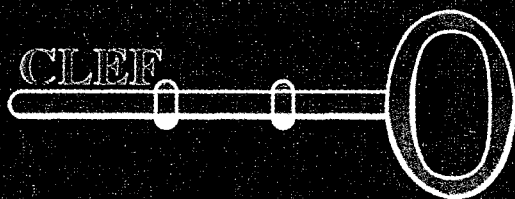


Carol Peters Paul Clough
Julio Gonzalo Gareth J. F. Jones
Michael Kluck Bernardo Magnini (Eds.)

LNCS 3491

Multilingual Information Access for Text, Speech and Images

5th Workshop of the
Cross-Language Evaluation Forum, CLEF 2004
Bath, UK, September 2004, 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.

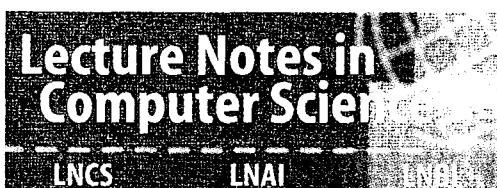
Detailed information on LNCS can be found at
<http://www.springeronline.com>

Proposals for publication should be sent to

LNCS Editorial, Tiergartenstr. 17, 69121 Heidelberg, Germany

E-mail: lncs@springer-sbm.com

ISSN 0302-9743



› springeronline.com

CLEF 2004 Steering Committee

- Maristella Agosti, University of Padua, Italy
- Eija Airio, University of Tampere, Finland
- Martin Braschler, Eurospider Information Technologies, Switzerland
- Hsin-Hsi Chen, National Taiwan University, Taipei, Taiwan
- Khalid Choukri, Evaluations and Language Resources Distribution Agency, ELDA, Paris, France
- Paul Clough, University of Sheffield, UK
- David A. Evans, Clairvoyance Corporation, USA
- Marcello Federico, ITC-irst, Trento, Italy
- Christian Fluhr, CEA-LIST, Fontenay-aux-Roses, France
- Norbert Fuhr, University of Duisburg, Germany
- Frederic C. Gey, UC Berkeley, USA
- Julio Gonzalo, LSI-UNED, Madrid, Spain
- Donna Harman, National Institute of Standards and Technology, USA
- Gareth Jones, Dublin City University, Ireland
- Franciska de Jong, University of Twente, Netherlands
- Noriko Kando, National Institute of Informatics, Tokyo, Japan
- Jussi Karlgren, Swedish Institute of Computer Science, Sweden
- Michael Kluck, Informationszentrum Sozialwissenschaften Bonn, Germany
- Natalia Loukachevitch, Moscow State University, Russia
- Bernardo Magnini, ITC-irst, Trento, Italy
- Paul McNamee, Johns Hopkins University, USA
- Henning Müller, University and University Hospitals, Geneva, Switzerland
- Douglas W. Oard, University of Maryland, USA
- Maarten de Rijke, University of Amsterdam, Netherlands
- Jacques Savoy, University of Neuchâtel, Switzerland
- Peter Schäuble, Eurospider Information Technologies, Switzerland
- Richard Sutcliffe, University of Limerick, Ireland
- Hans Uszkoreit, German Research Center for Artificial Intelligence (DFKI), Germany
- Felisa Verdejo, LSI-UNED, Madrid, Spain
- José Luis Vicedo, University of Alicante, Spain
- Ellen Voorhees, National Institute of Standards and Technology, USA
- Christa Womser-Hacker, University of Hildesheim, Germany

