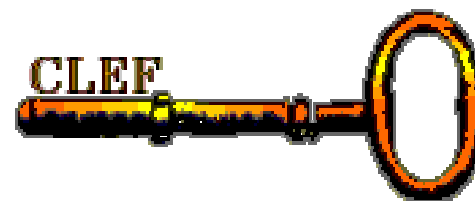
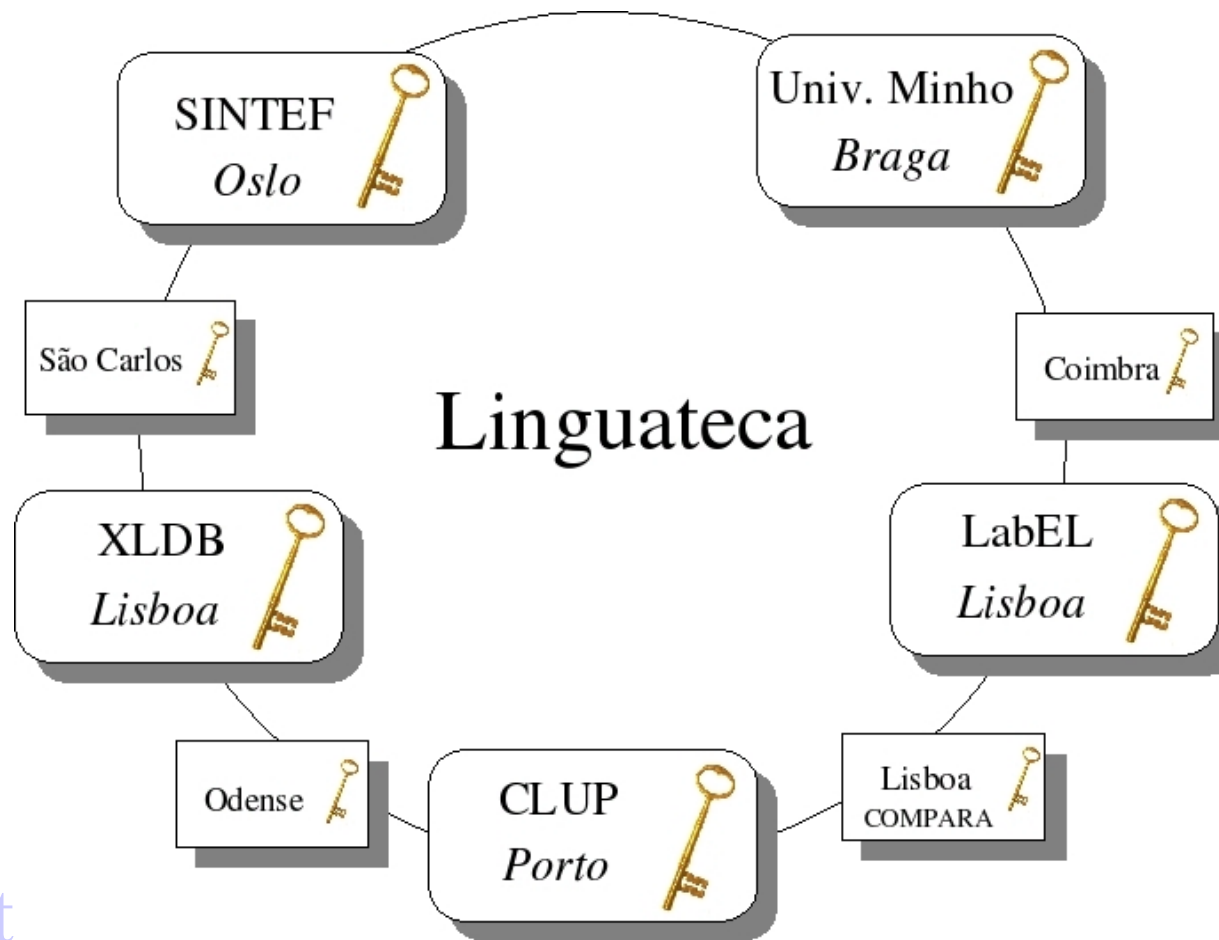

Cross-Language Evaluation Forum

CLEF Workshop 2005

Poster Boaster Session



Portuguese at CLEF 2005: Reflections and Challenges



Diana Santos

Nuno Cardoso

www.linguatca.pt

Hypotheses

- The **quality** of an evaluation campaign depends crucially on the **evaluation setup** (resources and method)
- The beauty of CLEF is **bringing multilinguality into the picture** as early as possible
- To prepare evaluation resources for Portuguese NLP one must know the challenges involved

A provocative presentation around these lines

Using Portuguese as example...

- Tracks where Portuguese was minor
 - WebCLEF
 - GeoCLEF
 - ImageCLEF

and the organizer's paradox
- Tracks where Portuguese was major
 - adhoc IR
 - QA@CLEF

Dictionary-based Amharic-French Information Retrieval

Atelach Alemu Argaw and Lars Asker

Department of Computer and Systems Sciences,
Stockholm University/KTH

Rickard Cöster, Jussi Karlgren and Magnus Sahlgren

Swedish Institute of Computer Science (SICS)

Ad-hoc, Cross-Language

CLEF-2005 Evaluation Campaign: Monolingual, Bilingual, and GIRT Information Retrieval

Jacques Savoy, Pierre-Yves Berger
Institut interfacultaire d'informatique
University of Neuchatel, Switzerland
www.unine.ch/info/clef/

Monolingual in FR, PT, BG and HU languages

- Okapi or Prosit presents the **best retrieval performance** across different languages (compared to the *tf idf* model)
FR +72%, PT +86%, HU +58%
- Rocchio's **blind-query expansion improves** the mean average precision
FR 5% to 14%, PT 3.5% to 14%
mixed results for BG and HU
- **Increasing the query size improves** the mean average precision
from T to TDN : PT +40%, FR +32%,
HU +25%, BG +19%

Bilingual in FR, PT, BG and HU languages

- FR : overall **good translation performance** (87% of a monolingual search), just one system at 77% for PT
- BG and HU : **few translation tools** and **low retrieval performance**
HU 49% of the monolingual search, BG 19%

Monolingual Domain-Specific Retrieval: GIRT

- the Okapi or Prosit model works well for GE and RU
- adding manually assigned descriptors improves the mean average precision
EN +36.5%, GE +14.4%



