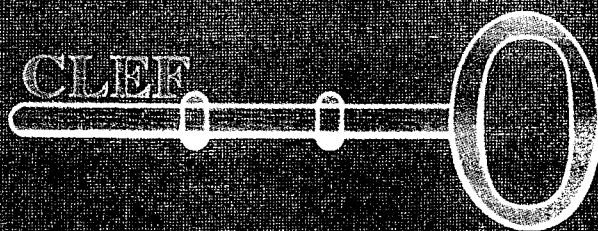


Carol Peters Fredric C. Gey
Julio Gonzalo Henning Müller
Gareth J.F. Jones Michael Kluck
Bernardo Magnini Maarten de Rijke (Eds.)

LINCS 4022

Accessing Multilingual Information Repositories

6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005
Vienna, Austria, September 2005
Revised Selected Papers



 Springer

Lecture Notes in Computer Science

The LNCS series reports state-of-the-art results in computer science research, development, and education, at a high level and in both printed and electronic form. Enjoying tight cooperation with the R&D community, with numerous individuals, as well as with prestigious organizations and societies, LNCS has grown into the most comprehensive computer science research forum available.

The scope of LNCS, including its subseries LNAI and LNBI, spans the whole range of computer science and information technology including interdisciplinary topics in a variety of application fields. The type of material published traditionally includes

- proceedings (published in time for the respective conference)
- post-proceedings (consisting of thoroughly revised final full papers)
- research monographs (which may be based on outstanding PhD work, research projects, technical reports, etc.)

More recently, several color-cover sublines have been added featuring, beyond a collection of papers, various added-value components; these sublines include

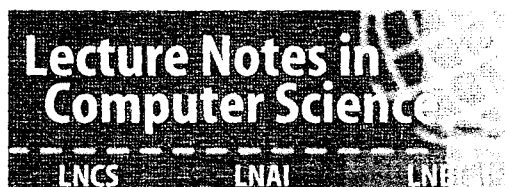
- tutorials (textbook-like monographs or collections of lectures given at advanced courses)
- state-of-the-art surveys (offering complete and mediated coverage of a topic)
- hot topics (introducing emergent topics to the broader community)

In parallel to the printed book, each new volume is published electronically in LNCS Online.

Detailed information on LNCS can be found at
www.springer.com/lncs

Proposals for publication should be sent to
LNCS Editorial, Tiergartenstr. 17, 69121 Heidelberg, Germany
E-mail: lncs@springer.com

ISSN 0302-9743



› springer.com

Table of Contents

What Happened in CLEF 2005	1
<i>Carol Peters</i>	
Part I. Multilingual Textual Document Retrieval (Ad Hoc)	
CLEF 2005: Ad Hoc Track Overview	11
<i>Giorgio M. Di Nunzio, Nicola Ferro, Gareth J.F. Jones, Carol Peters</i>	
Cross-Language and More	
Ad-Hoc Mono- and Bilingual Retrieval Experiments at the University of Hildesheim	37
<i>René Hackl, Thomas Mandl, Christa Womser-Hacker</i>	
MIRACLE at Ad-Hoc CLEF 2005: Merging and Combining Without Using a Single Approach	44
<i>José M. Goñi-Menoyo, José C. González-Cristóbal, Julio Villena-Román</i>	
The XLDB Group at the CLEF 2005 Ad-Hoc Task	54
<i>Nuno Cardoso, Leonardo Andrade, Alberto Simões, Mário J. Silva</i>	
Thomson Legal and Regulatory Experiments at CLEF-2005	61
<i>Isabelle Moulinier, Ken Williams</i>	
Using the X-IOTA System in Mono- and Bilingual Experiments at CLEF 2005	69
<i>Loïc Maisonnasse, Gilles Sérasset, Jean-Pierre Chevallet</i>	
Bilingual and Multilingual Experiments with the IR-n System	79
<i>Elisa Noguera, Fernando Llopis, Rafael Muñoz, Rafael M. Terol, Miguel A. García-Cumbreras, Fernando Martínez-Santiago, Arturo Montejo-Raez</i>	
Dictionary-Based Amharic-French Information Retrieval	83
<i>Atelach Alemu Argaw, Lars Asker, Rickard Cöster, Jussi Karlgren, Magnus Sahlgren</i>	

