

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Alfred Kobsa

University of California, Irvine, CA, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

University of Dortmund, Germany

Madhu Sudan

Massachusetts Institute of Technology, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max-Planck Institute of Computer Science, Saarbruecken, Germany

Carol Peters Valentin Jijkoun
Thomas Mandl Henning Müller
Douglas W. Oard Anselmo Peñas
Vivien Petras Diana Santos (Eds.)

Advances in Multilingual and Multimodal Information Retrieval

8th Workshop of the Cross-Language Evaluation Forum,
CLEF 2007
Budapest, Hungary, September 19-21, 2007
Revised Selected Papers



Springer

Volume Editors

Carol Peters

Istituto di Scienza e Tecnologie dell'Informazione, CNR, Pisa, Italy

carol.peters@isti.cnr.it

Valentin Jijkoun

University of Amsterdam, The Netherlands; jijkoun@science.uva.nl

Thomas Mandl

University of Hildesheim, Germany; mandl@uni-hildesheim.de

Henning Müller

University of Applied Sciences Western Switzerland, Sierre, Switzerland and

University Hospitals and University of Geneva, Switzerland

henning.mueller@sim.hcuge.ch

Douglas W. Oard

University of Maryland, USA; oard@glue.umd.edu

Anselmo Peñas

Universidad Nacional de Educación a Distancia, Madrid, Spain; anselmo@lsi.uned.es

Vivien Petras

GESIS Social Science Information Centre, Bonn, Germany; vivien.petras@gesis.org

Diana Santos

Linguateca, SINTEF, Norway; diana.santos@sintef.no

Managing Editor: Danilo Giampiccolo, CELCT, Trento, Italy; giampiccolo@celct.it

Library of Congress Control Number: 2008934678

CR Subject Classification (1998): I.2.7, H.5, I.2, H.2, I.7, H.4

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

ISSN 0302-9743

ISBN-10 3-540-85759-1 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-85759-4 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media

springer.com

© Springer-Verlag Berlin Heidelberg 2008

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12465019 06/3180 5 4 3 2 1 0

Preface

The eighth campaign of the Cross Language Evaluation Forum (CLEF) for European languages was held from January to September 2007. There were seven distinct evaluation tracks in CLEF 2007, designed to test the performance of a wide range of multilingual information access systems or system components. CLEF is by now an established international evaluation initiative and, in 2007, 81 groups from all over the world submitted results for one or more of the different evaluation tracks. Full details regarding the design of the tracks, the methodologies used for evaluation, and the results obtained by the participants can be found in the different sections of these proceedings.

As always the results of the campaign were reported and discussed at the annual workshop, held in Budapest, Hungary, 19-21 September, immediately following the eleventh European Conference on Digital Libraries. The workshop plays an important role by providing the opportunity for all the groups that have participated in the evaluation campaign to get together to compare approaches and exchange ideas.

The schedule of the workshop was divided between plenary track overviews and parallel, poster and breakout sessions presenting the 2007 experiments and discussing ideas for the future. There were also several invited talks. Noriko Kando, National Institute of Informatics, Tokyo, reported the lessons learned at NTCIR-6 and plans for NTCIR-7 (NTCIR is an evaluation initiative focussed on testing IR systems for Asian languages), while Mandar Mitra, Indian Statistical Institute Kolkata, presented FIRE, a new Forum for Information Retrieval Evaluation for Indian languages. Edouard Geoffrois of the French government described the objectives of the ambitious Quaero programme, which has the goal of developing multimedia and multilingual indexing and management tools for professional and general public applications. In two final talks, Martin Braschler of the Zurich University of Applied Sciences gave a summing up of what he felt were the major trends of the 2007 campaign in the light of what had emerged during the discussions at the workshop, and Donna Harman of the US National Institute of Standards and Technology made a number of recommendations for the future. In particular, she urged the participants to focus more on failure analysis; not just to recognise what methods were working and what had little impact but to try to really understand why and then to think about generalizing what has been learnt. The presentations given at the CLEF workshop can be found on the CLEF website at www.clef-campaign.org.