Research Computing Center
of Moscow State University



NCO Center for Information Research

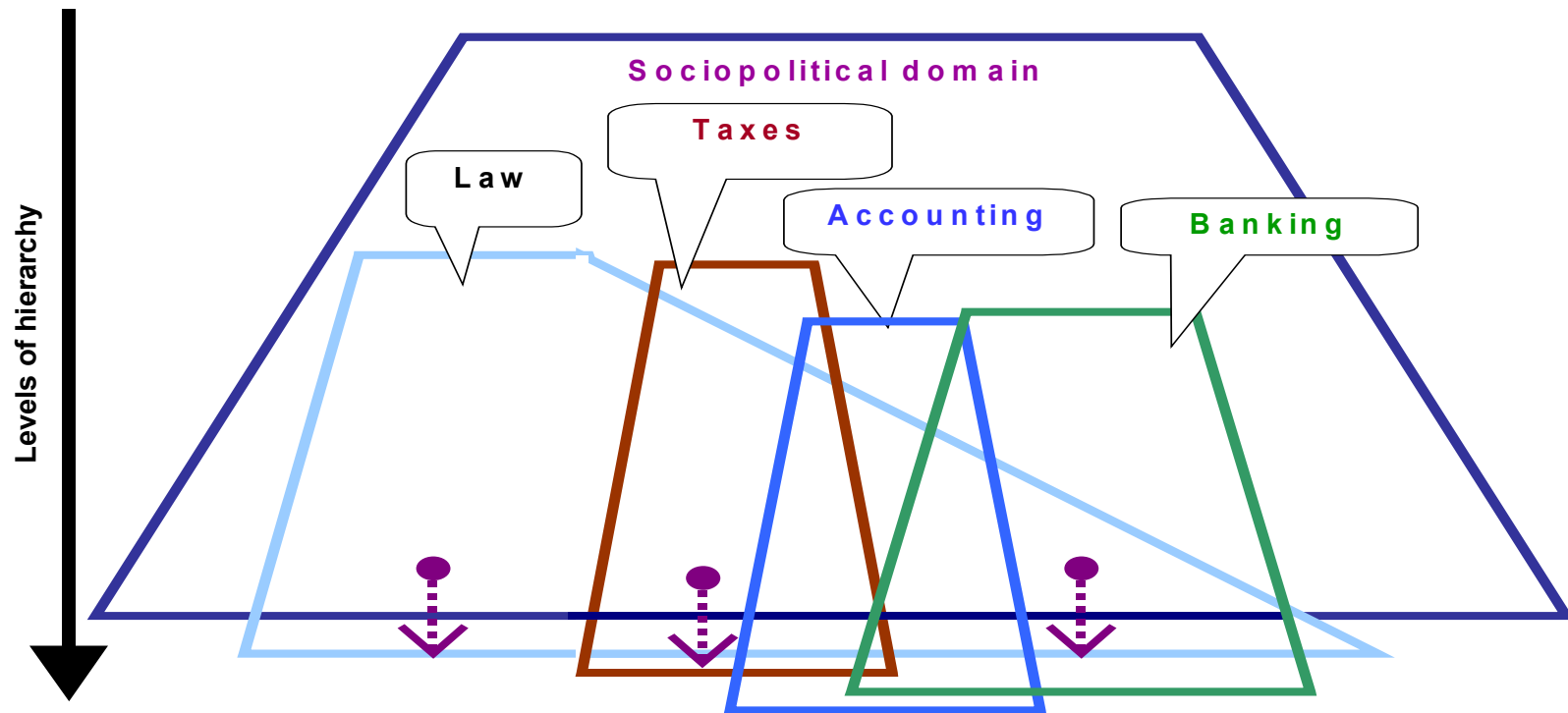
Ageev M., Dobrov B., Loukachevitch N.

**Sociopolitical Thesaurus in
Concept-based Information Retrieval:
Ad-hoc and Domain Specific Tasks**

Sociopolitical Domain

The thesaurus development is based on three methodologies:

- methods of construction of **information-retrieval thesauri** (information-retrieval context, analysis of terminology, terminology-based concepts, a small set of relation types)
- development of **wordnets** for various languages (word-based concepts, detailed sets of synonyms, description of ambiguous text expressions)
- ontology and **formal ontology** research (strictness of relations description, necessity of many-step inference)



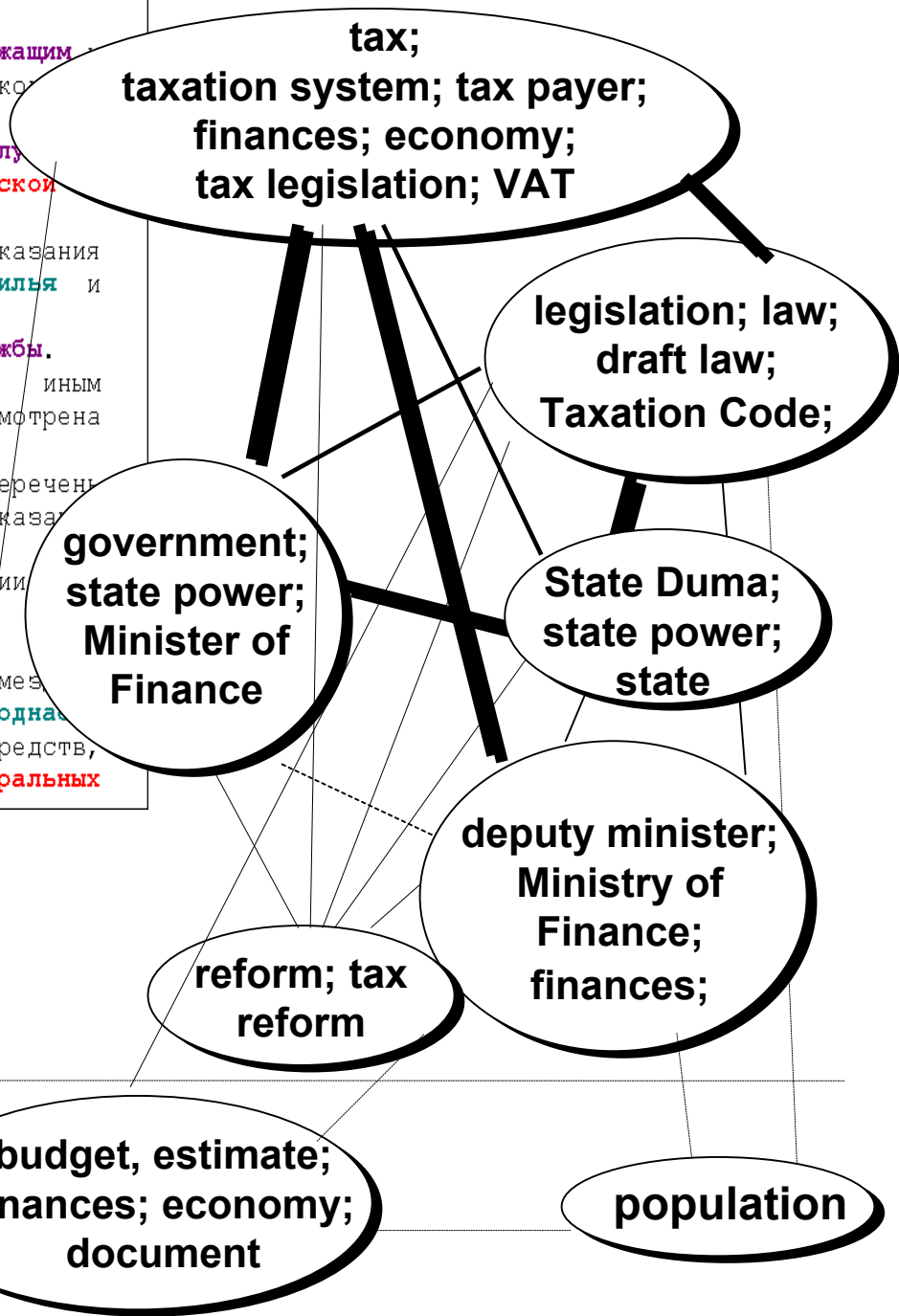
О порядке оказания безвозмездной **финансовой помощи** на **строительство (покупку) жилья** и выплаты **денежной компенсации** за наем (**поднаем**) **жилых помещений** **военнослужащим** **гражданам**, уволенным с **военной службы** Во исполнение Закона Российской Федерации "О статусе **военнослужащих**" и в целях обеспечения прав на **жилище военнослужащих** **граждан**, уволенных с **военной службы**, **Правительство Российской Федерации** постановляет:

1. Утвердить прилагаемое Положение о порядке оказания безвозмездной **финансовой помощи** на **строительство (покупку) жилья** и выплаты **денежной компенсации** за наем (**поднаем**) **жилых помещений** **военнослужащим** и **гражданам**, уволенным с **военной службы**.

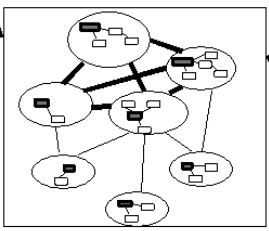
2. **Министерству обороны** Российской Федерации и иным **федеральным органам исполнительной власти**, в которых предусмотрена **военная служба**:

в месячный срок разработать и утвердить формы и перечни документов, необходимых для принятия решения об оказании **военнослужащим** безвозмездной **финансовой помощи** на **строительство (покупку) жилья** и о выплате **денежной компенсации** за наем (**поднаем**) **жилых помещений**;

расходы, связанные с оказанием **военнослужащим** безвозмездной **финансовой помощи** и выплатой **денежной компенсации** за наем (**поднаем**) **жилых помещений**, производить за счет и в пределах средств, выделяемых из федерального бюджета по сметам этих **федеральных**



Russian document



Thematic representation of document

PERSONAL DATA	WORK/ACTIVITY	PERSONAL DATA BY CATEGORY
NAME (LAST, FIRST, MIDDLE)	NAME (LAST, FIRST, MIDDLE)	NAME (LAST, FIRST, MIDDLE)
DATE OF BIRTH	DATE OF BIRTH	DATE OF BIRTH
PLACE OF BIRTH	PLACE OF BIRTH	PLACE OF BIRTH
EDUCATION	EDUCATION	EDUCATION
PROFESSION	PROFESSION	PROFESSION
EMPLOYMENT	EMPLOYMENT	EMPLOYMENT
RESIDENCE	RESIDENCE	RESIDENCE
MARRIAGE	MARRIAGE	MARRIAGE
CHILDREN	CHILDREN	CHILDREN
RELIGION	RELIGION	RELIGION
LANGUAGES	LANGUAGES	LANGUAGES
SKILLS	SKILLS	SKILLS
HOBBIES	HOBBIES	HOBBIES
CHARACTERISTICS	CHARACTERISTICS	CHARACTERISTICS
ADDITIONAL INFORMATION	ADDITIONAL INFORMATION	ADDITIONAL INFORMATION

Russian representation



English document

PUBLIC OFFICIAL	OFFICIAL	CIVIL SERVICE	STATE
ORGANIZATION	STAFF	PUBLIC OFFICE	GOVERNMENT MINISTER; DEPARTMENT...
POPULATION	CITIZEN	ORIGIN	EXTRACTION;
DISABILITY			
GOVERNMENT	BODY OF THE STATE POWER;	STATE POWER	FEDERAL BODY;
EXECUTIVE BODY	MANAGEMENT		
STATUS	OFFICIAL		
EUROPE	COUNCIL OF EUROPE;	PARLIAMENTARY ASSEMBLY OF COUNCIL OF EUROPEANS;	GENERAL

English representation

Ah Hoc – Multilingual Merging task

**Dublin City University at CLEF 2005:
Multilingual Merging Experiments**

**Adenike Lam-Adesina
Gareth J.F. Jones**

School of Computing

Dublin City University, Ireland

Summary

- Explore a range of standard and extended data fusion list merging techniques. (raw scores, various forms of score normalization, etc.)
- Apply to two sets of ranked lists provided by two laboratories.
- Investigate whether merging techniques behave consistently with respect to precision and recall for ranked lists from different sources.
- Results so far indicate that these techniques do not behave consistently.
- Further investigation is needed to try to understand the reasons for this.