A Hybrid Approach to Query and Document Translation Using a Pivot Language for Cross-Language Information Retrieval	93
<i>Kazuaki Kishida, Noriko Kando</i>	
Conceptual Indexing for Multilingual Information Retrieval	102
<i>Jacques Guyot, Saïd Radhouani, Gilles Falquet</i>	
SINAI at CLEF 2005: Multi-8 Two-Years-on and Multi-8 Merging-Only Tasks	113
<i>Fernando Martínez-Santiago, Miguel A. García-Cumbreras, L.A. Ureña-López</i>	
CLEF 2005: Multilingual Retrieval by Combining Multiple Multilingual Ranked Lists	121
<i>Luo Si, Jamie Callan</i>	
Monolingual, Bilingual, and GIRT Information Retrieval at CLEF-2005	131
<i>Jacques Savoy, Pierre-Yves Berger</i>	
Socio-Political Thesaurus in Concept-Based Information Retrieval	141
<i>Mikhail Ageev, Boris Dobrov, Natalia Loukachevitch</i>	
The Performance of a Machine Translation-Based English-Indonesian CLIR System	151
<i>Mirna Adriani, Ihsan Wahyu</i>	
Exploring New Languages with HAIRCUT at CLEF 2005	155
<i>Paul McNamee</i>	
Dublin City University at CLEF 2005: Multi-8 Two-Years-On Merging Experiments	165
<i>Adenike M. Lam-Adesina, Gareth J.F. Jones</i>	
Applying Light Natural Language Processing to Ad-Hoc Cross Language Information Retrieval	170
<i>Christina Lioma, Craig Macdonald, Ben He, Vassilis Plachouras, Iadh Ounis</i>	

Monolingual Experiments

Four Stemmers and a Funeral: Stemming in Hungarian at CLEF 2005	179
<i>Anna Tordai, Maarten de Rijke</i>	

ENSM-SE at CLEF 2005: Using a Fuzzy Proximity Matching Function	187
<i>Annabelle Mercier, Amélie Imafouo, Michel Beigbeder</i>	
Bulgarian and Hungarian Experiments with Hummingbird SearchServer™ at CLEF 2005	194
<i>Stephen Tomlinson</i>	
Combining Passages in the Monolingual Task with the IR-n System	204
<i>Fernando Llopis, Elisa Noguera</i>	
Weighting Query Terms Based on Distributional Statistics	208
<i>Jussi Karlgren, Magnus Sahlgren, Rickard Cöster</i>	

Part II. Domain-Specific Information Retrieval (Domain-Specific)

Domain-Specific Track CLEF 2005: Overview of Results and Approaches, Remarks on the Assessment Analysis	212
<i>Michael Kluck, Maximilian Stempfhuber</i>	
A Baseline for NLP in Domain-Specific IR	222
<i>Johannes Leveling</i>	
Domain-Specific CLIR of English, German and Russian Using Fusion and Subject Metadata for Query Expansion	226
<i>Vivien Petras, Fredric Gey, Ray R. Larson</i>	
Evaluating a Conceptual Indexing Method by Utilizing WordNet	238
<i>Mustapha Baziz, Mohand Boughanem, Nathalie Aussenac-Gilles</i>	
Domain Specific Mono- and Bilingual English to German Retrieval Experiments with a Social Science Document Corpus	247
<i>René Hackl, Thomas Mandl</i>	

Part III. Interactive Cross-Language Information Retrieval (iCLEF)

Overview of the CLEF 2005 Interactive Track	251
<i>Julio Gonzalo, Paul Clough, Alessandro Vallin</i>	
Use of Free On-Line Machine Translation for Interactive Cross-Language Question Answering	263
<i>Angel Zaragoza, Carlos G. Fiauerola, José Luis A. Berrocal,</i>	

MIRACLE's Cross-Lingual Question Answering Experiments with Spanish as a Target Language	488
<i>César de Pablo-Sánchez, Ana González-Ledesma, José Luis Martínez-Fernández, José María Guirao, Paloma Martínez, Antonio Moreno</i>	
The Role of Lexical Features in Question Answering for Spanish	492
<i>Manuel Pérez-Coutiño, Manuel Montes-y-Gómez, Aurelio López-López, Luis Villaseñor-Pineda</i>	
Cross-Language French-English Question Answering Using the DLT System at CLEF 2005	502
<i>Richard F.E. Sutcliffe, Michael Mulcahy, Igal Gabbay, Aoife O'Gorman, Darina Slattery</i>	
Finding Answers to Indonesian Questions from English Documents	510
<i>Mirna Adriani, Rinawati</i>	
BulQA: Bulgarian-Bulgarian Question Answering at CLEF 2005	517
<i>Kiril Simov, Petya Osenova</i>	
The Query Answering System PRODICOS	527
<i>Laura Monceaux, Christine Jacquin, Emmanuel Desmontils</i>	
Part V. Cross-Language Retrieval In Image Collections (ImageCLEF)	
The CLEF 2005 Cross-Language Image Retrieval Track	535
<i>Paul Clough, Henning Müller, Thomas Deselaers, Michael Grubinger, Thomas M. Lehmann, Jeffery Jensen, William Hersh</i>	
Linguistic Estimation of Topic Difficulty in Cross-Language Image Retrieval	558
<i>Michael Grubinger, Clement Leung, Paul Clough</i>	
Dublin City University at CLEF 2005: Experiments with the ImageCLEF St Andrew's Collection	567
<i>Gareth J.F. Jones, Kieran McDonald</i>	
A Probabilistic, Text and Knowledge-Based Image Retrieval System	574
<i>Rubén Izquierdo-Beviá, David Tomás, Maximiliano Saiz-Noeda, José Luis Vicedo</i>	