CLEF 2007 was an activity of the DELOS Network of Excellence for Digital Libraries, of the Information Society Technologies programme of the European Commission. However, synergy of activities has been established with other Networks. The CLEF workshop was thus preceded by two related events. On September 18, the ImageCLEF group, together with the MUSCLE Network

of Excellence held a joint meeting on the “Evaluation of Image and Video Retrieval”. On the morning of September 19, the MorphoChallenge 2007 meeting on “Unsupervised Morpheme Analysis” was held. MorphoChallenge 2007 was part of the EU Network of Excellence PASCAL Challenge Program and was organized in collaboration with CLEF.

At the time of writing, the organisation of CLEF 2008 is well underway. The campaign this year again includes seven main evaluation tracks. Pilot tasks are also proposed to assess the performance of systems working on cross-language video retrieval and multilingual information filtering. In addition, for the first time, in 2008 CLEF offers testing with target collections in non-European languages: Arabic in the filtering track and Persian for Ad Hoc retrieval tasks. This is an important new step for CLEF; the implications will be discussed in the next workshop, to be held in Aarhus, Denmark, September 17–19, 2008. CLEF 2008 is sponsored by TrebleCLEF, a Coordination Action of the Seventh Framework Programme of the European Commission.

These post-campaign proceedings represent extended and revised versions of the initial working notes distributed at the workshop. All papers have been subjected to a reviewing procedure. The final volume has been prepared with the assistance of the Center for the Evaluation of Language and Communication Technologies (CELCT), Trento, Italy, under the coordination of Danilo Giampiccolo. The support of CELCT is gratefully acknowledged. We should also like to thank all our reviewers for their careful refereeing.

May 2008

Carol Peters
Valentin Jijkoun
Thomas Mandl
Henning Müller
Douglas W. Oard
Anselmo Peñas
Vivien Petras
Diana Santos

Reviewers

The Editors express their gratitude to the colleagues listed below for their assistance in reviewing the papers in this volume:

- Mirna Adriani, Faculty of Computer Science, University of Indonesia, Indonesia
- Stefan Baerisch, GESIS Social Science Information Centre, Bonn, Germany
- Paul D. Clough, University of Sheffield, UK
- Thomas Deselaers, RWTH Aachen University, Germany
- Thomas M. Deserno, RWTH Aachen University, Germany
- Giorgio Di Nunzio, University of Padua, Italy
- Nicola Ferro, University of Padua, Italy
- Corina Forascu, Institute for Research in Artificial Intelligence, Romania
- Fredric C. Gey, University of California at Berkeley, USA
- Ingo Glöckner, FernUniversität in Hagen, Germany
- Michael Grubinger, Victoria University, Melbourne, Australia
- Danilo Giampiccolo, Centre for the Evaluation of Human Language and Multimodal Communication Technologies (CELCT), Trento, Italy
- Rene Hackl-Sommer, University of Hildesheim and Fachinformationszentrum Karlsruhe, Germany
- Allan Hanbury, Technical University of Vienna, Austria
- Donna Harman, National Institute of Standards and Technology, USA
- Sven Hartrumpf, FernUniversität in Hagen, Germany
- William Hersh, Oregon Health and Science University, USA
- Diana Inkpen, University of Ottawa, Canada
- Gareth Jones, Dublin City University, Ireland
- Jayashree Kalpathy-Cramer, Oregon Health and Science University, USA
- Ralph Kölle, University of Hildesheim, Germany
- Mikko Kurimo, Helsinki University of Technology, Finland
- Ray Larson, University of California at Berkeley, USA
- Matthew Lease, Brown University, USA
- Johannes Leveling, FernUniversität in Hagen, Germany
- Paul McNamee, Johns Hopkins University, USA
- Diego Molla, Macquarie University, Australia
- Manuel Montes, INAOE, Mexico
- Günter Neumann, German Research Centre for Artificial Intelligence, Germany
- Petya Osenova, Bulgarian Academy of Sciences
- Ross Purves, University of Zürich - Irchel, Switzerland
- Paulo Rocha, Linguateca, Sintef, Norway
- Alvaro Rodrigo, UNED, Madrid, Spain
- Paolo Rosso, Polytechnic University of Valencia, Spain