Ranking documents by using term proximity

Fuzzy proximity method

Annabelle MERCIER

Amélie IMAFOUO

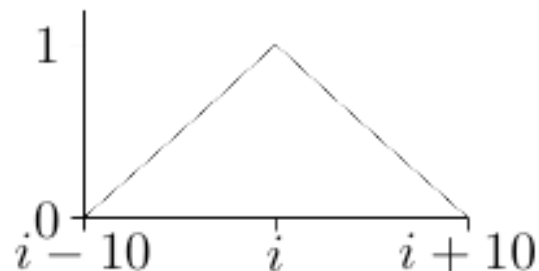
Michel BEIGBEDER

Monolingual French Ad Hoc

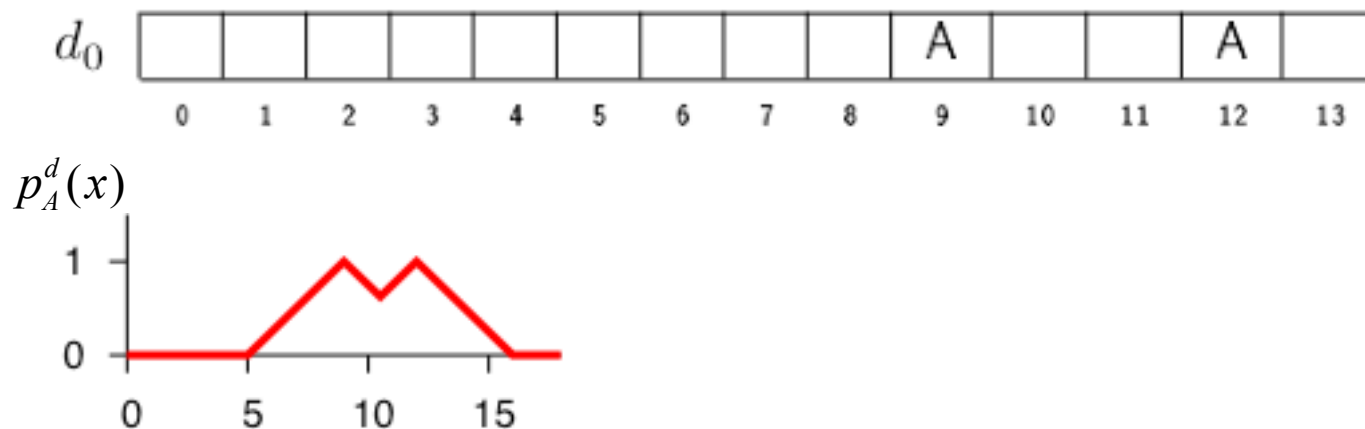


Fuzzy term proximity

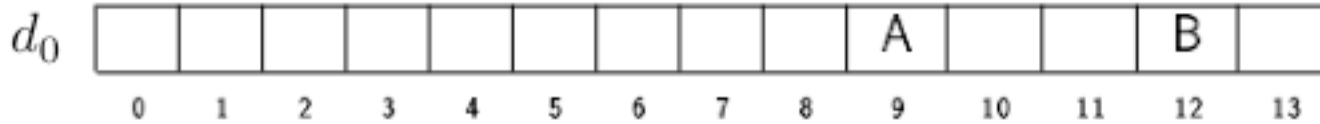
- Fuzzy proximity to a term occurrence at a position in the document is the value of an influence function (e.g. triangular) centered at the query term position occurrence



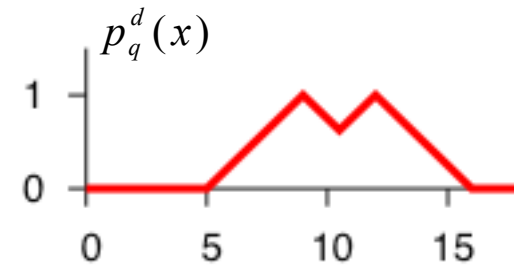
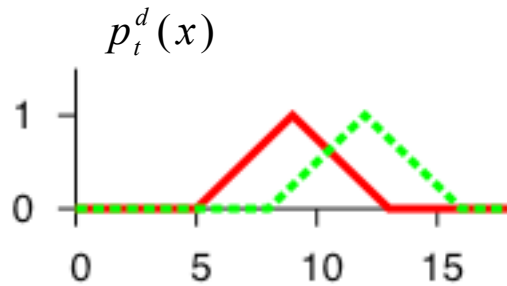
- Representation of fuzzy proximity to a term in the whole document



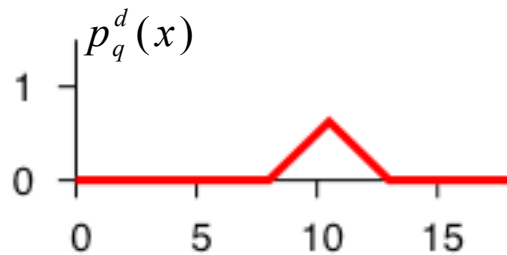
Fuzzy query proximity and document score



- Fuzzy proximity to the terms **A** and **B**
- Fuzzy proximity to the query **A** or **B**



- Fuzzy proximity to the query **A** and **B**
- Score is computed by :



$$\int_{-\infty}^{+\infty} p_q^d(x) dx$$



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

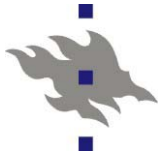
Question Answering with Semantic Annotation

Lili Aunimo

Department of Computer Science

University of Helsinki, Finland

Multiple Language Question Answering Track



Question Answering with Semantic Annotation

■ Monolingual Finnish

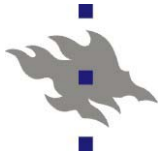
- Accuracy: 26.5% (C), 23.0% (1) and 19.0% (2)

■ Monolingual French

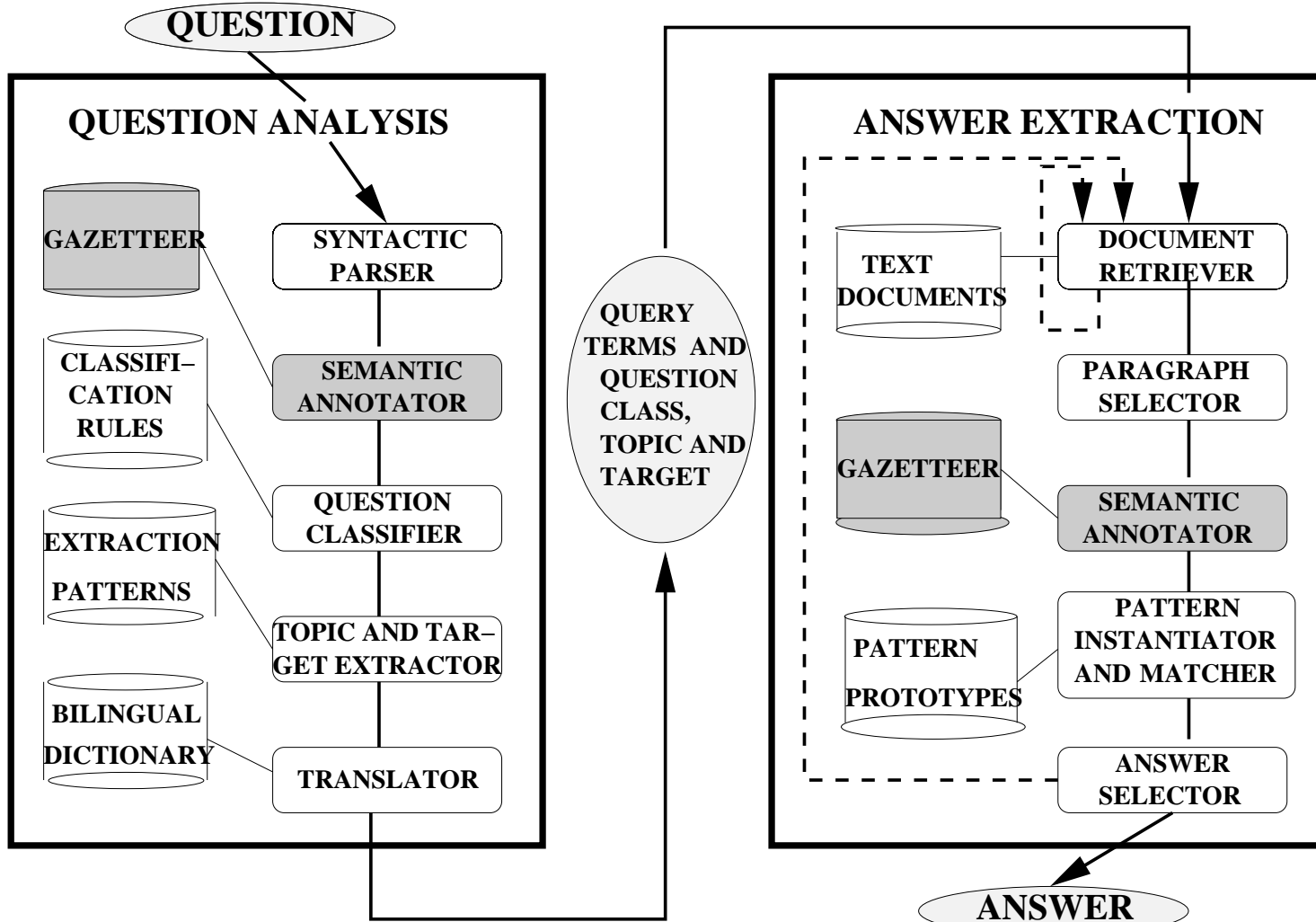
- Accuracy: 18.0% (C), 17.5% (1), 16.5 (2)