Table of Contents

What Happened in CLEF 2004? <i>Carol Peters</i>	1
Part I. Ad Hoc Text Retrieval Tracks	
CLEF 2004: Ad Hoc Track Overview and Results Analysis <i>Martin Braschler, Giorgio M. Di Nunzio, Nicola Ferro, Carol Peters</i>	10
Mainly Cross-Language Experiments	
Selection and Merging Strategies for Multilingual Information Retrieval <i>Jacques Savoy, Pierre-Yves Berger</i>	27
Using Surface-Syntactic Parser and Deviation from Randomness <i>Jean-Pierre Chevillet, Gilles Sérasset</i>	38
Cross-Language Retrieval Using HAIRCUT at CLEF 2004 <i>Paul McNamee, James Mayfield</i>	50
Experiments on Statistical Approaches to Compensate for Limited Linguistic Resources <i>Giorgio M. Di Nunzio, Nicola Ferro, Nicola Orio</i>	60
Application of Variable Length <i>N</i> -Gram Vectors to Monolingual and Bilingual Information Retrieval <i>Daniel Gayo-Avello, Darío Álvarez-Gutiérrez, José Gayo-Avello</i>	73
Integrating New Languages in a Multilingual Search System Based on a Deep Linguistic Analysis <i>Romaric Besançon, Olivier Ferret, Christian Fluhr</i>	83
IR-n r2: Using Normalized Passages <i>Fernando Llopis, Rafael Muñoz, Rafael M. Terol, Elisa Noguera</i>	90
Using COTS Search Engines and Custom Query Strategies at CLEF <i>David Nadeau, Mario Jarmasz, Caroline Barrière, George Foster, Claude St-Jacques</i>	100

Report on Thomson Legal and Regulatory Experiments at CLEF-2004 <i>Isabelle Moulinier, Ken Williams</i>	110
Effective Translation, Tokenization and Combination for Cross-Lingual Retrieval <i>Jaap Kamps, Sisay Fissaha Adafre, Maarten de Rijke</i>	123
Two-Stage Refinement of Transitive Query Translation with English Disambiguation for Cross-Language Information Retrieval: An Experiment at CLEF 2004 <i>Kazuaki Kishida, Noriko Kando, Kuang-Hua Chen</i>	135
Dictionary-Based Amharic -- English Information Retrieval <i>Atelach Alemu Argaw, Lars Asker, Rickard Cöster, Jussi Karlgren</i>	143
Dynamic Lexica for Query Translation <i>Jussi Karlgren, Magnus Sahlgren, Timo Järvinen, Rickard Cöster</i>	150
SINAI at CLEF 2004: Using Machine Translation Resources with a Mixed 2-Step RSV Merging Algorithm <i>Fernando Martínez-Santiago, Miguel A. García-Cumbreras, Manuel C. Díaz-Galiano, L. Alfonso Ureña</i>	156
Mono- and Crosslingual Retrieval Experiments at the University of Hildesheim <i>René Hackl, Thomas Mandl, Christa Womser-Hacker</i>	165
University of Chicago at CLEF2004: Cross-Language Text and Spoken Document Retrieval <i>Gina-Anne Levow, Irina Matveeva</i>	170
UB at CLEF2004: Cross Language Information Retrieval Using Statistical Language Models <i>Miguel E. Ruiz, Munirathnam Srikanth</i>	180
MIRACLE's Hybrid Approach to Bilingual and Monolingual Information Retrieval <i>José M. Goñi-Menoyo, José C. González, José Luis Martínez-Fernández, Julio Villena</i>	188
Searching a Russian Document Collection Using English, Chinese and Japanese Queries <i>Fredric C. Gey</i>	200

Dublin City University at CLEF 2004: Experiments in Monolingual, Bilingual and Multilingual Retrieval <i>Gareth J.F. Jones, Michael Burke, John Judge, Anna Khasin, Adenike Lam-Adesina, Joachim Wagner</i>	207
--	-----

Monolingual Experiments

Finnish, Portuguese and Russian Retrieval with Hummingbird SearchServer™ at CLEF 2004 <i>Stephen Tomlinson</i>	221
Data Fusion for Effective European Monolingual Information Retrieval <i>Jacques Savoy</i>	233
The XLDB Group at CLEF 2004 <i>Nuno Cardoso, Mário J. Silva, Miguel Costa</i>	245
The University of Glasgow at CLEF 2004: French Monolingual Information Retrieval with Terrier <i>Christina Lioma, Ben He, Vassilis Plachouras, Iadh Ounis</i>	253