VIII Reviewers

- Miguel Ruiz, University of North Texas, USA
- Mark Sanderson, University of Sheffield, UK
- Jacques Savoy, University of Neuchatel, Switzerland
- Mário J. Silva, University of Lisbon, Portugal
- Stephen Tomlinson, Open Text Corporation, Canada
- Jordi Turmo, Polytechnic of Catalonia, Spain
- José-Luis Vicedo, University of Alicante, Spain
- Wouter Weerkamp, University of Amsterdam, The Netherlands
- Christa Womser-Hacker, University of Hildesheim, Germany
- Xing Xie, Microsoft Research Asia, China
- Fabio Massimo Zanzotto, University of Rome “Tor Vergata”, Italy

CLEF 2007 Coordination

CLEF is coordinated by the Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa. The following Institutions contributed to the organisation of the different tracks of the CLEF 2007 campaign:

- Centre for the Evaluation of Human Language and Multimodal Communication Technologies (CELCT), Italy
- College of Information Studies and Institute for Advanced Computer Studies, University of Maryland, USA
- Department of Computer Science, University of Indonesia, Indonesia
- Department of Computer Science and Information Systems, University of Limerick, Ireland
- Department of Computer Science and Information Engineering, National University of Taiwan, Taiwan
- Department of Information Engineering, University of Padua, Italy
- Department of Information Science, University of Hildesheim, Germany
- Department of Information Studies, University of Sheffield, UK
- Department of Medical Informatics and Computer Science, RWTH Aachen University, Germany
- Evaluations and Language Resources Distribution Agency (ELDA), France
- Fondazione Bruno Kessler FBK-irst, Trento, Italy
- German Research Centre for Artificial Intelligence, DFKI, Saarbrücken, Germany
- Information and Language Processing Systems, University of Amsterdam, The Netherlands
- InformationsZentrum Sozialwissenschaften, Bonn, Germany
- Institute for Information Technology, Hyderabad, India
- Institute of Formal and Applied Linguistics, Charles University, Czech Rep.
- Universidad Nacional de Educación a Distancia, Madrid, Spain
- Linguateca, Sintef, Oslo, Norway
- Linguistic Modelling Laboratory, Bulgarian Academy of Sciences, Bulgaria
- Microsoft Research Asia
- National Institute of Standards and Technology, Gaithersburg, USA
- Department of Medical Informatics & Clinical Epidemiology, Oregon Health and Science University, USA
- Research Computing Center of Moscow State University, Russia
- Research Institute for Linguistics, Hungarian Academy of Sciences, Hungary
- School of Computing, Dublin City University, Ireland
- School of Computer Science and Mathematics, Victoria Univ., Australia
- School of Information Management and Systems, UC Berkeley, USA
- Universitat Politècnica de Catalunya, Barcelona, Spain
- University “Alexandru Ioan Cuza”, IASI, Romania
- University and University Hospitals of Geneva, Switzerland
- Vienna University of Technology, Austria

CLEF 2007 Steering Committee

- Maristella Agosti, University of Padua, Italy
- Martin Braschler, Zurich University of Applied Sciences, Switzerland
- Amedeo Cappelli, ISTI-CNR and CELCT, Italy
- Hsin-Hsi Chen, National Taiwan University, Taipei, Taiwan
- Khalid Choukri, Evaluations and Language resources Distribution Agency, Paris, France
- Paul Clough, University of Sheffield, UK
- Thomas Deselaers, RWTH Aachen University, Germany
- Giorgio Di Nunzio, University of Padua, Italy
- David A. Evans, Clairvoyance Corporation, USA
- Marcello Federico, FBK-irst, Trento, Italy
- Nicola Ferro, University of Padua, Italy
- Christian Fluhr, CEA-LIST, Fontenay-aux-Roses, France
- Norbert Fuhr, University of Duisburg, Germany
- Frederic C. Gey, U.C. 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, FBK-irst, Trento, Italy
- Paul McNamee, Johns Hopkins University, USA
- Henning Müller, University and Hospitals of Geneva, Switzerland
- Douglas W. Oard, University of Maryland, USA
- Anselmo Peñas, LSI-UNED, Madrid, Spain
- Maarten de Rijke, University of Amsterdam, Netherlands
- Diana Santos, Linguateca, Sintef, Oslo, Norway
- Jacques Savoy, University of Neuchatel, Switzerland
- Peter Schäuble, Eurospider Information Technologies, Switzerland
- Max Stempfhuber, Informationszentrum Sozialwissenschaften Bonn, Germany
- 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