■ Bilingual Finnish-English

- Accuracy: 14.0% (C), 12.5% (1), 10.0 (2)



System Architecture



Exploiting Linguistic Indices and Syntactic Structures for Multilingual QA

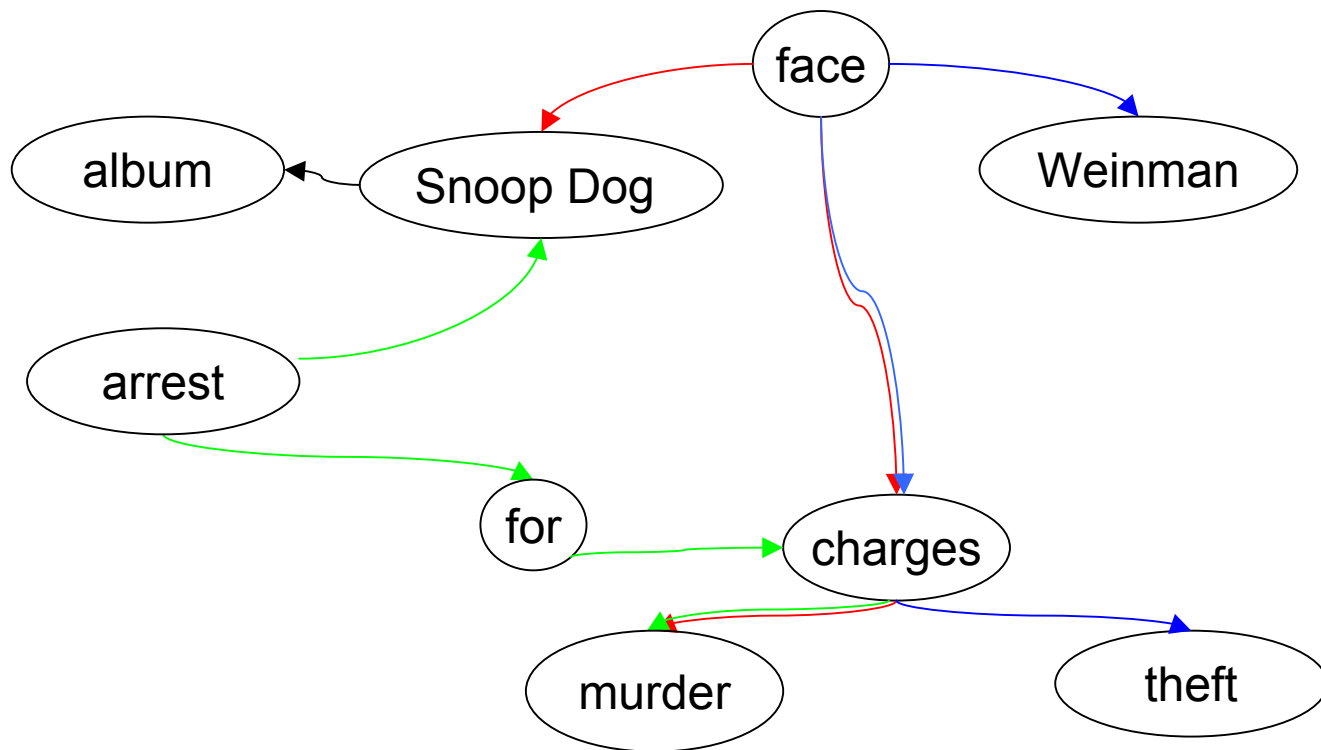
Tanev H., Kouylekov M., Magnini B.,
Negri M., and Simov K*.

ITC-irst, Trento, Italy

*Bulgarian Academy of Sciences

Syntactic Network Model

SyntNet



Exploiting Linguistic Structures

- Edit-distance algorithm on syntactic trees for answer extraction. The algorithm is adaptation of the [Zhang and Shasha'90]
- Monolingual Bulgarian task - sentence-level linguistic index with:
 - word lemmas
 - part-of-speech tags

Question Answering as Semistructured Information Retrieval

David Ahn, Valentin Jijkoun, Karin Müller,
Maarten de Rijke, Erik Tjong Kim Sang

- ▶ More challenging questions give rise to additional types of markup
- ▶ Our strategy: move tagging, parsing steps and other costly evaluations off-line for performance reasons
- ▶ Use XML for storing individual annotation layers
- ▶ Result: question answering becomes semi-structured information retrieval



- ▶ Question answering as Semistructured Information Retrieval
 - ▶ Off-line processing
 - ▶ Not XML
 - ▶ Query language
 - ▶ Question-to-query mapping
 - ▶ Query evaluation, result re-ranking, ...
- ▶ Current state
- ▶ Next steps



20th Century Esfinge (Sphinx) solving the riddles at CLEF 2005

PT-PT and EN-PT question answering tasks

Luís Costa

Luis.costa@sintef.no

Linguatca / SINTEF ICT

<http://www.linguatca.pt>





Esfinge overview

- ❖ QA system exploring the redundancy existent in the Web and the fact that Portuguese is one of the most used languages in the Web.
- ❖ Two strategies tested:
 - Searching the answers in the Web and using the CLEF doc. collection to confirm them.
 - Searching the answers only in CLEF doc. collection.

What was new at CLEF 2005

- ✓ Use of NER system SIEMES.
- ✓ List of not interesting websites (jokes, blogs, etc.).
- ✓ Use Lingua::PT::Stemmer for the generalization of search patterns.
- ✓ Filtering of undesired answers.
- ✓ Searching longer answers.
- ✓ Participation in the EN-PT multilingual task.