Part II. Domain-Specific Document Retrieval

The Domain-Specific Track in CLEF 2004: Overview of the Results and Remarks on the Assessment Process <i>Michael Kluck</i>	260
University of Hagen at CLEF 2004: Indexing and Translating Concepts for the GIRT Task <i>Johannes Leveling, Sven Hartrumpf</i>	271
IRIT at CLEF 2004: The English GIRT Task <i>Mustapha Baziz, Mohand Boughanem, Nathalie Aussenac-Gilles</i>	283
Ricoh at CLEF 2004 <i>Yuichi Kojima</i>	292
GIRT and the Use of Subject Metadata for Retrieval <i>Walter Rada</i>	299

Part III. Interactive Cross-Language Information Retrieval

iCLEF 2004 Track Overview: Pilot Experiments in Interactive Cross-Language Question Answering <i>Julio Gonzalo, Douglas W. Oard</i>	310
Interactive Cross-Language Question Answering: Searching Passages Versus Searching Documents <i>Fernando López-Ostenero, Julio Gonzalo, Víctor Peinado, Felisa Verdejo</i>	323
Improving Interaction with the User in Cross-Language Question Answering Through Relevant Domains and Syntactic Semantic Patterns <i>Borja Navarro, Lorenza Moreno, Sonia Vázquez, Fernando Llopis, Andrés Montoyo, Miguel Ángel Varó</i>	334
Cooperation, Bookmarking, and Thesaurus in Interactive Bilingual Question Answering <i>Preben Hansen, Jussi Karlgren, Magnus Sahlgrén</i>	343
Summarization Design for Interactive Cross-Language Question Answering <i>Daqing He, Jianqiang Wang, Jun Luo, Douglas W. Oard</i>	348
Interactive and Bilingual Question Answering Using Term Suggestion and Passage Retrieval <i>Carlos G. Figuerola, Angel F. Zazo, José L. Alonso Berrocal, Emilio Rodríguez Vázquez de Aldana</i>	363

Part IV. Multiple Language Question Answering

Overview of the CLEF 2004 Multilingual Question Answering Track <i>Bernardo Magnini, Alessandro Vallin, Christelle Ayache, Gregor Erbach, Anselmo Peñas, Maarten de Rijke, Paulo Rocha, Kiril Simov, Richard F.E. Sutcliffe</i>	371
A Question Answering System for French <i>Laura Perret</i>	392
Cross-Language French-English Question Answering Using the DLT System at CLEF 2004 <i>Richard F.E. Sutcliffe, Igal Gabbay, Michael Mulcahy, Aoife O'Gorman</i>	404

Experiments on Robust NL Question Interpretation and Multi-layered Document Annotation for a Cross-Language Question/Answering System <i>Günter Neumann, Bogdan Sacaleanu</i>	411
Making Stone Soup: Evaluating a Recall-Oriented Multi-stream Question Answering System for Dutch <i>David Ahn, Valentin Jijkoun, Karin Müller, Maarten de Rijke, Stefan Schlobach, Gilad Mishne</i>	423
The DIOGENE Question Answering System at CLEF-2004 <i>Hristo Tanev, Matteo Negri, Bernardo Magnini, Milen Kouylekov</i>	435
Cross-Lingual Question Answering Using Off-the-Shelf Machine Translation <i>Kisuh Ahn, Beatrice Alex, Johan Bos, Tiphaine Dalmas, Jochen L. Leidner, Matthew B. Smillie</i>	446
Bulgarian-English Question Answering: Adaptation of Language Resources <i>Petya Osenova, Alexander Simov, Kiril Simov, Hristo Tanev, Milen Kouylekov</i>	458
Answering French Questions in English by Exploiting Results from Several Sources of Information <i>Brigitte Grau, Gabriel Illouz, Laura Monceaux, Isabelle Robba, Anne Vilnat, Guillaume Bourdil, Faïza Elkateb-Gara, Olivier Ferret, Benoît Mathieu</i>	470
Finnish as Source Language in Bilingual Question Answering <i>Lili Aunimo, Reeta Kuuskoski, Juha Makkonen</i>	482
miraQA: Experiments with Learning Answer Context Patterns from the Web <i>César de Pablo-Sánchez, José Luis Martínez-Fernández, Paloma Martínez, Julio Villena</i>	494
Question Answering for Spanish Supported by Lexical Context Annotation <i>Manuel Pérez-Coutiño, Tamar Solorio, Manuel Montes-y-Gómez, Aurelio López-López, Luis Villaseñor-Pineda</i>	502
Question Answering Using Sentence Parsing and Semantic Network Matching	