UNED at ImageCLEF 2005: Automatically Structured Queries with Named Entities over Metadata	578
<i>Víctor Peinado, Fernando López-Ostenero, Julio Gonzalo, Felisa Verdejo</i>	
Easing Erroneous Translations in Cross-Language Image Retrieval Using Word Associations	582
<i>Masashi Inoue</i>	
A Corpus-Based Relevance Feedback Approach to Cross-Language Image Retrieval	592
<i>Yih-Chen Chang, Wen-Cheng Lin, Hsin-Hsi Chen</i>	
CUHK at ImageCLEF 2005: Cross-Language and Cross-Media Image Retrieval	602
<i>Steven C.H. Hoi, Jianke Zhu, Michael R. Lyu</i>	
The University of Jaén at ImageCLEF 2005: Adhoc and Medical Tasks	612
<i>M.T. Martín-Valdivia, M.A. García-Cumbreras, Miguel C. Díaz-Galiano, L.A. Ureña-López, A. Montejo-Raez</i>	
Data Fusion of Retrieval Results from Different Media: Experiments at ImageCLEF 2005	622
<i>Romarc Besançon, Christophe Millet</i>	
Combining Visual Features for Medical Image Retrieval and Annotation	632
<i>Wei Xiong, Bo Qiu, Qi Tian, Changsheng Xu, S.H. Ong, Kelvin Foong</i>	
A Structured Visual Learning Approach Mixed with Ontology Dimensions for Medical Queries	642
<i>Jean-Pierre Chevallet, Joo-Hwee Lim, Saïd Radhouani</i>	
FIRE in ImageCLEF 2005: Combining Content-Based Image Retrieval with Textual Information Retrieval	652
<i>Thomas Deselaers, Tobias Weyand, Daniel Keysers, Wolfgang Macherey, Hermann Ney</i>	
A Clustered Retrieval Approach for Categorizing and Annotating Images	662
<i>Lisa Ballesteros, Desislava Petkova</i>	

Manual Query Modification and Data Fusion for Medical Image Retrieval	673
<i>Jeffery R. Jensen, William R. Hersh</i>	
Combining Textual and Visual Features for Image Retrieval	680
<i>José Luis Martínez-Fernández, Julio Villena Román, Ana M. García-Serrano, José Carlos González-Cristóbal</i>	
Supervised Machine Learning Based Medical Image Annotation and Retrieval in ImageCLEFmed 2005	692
<i>Md. Mahmudur Rahman, Bipin C. Desai, Prabir Bhattacharya</i>	
Content-Based Retrieval of Medical Images by Combining Global Features	702
<i>Mark O Güld, Christian Thies, Benedikt Fischer, Thomas M. Lehmann</i>	
Combining Textual and Visual Features for Cross-Language Medical Image Retrieval	712
<i>Pei-Cheng Cheng, Been-Chian Chien, Hao-Ren Ke, Wei-Pang Yang</i>	
The Use of MedGIFT and EasyIR for ImageCLEF 2005	724
<i>Henning Müller, Antoine Geissbühler, Johan Marty, Christian Lovis, Patrick Ruch</i>	
Retrieving Images Using Cross-Language Text and Image Features	733
<i>Mirna Adriani, Framadhan Arnely</i>	
UB at CLEF 2005: Bilingual CLIR and Medical Image Retrieval Tasks	737
<i>Miguel E. Ruiz, Silvia B. Southwick</i>	

Part VI. Cross-Language Speech Retrieval (CL-SR)

Overview of the CLEF-2005 Cross-Language Speech Retrieval Track	744
<i>Ryen W. White, Douglas W. Oard, Gareth J.F. Jones, Dagobert Soergel, Xiaoli Huang</i>	
Using Various Indexing Schemes and Multiple Translations in the CL-SR Task at CLEF 2005	760
<i>Diana Inkpen, Muath Alzghool, Aminul Islam</i>	
The University of Alicante at CL-SR Track	769
<i>Rafael M. Terol, Manuel Palomar, Patricio Martínez-Barco,</i>	