ImageCLEF (ad hoc)

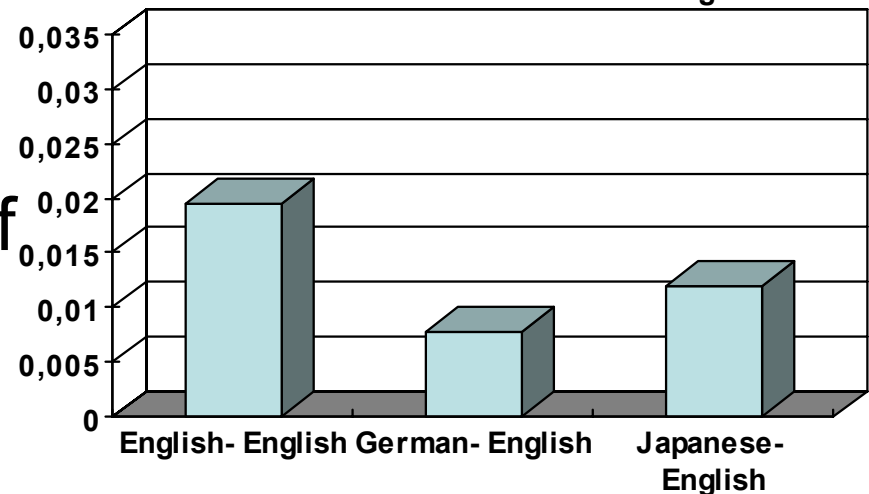
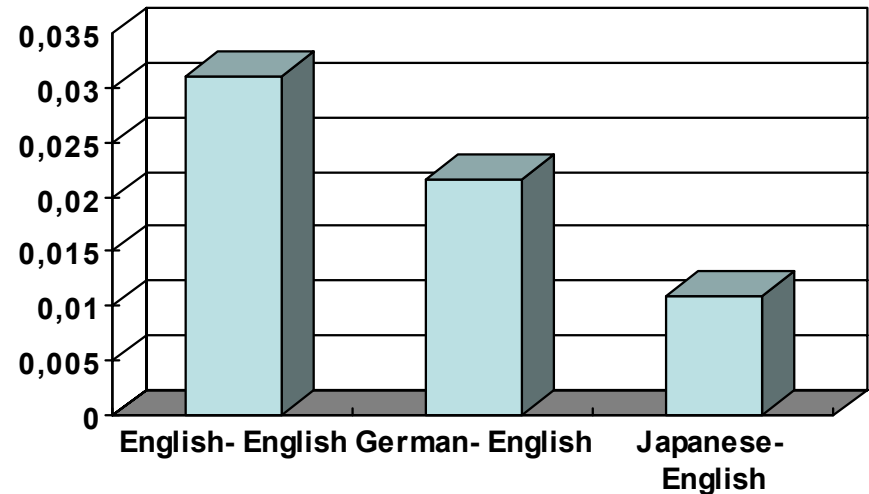
Recovering translation errors in
cross-language image retrieval using
word-association models

INOUE, Masashi

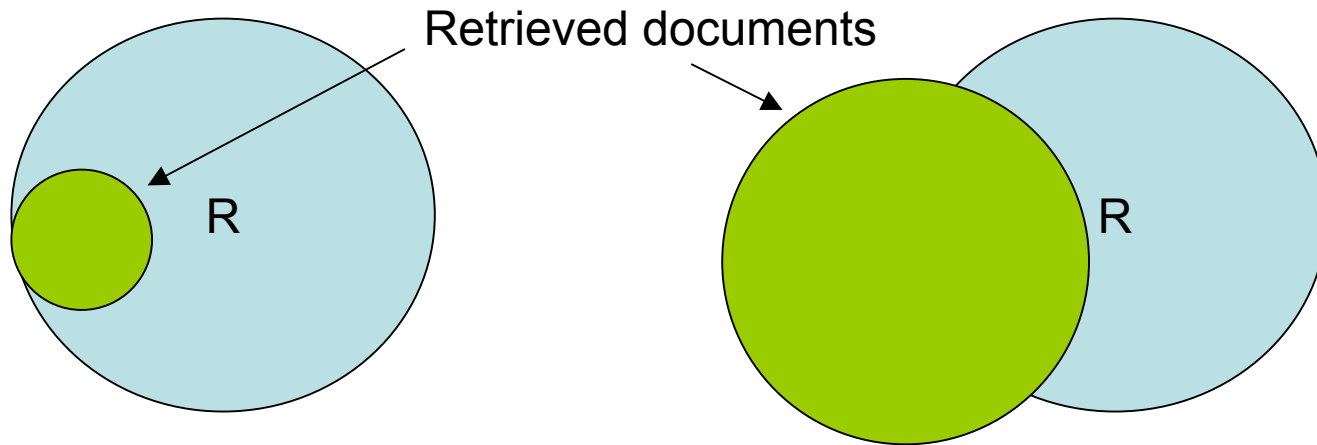
NII, Tokyo

I. Effects of Translation Errors

- Worse machine translation
->
Worse retrieval performance
- Retrieval strategy under the presence of translation errors?



II. Combination of Two Models



- Precision-oriented model (misses many)

- Recall-oriented model (includes noises)

Combine their outputs

Manual Query Modification and Automated Translation to Improve Cross-language Medical Image Retrieval

Jeffery R. Jensen

William R. Hersh

Department of Medical Informatics & Clinical Epidemiology

Oregon Health & Science University

Portland, OR, USA

{jensejef, hersh}@ohsu.edu

OHSU Participation in the ImageCLEF Medical Task

- Experiments based on a context-based image retrieval system we developed
- We also attempted to improve our performance by augmenting the image output with results made available from a content-based search
- Our three experiments included
 - Automatic query
 - Manual query
 - Manual/Visual query

Best Run – OHSUman

- Manual modification of the query for each topic
 - Keywords were expanded or mapped to more specific terms
 - Translations were obtained from Babelfish
 - <http://babelfish.altavista.com>
 - Used “term boosting”
- MAP of 0.2116 (compared to 0.2821 best overall)
- Next best run used output from OHSUman
 - excluded all documents that were not retrieved by the University of Geneva content-based “baseline” run (GE_M_4g.txt)



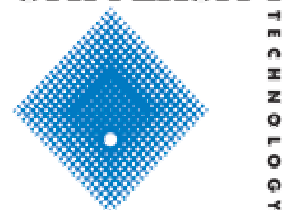
Towards a Topic Complexity Measure for Cross-Language Image Retrieval

Michael Grubinger¹, Clement Leung¹, Paul Clough²

¹School of Computer Science and Mathematics, Victoria University, Melbourne, Australia
{michael.grubinger@research.vu.edu.au}, {clement.leung@vu.edu.au}

²Department of Information Studies, Sheffield University, Sheffield, UK
{p.d.clough@sheffield.ac.uk}

**VICTORIA
UNIVERSITY**





MOTIVATION

Topic selection process → important part of any benchmarking event:

- representative for (image) collection
- reflect realistic user interests/needs

Topics are generated against certain dimensions, including

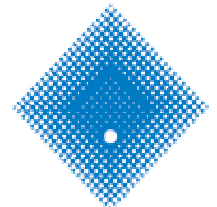
- estimated number of relevant images for each topic
- variation of task parameters to test different translation problems
- scope of each topic (e.g. broad or narrow, general or specific)
- difficulty of the topic (topic complexity)