First Evaluation of Esfinge -- A Question Answering System for Portuguese <i>Luís Costa</i>	522
University of Évora in QA@CLEF-2004 <i>Paulo Quaresma, Luís Quintano, Irene Rodrigues, José Saias, Pedro Salgueiro</i>	534
COLE Experiments at QA@CLEF 2004 Spanish Monolingual Track <i>Enrique Méndez Díaz, Jesús Vilares Ferro, David Cabrero Souto</i>	544
Does English Help Question Answering in Spanish? <i>José Luis Vicedo, Maximiliano Saiz, Rubén Izquierdo, Fernando Llopis</i>	552
The TALP-QA System for Spanish at CLEF 2004: Structural and Hierarchical Relaxing of Semantic Constraints <i>Daniel Ferrés, Sumir Kanaan, Alicia Ageno, Edgar González, Horacio Rodríguez, Mihai Surdeanu, Jordi Turmo</i>	557
ILC-UniPI Italian QA <i>Francesca Bertagna, Luminita Chiran, Maria Simi</i>	569
Question Answering Pilot Task at CLEF 2004 <i>Jesús Herrera, Anselmo Peñas, Felisa Verdejo</i>	581
Evaluation of Complex Temporal Questions in CLEF-QA <i>Estela Saquete, José Luis Vicedo, Patricio Martínez-Barco, Rafael Muñoz, Fernando Llopis</i>	591
Part V. Cross-Language Retrieval in Image Collections	
The CLEF 2004 Cross-Language Image Retrieval Track <i>Paul Clough, Henning Müller, Mark Sanderson</i>	597
Caption and Query Translation for Cross-Language Image Retrieval <i>Paul Clough</i>	614
Pattern-Based Image Retrieval with Constraints and Preferences on ImageCLEF 2004 <i>Maximiliano Saiz, José Luis Vicedo, Rubén Izquierdo</i>	626

How to Visually Retrieve Images from the St. Andrews Collection Using GIFT <i>Henning Müller, Antoine Geissbühler</i>	638
UNED at ImageCLEF 2004: Detecting Named Entities and Noun Phrases for Automatic Query Expansion and Structuring <i>Víctor Peinado, Javier Artiles, Fernando López-Ostenero, Julio Gonzalo, Felisa Verdejo</i>	648
Dublin City University at CLEF 2004: Experiments with the ImageCLEF St. Andrew's Collection <i>Gareth J.F. Jones, Declan Groves, Anna Khasin, Adenike Lam-Adesina, Bart Mellebeek, Andy Way</i>	658
From Text to Image: Generating Visual Query for Image Retrieval <i>Wen-Cheng Lin, Yih-Chen Chang, Hsin-Hsi Chen</i>	664
Toward Cross-Language and Cross-Media Image Retrieval <i>Carmen Alvarez, Ahmed Id Oumohmed, Max Mignotte, Jian-Yun Nie</i>	676
FIRE -- Flexible Image Retrieval Engine: ImageCLEF 2004 Evaluation <i>Thomas Deselaers, Daniel Keysers, Hermann Ney</i>	688
MIRACLE Approach to ImageCLEF 2004: Merging Textual and Content-Based Image Retrieval <i>José Luis Martínez-Fernández, Ana García-Serrano, Julio Villena, Víctor Méndez-Sáenz</i>	698
Cross-Media Feedback Strategies: Merging Text and Image Information to Improve Image Retrieval <i>Romaric Besançon, Patrick Hède, Pierre-Alain Moellic, Christian Fluhr</i>	708
ImageCLEF 2004: Combining Image and Multi-lingual Search for Medical Image Retrieval <i>Henning Müller, Antoine Geissbühler, Patrick Ruch</i>	718
Multi-modal Information Retrieval Using FINT <i>Menno van Zaanen, Guido de Croon</i>	728
Medical Image Retrieval Using Texture, Locality and Colour <i>Peter Howarth, Alexei Yavilinsky, Daniel Heesch, Stefan Rüger</i>	740