Introduction

What Happened in CLEF 2007	1
<i>Carol Peters</i>	

Part I: Multilingual Textual Document Retrieval (Ad Hoc)

CLEF 2007: Ad Hoc Track Overview	13
<i>Giorgio M. Di Nunzio, Nicola Ferro, Thomas Mandl, and Carol Peters</i>	

Monolingual

Charles University at CLEF 2007 Ad-Hoc Track	33
<i>Pavel Češka and Pavel Pecina</i>	

Stemming Approaches for East European Languages	37
<i>Ljiljana Dolamic and Jacques Savoy</i>	

Applying Query Expansion Techniques to Ad Hoc Monolingual Tasks with the IR-n System	45
<i>Elisa Noguera and Fernando Llopis</i>	

Bulgarian, Hungarian and Czech Stemming Using YASS	49
<i>Prasenjit Majumder, Mandar Mitra, and Dipasree Pal</i>	

Sampling Precision to Depth 10000 at CLEF 2007	57
<i>Stephen Tomlinson</i>	

Cross-Language: European

Disambiguation and Unknown Term Translation in Cross Language Information Retrieval	64
<i>Dong Zhou, Mark Truran, and Tim Brailsford</i>	

Cross-Language Retrieval with Wikipedia	72
<i>Péter Schönhofen, András Benczúr, István Bíró, and Károly Csalogány</i>	

Cross-Language: Non-European

Cross-Lingual Information Retrieval System for Indian Languages	80
<i>Jagadeesh Jagarlamudi and A. Kumaran</i>	

Bengali, Hindi and Telugu to English Ad-Hoc Bilingual Task at CLEF 2007	88
<i>Sivaji Bandyopadhyay, Tapabrata Mondal, Sudip Kumar Naskar, Asif Ekbal, Rejuwanul Haque, and Srinivasa Rao Godhavarthy</i>	
Bengali and Hindi to English CLIR Evaluation	95
<i>Debasis Mandal, Mayank Gupta, Sandipan Dandapat, Pratyush Banerjee, and Sudeshna Sarkar</i>	
Improving Recall for Hindi, Telugu, Oromo to English CLIR	103
<i>Prasad Pingali, Kula Kekeba Tune, and Vasudeva Varma</i>	
Hindi to English and Marathi to English Cross Language Information Retrieval Evaluation	111
<i>Manoj Kumar Chinnakotla, Sagar Ranadive, Om P. Damani, and Pushpak Bhattacharyya</i>	
Amharic-English Information Retrieval with Pseudo Relevance Feedback	119
<i>Atelach Alemu Argaw</i>	
Indonesian-English Transitive Translation for Cross-Language Information Retrieval	127
<i>Mirna Adriani, Herika Hayurani, and Syandra Sari</i>	
Robust	
Robust Retrieval Experiments at the University of Hildesheim	134
<i>Ben Heuwing and Thomas Mandl</i>	
SINAI at CLEF Ad-Hoc Robust Track 2007: Applying Google Search Engine for Robust Cross-Lingual Retrieval	137
<i>F. Martínez-Santiago, A. Montejo-Ráez, and M.Á. García-Cumbreras</i>	
Improving Robustness Using Query Expansion	143
<i>Angel F. Zazo, José L. Alonso Berrocal, and Carlos G. Figuerola</i>	
English-to-French CLIR: A Knowledge-Light Approach through Character <i>N</i> -Grams Alignment	148
<i>Jesús Vilares, Michael P. Oakes, and Manuel Vilares</i>	
MIRACLE Progress in Monolingual Information Retrieval at Ad-Hoc CLEF 2007	156
<i>José-Carlos González-Cristóbal, José Miguel Goñi-Menoyo, Julio Villena-Román, and Sara Lana-Serrano</i>	
Part II: Domain-Specific Information Retrieval (Domain-Specific)	
The Domain-Specific Track at CLEF 2007	160
<i>Vivien Petras, Stefan Baierisch, and Maximillian Stempfhuber</i>	