If topics are too easy

- ✗ results are not meaningful
- ✗ no challenge for returning participants

If topics are too difficult

- ✗ results are not meaningful
- ✗ hard for new participants to obtain good results
- ✗ prevent them from presenting results



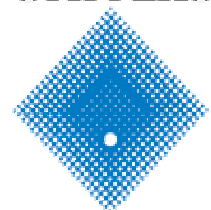


THE POSTER

- need for a **measure of topic complexity** to ground the topic generation process within a methodical and reliable framework

CONTENTS

- ✓ Definition of the Topic Complexity Measure
- ✓ Analysis of ad-hoc topics (ImageCLEF 2005)
- ✓ Comparison with ad-hoc results (ImageCLEF 2005)
- ✓ Correlation of MAP and Topic Complexity Level
- ✓ Further verification with results from ImageCLEF 2003 and 2004
- ✓ Future perspectives for ImageCLEF 2006



Dublin City University at CLEF 2005: Experiments with the ImageCLEF St Andrew's Collection

Gareth J. F. Jones Kieran McDonald

Centre for Digital Video Processing
Dublin City University, Ireland

Outline

C e n t r e f o r D i g i t a l V i d e o P r o c e s s i n g

- Text-retrieval using standard Okapi BM25 with pseudo relevance feedback (PRF).
- Images indexed using standard low-level features: colour, edge and texture.
- Research question: can we combine evidence from image and text retrieval to improve overall retrieval effectiveness?
- DCU CLEF 2004 found that merging ranked lists from text and image retrieval is little different to text only retrieval.
- Hypothesis: documents retrieved by both text and image systems are more likely to be relevant.
- Method investigate. Modify text retrieval with PRF to assume documents retrieved using both text and image relevant.
- Results so far indicate that this method again gives little difference to text only retrieval with PRF.

CLEF-2005 CL-SR at Maryland: Document and Query Expansion Using Side Collections and Thesauri

Jianqiang Wang and Douglas W. Oard

College of Information Studies and UMIACS
University of Maryland, USA

Track: CL-SR

Techniques and results

- Document expansion using side collections
 - 24% (significant) improvement on the training set
 - 6% (insignificant) improvement on the evaluation set
 - Parameter tuning is important
- Keyword “synonym” expansion using thesauri
 - Marginal (insignificant) improvement on the training set
- Query expansion using blind relevance feedback
 - 45% (significant) improvement on CLIR with ASR text
 - Marginal (insignificant) improvement on searching metadata
 - Parameter tuning could be productive
- Corpus-based CLIR
 - 674,247 sentence aligned corpus (Europarl)
 - One-best French-to-English query translation
 - 79% monolingual mean average precision

Error Analysis

In ASR

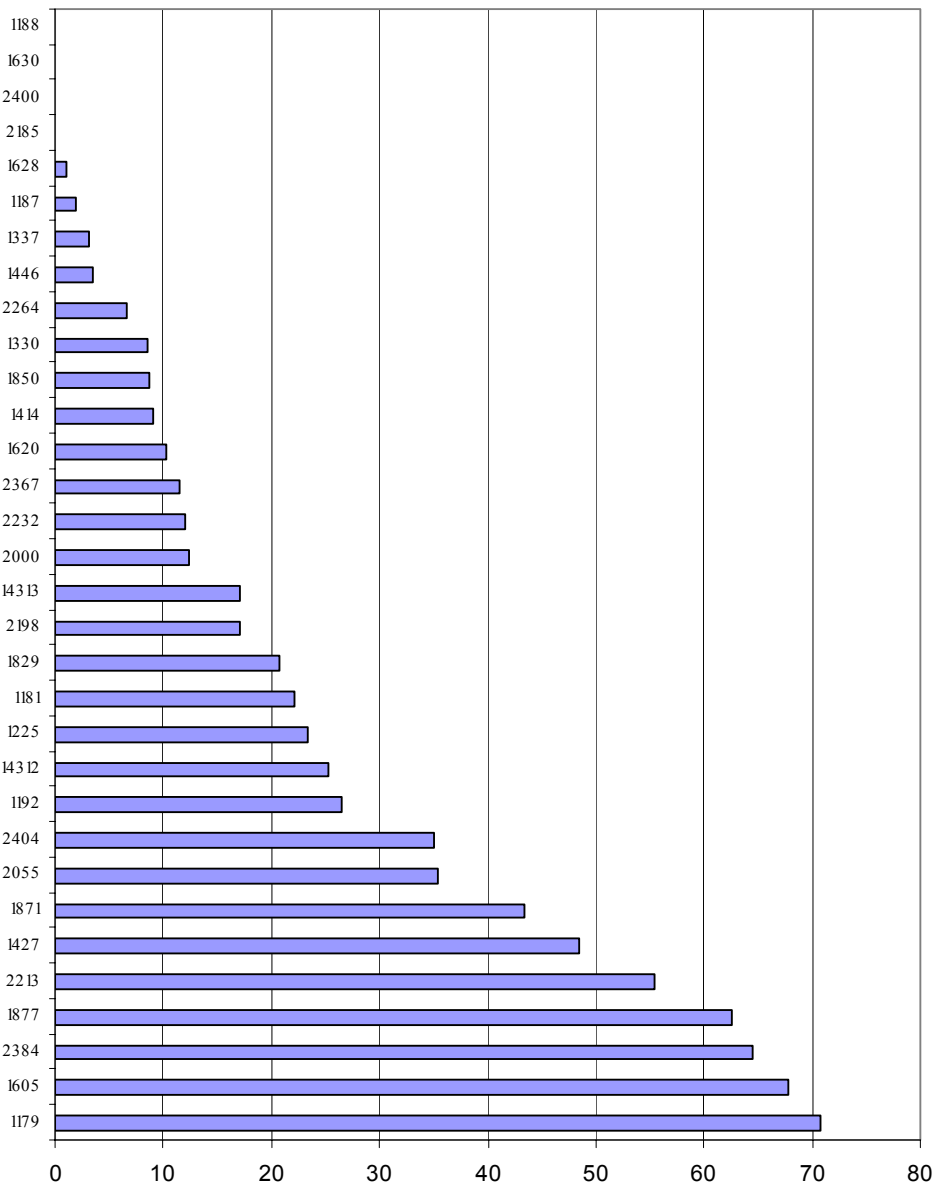
(Segments: ASR / Metadata)

jews
witness
sobibor(5/13) death camp
holocaust
slave labor
labor camps
art auschwitz
minsk(21/71) ghetto underground
abusive female(8/81) personnel
wallenberg(3/16)
wallenberg(3/16) rescue jews
fort ontario refugee camp
jewish nurses concentration camps
war crime trial participants
french internment camps
post liberation experience
birkenau daily life
sonderkommando
jewish gentile relations poland
sonderkommando auschwitz
liberation buchenwald dachau
jewish kapos
kindertransport
decision migration australia
flight denmark sweden
dp camps american zone
rescue danish children
persecution jews italy
kindertransport possessions
red cross holocaust
jews shanghai
bulgaria saved jews

Only in Metadata

volkswagen
eichmann
sinti roma
telefonken
ig farben
eichmann

ASR of % Metadata



Using various indexing schemes and multiple translations in the CL-SR task at CLEF 2005

Track: Cross Language Spoken Retrieval

Diana Inkpen, Muath Alzghool, and Aminul Islam

University of Ottawa

{diana,alzghool,mdislam}@site.uottawa.ca

System overview. Results.