SMIRE: Similar Medical Image Retrieval Engine <i>Pei-Cheng Cheng, Been-Chian Chien, Hao-Ren Ke, Wei-Pang Yang</i>	750
A Probabilistic Approach to Medical Image Retrieval <i>Koen Lubbers, Arjen P. de Vries, Theo Huibers, Paul van der Vet</i> ...	761
UB at CLEF2004 Cross Language Medical Image Retrieval <i>Miguel E. Ruiz, Munirathnam Srikanth</i>	773
Content-Based Queries on the CasImage Database Within the IRMA Framework <i>Christian Thies, Mark Oliver Güld, Benedikt Fischer, Thomas M. Lehmann</i>	781
Comparison and Combination of Textual and Visual Features for Interactive Cross-Language Image Retrieval <i>Pei-Cheng Cheng, Jen-Yuan Yeh, Hao-Ren Ke, Been-Chian Chien, Wei-Pang Yang</i>	793
MSU at ImageCLEF: Cross Language and Interactive Image Retrieval <i>Vineet Bansal, Chen Zhang, Joyce Y. Chai, Rong Jin</i>	805
Part VI. Cross-Language Spoken Document Retrieval	
CLEF 2004 Cross-Language Spoken Document Retrieval Track <i>Marcello Federico, Nicola Bertoldi, Gina-Anne Levow, Gareth J.F. Jones</i>	816
Part VII. Issues in CLIR and in Evaluation	
The Key to the First CLEF with Portuguese: Topics, Questions and Answers in CHAVE <i>Diana Santos, Paulo Rocha</i>	821
How Do Named Entities Contribute to Retrieval Effectiveness? <i>Thomas Mandl, Christa Womser-Hacker</i>	833
Author Index	843

What Happened in CLEF 2004?

Carol Peters

Istituto di Scienza e Tecnologie dell'Informazione (ISTI-CNR), Pisa, Italy
carol.peters@isti.cnr.it

Abstract. The organization of the CLEF 2004 evaluation campaign is described and details are provided concerning the tracks, test collections and participation. Information on new activities for CLEF 2005 is also given.

1 Introduction

This volume reports the results of the fifth in a series of annual system evaluation campaigns organised by the Cross-Language Evaluation Forum (CLEF)¹. The main objectives of CLEF are (i) to provide an infrastructure that facilitates testing of all kinds of multilingual information access systems – from monolingual retrieval for multiple languages to the implementation of complete multilingual multimedia search services, and (ii) to construct test-suites of reusable data that can be used for benchmarking purposes. These objectives are achieved through the organisation of evaluation campaigns that culminate each year in a workshop in which the groups that participated in the campaign can report and discuss their experiments. An additional aim of CLEF is to encourage contacts between the R&D and the application communities and promote the industrial take-up of research results.

The main features of the 2004 campaign are briefly outlined below in order to provide the necessary background to the experiments reported in these post-campaign proceedings.

2 Tracks and Tasks in CLEF 2004

In recent years, CLEF distinguished between the core tracks, which were those offered regularly each campaign (the monolingual, bilingual, multilingual and domain-specific tracks), and additional tracks, which were organised on an experimental basis with the objective of identifying new requirements and appropriate methodologies for their testing in a cross-language context. This distinction no longer held in 2004. The interactive track, run since 2001, was finally recognised as part of the main activity, and the great success of the pilot tracks in CLEF 2003, and in particular the cross-language question answering and image retrieval activities, led to their inclusion as regular tracks. This meant that the scope of CLEF 2004 was considerably widened with respect to previous years, with much attention being given to tasks involving information extraction (question answering) and retrieval from multimedia.

¹ CLEF 2004 is included in the activities of the DELOS Network of Excellence on Digital Libraries, funded by the Sixth Framework Programme of the European Commission. DELOS is