The XTRIEVAL Framework at CLEF 2007: Domain-Specific Track	174
<i>Jens Kürsten, Thomas Wilhelm, and Maximilian Eibl</i>	
Query Translation through Dictionary Adaptation	182
<i>Stephane Clinchant and Jean-Michel Renders</i>	
Experiments in Classification Clustering and Thesaurus Expansion for Domain Specific Cross-Language Retrieval	188
<i>Ray R. Larson</i>	
Domain-Specific IR for German, English and Russian Languages	196
<i>Claire Fautsch, Ljiljana Dolamic, Samir Abdou, and Jacques Savoy</i>	

Part III: Multiple Language Question Answering (QA@CLEF)

Overview of the CLEF 2007 Multilingual Question Answering Track	200
<i>Danilo Giampiccolo, Pamela Forner, Jesús Herrera, Anselmo Peñas, Christelle Ayache, Corina Forascu, Valentin Jijkoun, Petya Osenova, Paulo Rocha, Bogdan Sacaleanu, and Richard Sutcliffe</i>	
Overview of the Answer Validation Exercise 2007	237
<i>Anselmo Peñas, Álvaro Rodrigo, and Felisa Verdejo</i>	
Overview of QAST 2007	249
<i>Jordi Turmo, Pere R. Comas, Christelle Ayache, Djamel Mostefa, Sophie Rosset, and Lori Lamel</i>	

Main Task: Mono- and Bilingual QA

Question Answering with Joost at CLEF 2007	257
<i>Gosse Bouma, Geert Kloosterman, Jori Mur, Gertjan van Noord, Lonneke van der Plas, and Jörg Tiedemann</i>	
What Happened to Esfinge in 2007?	261
<i>Luís Miguel Cabral, Luís Fernando Costa, and Diana Santos</i>	
Coreference Resolution for Questions and Answer Merging by Validation	269
<i>Sven Hartrumpf, Ingo Glöckner, and Johannes Leveling</i>	
Multilingual Question Answering through Intermediate Translation: LCC's PowerAnswer at QA@CLEF 2007	273
<i>Mitchell Bowden, Marian Olteanu, Pasin Suriyentrakorn, Thomas d'Silva, and Dan Moldovan</i>	
RACAI's Question Answering System at QA@CLEF2007	284
<i>Dan Tufiș, Dan Stefanescu, Radu Ion, and Alexandru Ceașu</i>	

DFKI-LT at QA@CLEF 2007	292
<i>Bogdan Sacaleanu, Günter Neumann, and Christian Spurk</i>	
University of Wolverhampton at CLEF 2007.....	300
<i>Georgiana Pușcașu and Constantin Orăsan</i>	
Bilingual Question Answering Using CINDI-QA at QA@CLEF 2007....	308
<i>Chedid Haddad and Bipin C. Desai</i>	
The University of Évora's Participation in QA@CLEF-2007	316
<i>José Saias and Paulo Quaresma</i>	
Web-Based Anaphora Resolution for the QUASAR Question Answering System	324
<i>Davide Buscaldi, Yassine Benajiba, Paolo Rosso, and Emilio Sanchis</i>	
A Lexical Approach for Spanish Question Answering	328
<i>Alberto Téllez, Antonio Juárez, Gustavo Hernández, Claudia Denicia, Esaú Villatoro, Manuel Montes, and Luis Villaseñor</i>	
Finding Answers Using Resources in the Internet	332
<i>Mirna Adriani and Septian Adiwibowo</i>	
UAIC Romanian QA System for QA@CLEF	336
<i>Adrian Iftene, Diana Trandabăt, Ionuț Pistol, Alex Moruz, Alexandra Balahur, Diana Cotelea, Iustin Dornescu, Iuliana Drăghici, and Dan Cristea</i>	
The University of Amsterdam's Question Answering System at QA@CLEF 2007	344
<i>Valentin Jijkoun, Katja Hofmann, David Ahn, Mahboob Alam Khalid, Joris van Rantwijk, Maarten de Rijke, and Erik Tjong Kim Sang</i>	
Combining Wikipedia and Newswire Texts for Question Answering in Spanish	352
<i>César de Pablo-Sánchez, José L. Martínez-Fernández, Ana González-Ledesma, Doaa Samy, Paloma Martínez, Antonio Moreno-Sandoval, and Harith Al-Jumaily</i>	
QA@L ² F, First Steps at QA@CLEF	356
<i>Ana Mendes, Luísa Coheur, Nuno J. Mamede, Ricardo Ribeiro, Fernando Batista, and David Martins de Matos</i>	
Priberam's Question Answering System in QA@CLEF 2007	364
<i>Carlos Amaral, Adán Cassan, Helena Figueira, André Martins, Afonso Mendes, Pedro Mendes, Cláudia Pinto, and Daniel Vidal</i>	

