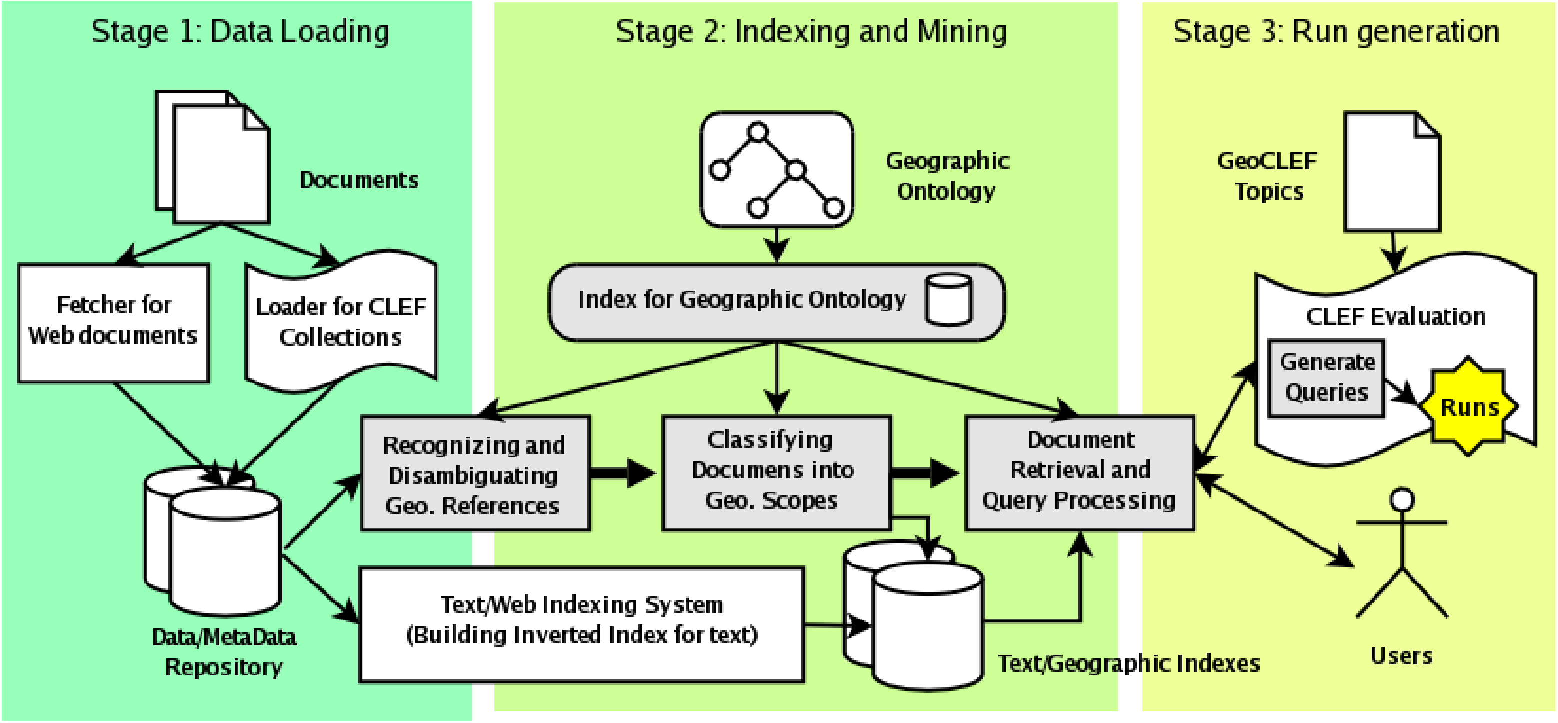
- **Goal:** Compare 2 strategies for GIR against conventional IR approach:

  #1: Geographic text mining

  #2: Augmentation of geographic terms

FACULDADE · DE · CIÊNCIAS UNIVERSIDADE · DE · LISBOA

The University of Lisbon at CLEF 2006



Geographic text mining (strategy #1 at GeoCLEF 2006):

1) Mining geographic references from text.
2) Assign a scope to each document.
3) Convert topics into <*what*, *relation*, *where*> triples.
4) assign scopes to <*where*> terms.
5) BM25 + geographic similarity