Pitt at CLEF05: Data Fusion for Spoken Document Retrieval	773
<i>Daqing He, Jae-Wook Ahn</i>	
UNED@CL-SR CLEF 2005: Mixing Different Strategies to Retrieve Automatic Speech Transcriptions	783
<i>Fernando López-Ostenero, Víctor Pemado, Valentín Sama, Felisa Verdejo</i>	
Dublin City University at CLEF 2005: Cross-Language Speech Retrieval (CL-SR) Experiments	792
<i>Adenike M. Lam-Adesina, Gareth J.F. Jones</i>	
CLEF-2005 CL-SR at Maryland: Document and Query Expansion Using Side Collections and Thesauri	800
<i>Jianqiang Wang, Douglas W. Oard</i>	

Part VII. Multilingual Web Track (WebCLEF)

Overview of WebCLEF 2005	810
<i>Börkur Sigurbjörnsson, Jaap Kamps, Maarten de Rijke</i>	
EuroGOV: Engineering a Multilingual Web Corpus	825
<i>Börkur Sigurbjörnsson, Jaap Kamps, Maarten de Rijke</i>	
Web Retrieval Experiments with the EuroGOV Corpus at the University of Hildesheim	837
<i>Niels Jensen, René Hackl, Thomas Mandl, Robert Strötgen</i>	
Danish and Greek Web Search Experiments with Hummingbird SearchServer™ at CLEF 2005	846
<i>Stephen Tomlinson</i>	
Combination Methods for Crosslingual Web Retrieval	856
<i>Jaap Kamps, Maarten de Rijke, Börkur Sigurbjörnsson</i>	
University of Alicante at the CLEF 2005 WebCLEF Track	865
<i>Trinitario Martínez, Elisa Noguera, Rafael Muñoz, Fernando Llopis</i>	
MIRACLE at WebCLEF 2005: Combining Web Specific and Linguistic Information	869
<i>Ángel Martínez-González, José Luis Martínez-Fernández, César de Pablo-Sánchez, Julio Villena-Román</i>	

BUAP-UPV TPIRS: A System for Document Indexing Reduction at WebCLEF	873
<i>David Pinto, Héctor Jiménez-Salazar, Paolo Rosso, Emilio Sanchis</i>	
Web Page Retrieval by Combining Evidence	880
<i>Carlos G. Figuerola, José L. Alonso Berrocal, Angel F. Zazo, Emilio Rodríguez Vázquez de Aldana</i>	
UNED at WebCLEF 2005	888
<i>Javier Artiles, Víctor Peinado, Anselmo Peñas, Julio Gonzalo, Felisa Verdejo</i>	
Using the Web Information Structure for Retrieving Web Pages	892
<i>Mirna Adriani, Rama Pandugita</i>	
University of Glasgow at WebCLEF 2005: Experiments in Per-Field Normalisation and Language Specific Stemming	898
<i>Craig Macdonald, Vassilis Plachouras, Ben He, Christina Lioma, Iadh Ounis</i>	
Part VIII. Cross-Language Geographical Retrieval (GeoCLEF)	
GeoCLEF: The CLEF 2005 Cross-Language Geographic Information Retrieval Track Overview	908
<i>Fredric Gey, Ray Larson, Mark Sanderson, Hideo Joho, Paul Clough, Vivien Petras</i>	
MIRACLE at GeoCLEF 2005: First Experiments in Geographical IR	920
<i>Sara Lana-Serrano, José M. Goñi-Menoyo, José C. González-Cristóbal</i>	
University of Alicante at GeoCLEF 2005	924
<i>Óscar Ferrández, Zornitsa Kozareva, Antonio Toral, Elisa Noguera, Andrés Montoyo, Rafael Muñoz, Fernando Llopis</i>	
Evaluating Geographic Information Retrieval	928
<i>András Kornai</i>	
Using the WordNet Ontology in the GeoCLEF Geographical Information Retrieval Task	939
<i>Davide Buscaldi, Paolo Rosso, Emilio Sanchis Arnal</i>	

The GeoTALP-IR System at GeoCLEF 2005: Experiments Using a QA-Based IR System, Linguistic Analysis, and a Geographical Thesaurus	947
<i>Daniel Ferrés, Alicia Ageno, Horacio Rodríguez</i>	
CSUSM Experiments in GeoCLEF2005: Monolingual and Bilingual Tasks	956
<i>Rocio Guillén</i>	
Berkeley at GeoCLEF: Logistic Regression and Fusion for Geographic Information Retrieval	963
<i>Ray R. Larson, Fredric C. Gey, Vivien Petras</i>	
Using Semantic Networks for Geographic Information Retrieval	977
<i>Johannes Leveling, Sven Hartrumpf, Dirk Veiel</i>	
Experiments with Geo-Filtering Predicates for IR	987
<i>Jochen L. Leidner</i>	
The XLDB Group at GeoCLEF 2005	997
<i>Nuno Cardoso, Bruno Martins, Marcirio Chaves, Leonardo Andrade, Mário J. Silva</i>	
Evaluation Issues	
Portuguese at CLEF 2005	1007
<i>Diana Santos, Nuno Cardoso</i>	
Author Index	1011