Answer Validation Exercise (AVE)

Combining Logic and Aggregation for Answer Selection	372
<i>Ingo Glöckner</i>	
On the Application of Lexical-Syntactic Knowledge to the Answer Validation Exercise	377
<i>Óscar Ferrández, Daniel Micol, Rafael Muñoz, and Manuel Palomar</i>	
Combining Lexical Information with Machine Learning for Answer Validation at QA@CLEF 2007	381
<i>M.Á. García-Cumbreras, J.M. Perea-Ortega, F. Martínez-Santiago, and L. Alfonso Ureña-López</i>	
Using Recognizing Textual Entailment as a Core Engine for Answer Validation	387
<i>Rui Wang and Günter Neumann</i>	
A Supervised Learning Approach to Spanish Answer Validation	391
<i>Alberto Téllez-Valero, Manuel Montes-y-Gómez, and Luis Villaseñor-Pineda</i>	
UAIC Participation at AVE 2007	395
<i>Adrian Iftene and Alexandra Balahur-Dobrescu</i>	
UNED at Answer Validation Exercise 2007	404
<i>Álvaro Rodrigo, Anselmo Peñas, and Felisa Verdejo</i>	

Question Answering on Speech Transcription (QAST)

Adapting QA Components to Mine Answers in Speech Transcripts	410
<i>Günter Neumann and Rui Wang</i>	
The LIMSI Participation in the QAst Track	414
<i>Sophie Rosset, Olivier Galibert, Gilles Adda, and Eric Bilinski</i>	
Robust Question Answering for Speech Transcripts Using Minimal Syntactic Analysis	424
<i>Pere R. Comas, Jordi Turmo, and Mihai Surdeanu</i>	

Part IV: Cross-Language Retrieval in Image Collections (ImageCLEF)

Overview of the ImageCLEFphoto 2007 Photographic Retrieval Task	433
<i>Michael Grubinger, Paul Clough, Allan Hanbury, and Henning Müller</i>	

Overview of the ImageCLEF 2007 Object Retrieval	445
<i>Thomas Deselaers, Allan Hanbury, Ville Viitaniemi, András Benczúr, Mátyás Brendel, Bálint Daróczy, Hugo Jair Escalante Balderas, Theo Gevers, Carlos Arturo Hernández Gracidas, Steven C.H. Hoi, Jorma Laaksonen, Mingjing Li, Heidy Marisol Marín Castro, Hermann Ney, Xiaoguang Rui, Nicu Sebe, Julian Stöttinger, and Lei Wu</i>	
Overview of the ImageCLEFmed 2007 Medical Retrieval and Medical Annotation Tasks	472
<i>Henning Müller, Thomas Deselaers, Thomas M. Deserno, Jayashree Kalpathy-Cramer, Eugene Kim, and William Hersch</i>	
ImageCLEFphoto	
FIRE in ImageCLEF 2007: Support Vector Machines and Logistic Models to Fuse Image Descriptors for Photo Retrieval	492
<i>Tobias Gass, Tobias Weyand, Thomas Deselaers, and Hermann Ney</i>	
MIRACLE at ImageCLEFphoto 2007: Evaluation of Merging Strategies for Multilingual and Multimedia Information Retrieval	500
<i>Julio Villena-Román, Sara Lana-Serrano, José Luis Martínez-Fernández, and José Carlos González-Cristóbal</i>	
Using an Image-Text Parallel Corpus and the Web for Query Expansion in Cross-Language Image Retrieval	504
<i>Yih-Chen Chang and Hsin-Hsi Chen</i>	
SINAI System: Combining IR Systems at ImageCLEFPhoto 2007	512
<i>M.Á. García-Cumbreras, M.C. Díaz-Galiano, M.T. Martín-Valdivia, A. Montejo-Ráez, and L.A. Ureña-López</i>	
Multimodal Retrieval by Text-Segment Biclustering	518
<i>András Benczúr, István Bíró, Mátyás Brendel, Károly Csalogány, Bálint Daróczy, and Dávid Siklósi</i>	
Analysing an Approach to Information Retrieval of Visual Descriptions with IR-n, a System Based on Passages	522
<i>Sergio Navarro, Fernando Llopis, Rafael Muñoz Guillena, and Elisa Noguera</i>	
DCU and UTA at ImageCLEFPhoto 2007	530
<i>Anni Järvelin, Peter Wilkins, Tomasz Adamek, Eija Airio, Gareth J.F. Jones, Alan F. Smeaton, and Eero Sormunen</i>	