- Off the shelf components:
 - Smart IR system with many indexing schemes.
 - 7 online MT tools, combined the results

Results of the five submitted runs, for topics in English, Spanish, French, and German (**best results out of the 7 teams**):

Language	Run	map	bpref	Fields	Description
English	uoEnTDN	0.2176	0.2005	TDN	Weighting scheme: mpc/ntn
Spanish	uoSpTDN	0.1863	0.1750	TDN	Weighting scheme: mpc/ntn
French	uoFrTD	0.1685	0.1599	TD	Weighting scheme: mpc/ntn
English	uoEnTD	0.1653	0.1705	TD	Weighting scheme: mpc/ntn
German	uoGrTDN	0.1281	0.1331	TDN	Weighting scheme: mpc/ntn

Next best system: map 0.1288, bpref 0.1719 on English TD

Results of additional experiments

- Results on the output of each Machine Translation system: Spanish, French, German, and Czech. Combination is better.
- Results of the various indexing schemes (best one: mpc/ntn on TDN).
- Results on phonetic n-grams, and combination text plus phonetic n-grams.
 - map 0.2590, bpref 0.1585
- Results of indexing all fields: manual keywords and summaries, ASR transcripts.
 - map 0.4647, bpref 0.3660

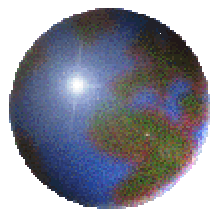
Cross-Language Spoken Document Retrieval Task

Dublin City University at CLEF 2005: Cross-Language Spoken Document Retrieval (CL-SR) Experiments

Adenike Lam-Adesina Gareth J.F. Jones
School of Computing
Dublin City University, Ireland

Summary

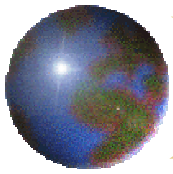
- Retrieval based on Okapi BM25 with standard stopping and stemming.
- Summary-based pseudo relevance feedback.
 - Sentence-based summary method extended to word clusters in running transcriptions containing no punctuation markers..
- Topic translation to English using Systran.
- Exploration of retrieval using running transcription combined with various metadata fields.
- Results indicate that manually generated metadata is more effective for retrieval.



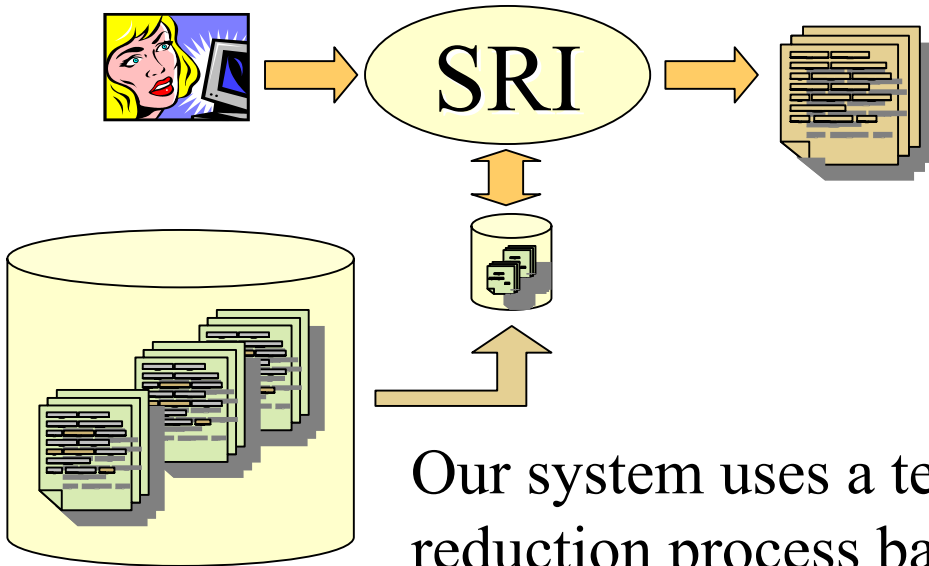
TPIRS: A System for Document Indexing Reduction on WebCLEF

David Pinto & Héctor Jiménez-Salazar
Faculty of Computer Science, BUAP, Mexico

Paolo Rosso & Emilio Sanchis
Department of Information Systems and Computation, UPV, Spain



TPIRS (Transition Point IRS)

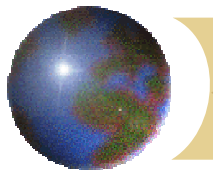


Our model is based on the boolean model with Jaccard similarity function for ranking.

Our system uses a term reduction process based on the Transition Point Technique.

$$TP = \frac{\sqrt{8 * I_1 + 1} - 1}{2}$$

$$TP_{SET} = \{t_i | (t_i, f_i) \in V_{TP}, U_1 \leq f_i \leq U_2\}$$



Results (BiEnEs task)

Corpus	Size (Kb)	% of Reduction
Full	117,345	0%
TP10	12,616	89.25%
TP20	19,660	83.25%
TP40	20,477	82.55%
TP60	28,903	75.37%

Corpus	Average Success at					Mean Reciprocal Rank
	1	5	10	20	50	
Full	0.0224	0.0672	0.1119	0.1418	0.1866	0.0465
TP10	0.0224	0.0373	0.0672	0.0821	0.1119	0.0331
TP20	0.0299	0.0448	0.0672	0.1045	–	0.0446
TP40	0.0597	0.0970	0.1119	0.1418	0.2164	0.0844
TP60	0.0522	0.1045	0.1269	0.1642	0.2090	0.0771



The XLDB Group at GeoCLEF 2005

Nuno Cardoso, Bruno Martins, Marcirio
Chaves, Leonardo Andrade, 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 XLDB Group at GeoCLEF 2005



+ geographic reasoning = **GeoTumba**

- **Geo-scope** = footprint = focus = ...

- Documents have geo-scopes

(One sense per discourse)

- Queries have geo-scopes

- Search: similarity using index

terms + geo-scopes

GeoCLEF participation

- Geo-IR system with components from GeoTumba

- In GeoCLEF we experimented:

- **Ranking Geo-scopes**

- **Scope Assignment Algorithms**

- **Location Terms Expansion**

- **Topic Translation**

The XLDB Group at GeoCLEF 2005

Geo-IR architecture