Cross-Language and Cross-Media Image Retrieval: An Empirical Study at ImageCLEF2007	538
<i>Steven C.H. Hoi</i>	
Towards Annotation-Based Query and Document Expansion for Image Retrieval	546
<i>Hugo Jair Escalante, Carlos Hernández, Aurelio López, Heidy Marín, Manuel Montes, Eduardo Morales, Enrique Sucar, and Luis Villaseñor</i>	
Content-Based Image Retrieval Using Combined 2D Attribute Pattern Spectra	554
<i>Florence Tushabe and Michael. H.F. Wilkinson</i>	
Text-Based Clustering of the ImageCLEFphoto Collection for Augmenting the Retrieved Results	562
<i>Osama El Demerdash, Leila Kosseim, and Sabine Bergler</i>	
Trans-Media Pseudo-Relevance Feedback Methods in Multimedia Retrieval	569
<i>Stephane Clinchant, Jean-Michel Renders, and Gabriela Csurka</i>	
ImageCLEFmed	
Cue Integration for Medical Image Annotation	577
<i>Tatiana Tommasi, Francesco Orabona, and Barbara Caputo</i>	
Multiplying Concept Sources for Graph Modeling	585
<i>Loïc Maisonnasse, Eric Gaussier, and Jean Pierre Chevallet</i>	
MIRACLE at ImageCLEFmed 2007: Merging Textual and Visual Strategies to Improve Medical Image Retrieval	593
<i>Julio Villena-Román, Sara Lana-Serrano, and José Carlos González-Cristóbal</i>	
MIRACLE at ImageCLEFmed 2007: Machine Learning Experiments on Medical Image Annotation	597
<i>Sara Lana-Serrano, Julio Villena-Román, José Carlos González-Cristóbal, and José Miguel Goñi-Menoyo</i>	
Integrating MeSH Ontology to Improve Medical Information Retrieval	601
<i>M.C. Díaz-Galiano, M.Á. García-Cumbreras, M.T. Martín-Valdivia, A. Montejo-Ráez, and L.A. Ureña-López</i>	
Speeding Up IDM without Degradation of Retrieval Quality	607
<i>Michael Springmann and Heiko Schuldt</i>	

XVIII Table of Contents

Content-Based Medical Image Retrieval Using Low-Level Visual Features and Modality Identification	615
<i>Juan C. Caicedo, Fabio A. Gonzalez, and Eduardo Romero</i>	
Medical Image Retrieval and Automatic Annotation: OHSU at ImageCLEF 2007	623
<i>Jayashree Kalpathy-Cramer and William Hersh</i>	
Using Bayesian Network for Conceptual Indexing: Application to Medical Document Indexing with UMLS Metathesaurus	631
<i>Thi Hoang Diem Le, Jean-Pierre Chevallet, and Joo Hwee Lim</i>	
Baseline Results for the ImageCLEF 2007 Medical Automatic Annotation Task Using Global Image Features	637
<i>Mark O. Güld and Thomas M. Deserno</i>	
Evaluation of Automatically Assigned MeSH Terms for Retrieval of Medical Images	641
<i>Miguel E. Ruiz and Aurélie Névéol</i>	
ImageCLEF photo and med	
University and Hospitals of Geneva Participating at ImageCLEF 2007	649
<i>Xin Zhou, Julien Gobet, Patrick Ruch, and Henning Müller</i>	
An Interactive and Dynamic Fusion-Based Image Retrieval Approach by CINDI	657
<i>M.M. Rahman, B.C. Desai, and P. Bhattacharya</i>	
Using Pseudo-Relevance Feedback to Improve Image Retrieval Results	665
<i>Mouna Torjmen, Karen Pinel-Sauvagnat, and Mohand Boughanem</i>	
Part V: Cross-Language Speech Retrieval (CL-SR)	
Overview of the CLEF-2007 Cross-Language Speech Retrieval Track	674
<i>Pavel Pecina, Petra Hoffmannová, Gareth J.F. Jones, Ying Zhang, and Douglas W. Oard</i>	
A Dirichlet-Smoothed Bigram Model for Retrieving Spontaneous Speech	687
<i>Matthew Lease and Eugene Charniak</i>	
Model Fusion Experiments for the CLSR Task at CLEF 2007	695
<i>Muath Alzghool and Diana Inkpen</i>	

Dublin City University at CLEF 2007: Cross-Language Speech Retrieval Experiments	703
<i>Ying Zhang, Gareth J.F. Jones, and Ke Zhang</i>	
What Can and Cannot Be Found in Czech Spontaneous Speech Using Document-Oriented IR Methods—UWB at CLEF 2007 CL-SR Track.....	712
<i>Pavel Ircing, Josef Psutka, and Jan Vavruška</i>	
Using Information Gain to Filter Information in CLEF CL-SR Track ...	719
<i>M.C. Díaz-Galiano, M.T. Martín-Valdivia, M.Á. García-Cumbreras, and L.A. Ureña-López</i>	

Part VI: Multilingual Web Retrieval (WebCLEF)

Overview of WebCLEF 2007	725
<i>Valentin Jijkoun and Maarten de Rijke</i>	
Segmentation of Web Documents and Retrieval of Useful Passages	732
<i>Carlos G. Figuerola, José L. Alonso Berrocal, and Angel F. Zazo Rodríguez</i>	
Using Centrality to Rank Web Snippets.....	737
<i>Valentin Jijkoun and Maarten de Rijke</i>	
Using Web-Content for Retrieving Snippets	742
<i>Okky Hendriansyah, Tri Firgantoro, and Mirna Adriani</i>	

Part VII: Cross-Language Geographical Retrieval (GeoCLEF)

GeoCLEF 2007: The CLEF 2007 Cross-Language Geographic Information Retrieval Track Overview	745
<i>Thomas Mandl, Fredric Gey, Giorgio Di Nunzio, Nicola Ferro, Ray Larson, Mark Sanderson, Diana Santos, Christa Womser-Hacker, and Xing Xie</i>	
Inferring Location Names for Geographic Information Retrieval	773
<i>Johannes Leveling and Sven Hartrumpf</i>	
GeoParsing Web Queries	781
<i>Rocio Guillén</i>	
MIRACLE at GeoCLEF Query Parsing 2007: Extraction and Classification of Geographical Information.....	786
<i>Sara Lana-Serrano, Julio Villena-Román, José Carlos González-Cristóbal, and José Miguel Goñi-Menoyo</i>	

Relevance Measures Using Geographic Scopes and Types	794
<i>Geoffrey Andogah and Gosse Bouma</i>	
Using Geographic Signatures as Query and Document Scopes in Geographic IR	802
<i>Nuno Cardoso, David Cruz, Marcirio Chaves, and Mário J. Silva</i>	
Cheshire at GeoCLEF 2007: Retesting Text Retrieval Baselines	811
<i>Ray R. Larson</i>	
On the Relative Importance of Toponyms in GeoCLEF	815
<i>Davide Buscaldi and Paolo Rosso</i>	
Filtering for Improving the Geographic Information Search	823
<i>José M. Perea-Ortega, Miguel A. García-Cumbreras, Manuel García-Vega, and L.A. Ureña-López</i>	
TALP at GeoCLEF 2007: Results of a Geographical Knowledge Filtering Approach with Terrier	830
<i>Daniel Ferrés and Horacio Rodríguez</i>	
TALP at GeoQuery 2007: Linguistic and Geographical Analysis for Query Parsing	834
<i>Daniel Ferrés and Horacio Rodríguez</i>	
Applying Geo-feedback to Geographic Information Retrieval	838
<i>Mirna Adriani and Nasikhin</i>	
Exploring LDA-Based Document Model for Geographic Information Retrieval	842
<i>Zhisheng Li, Chong Wang, Xing Xie, Xufa Wang, and Wei-Ying Ma</i>	
Mono- and Crosslingual Retrieval Experiments with Spatial Restrictions at GeoCLEF 2007	850
<i>Ralph Kölle, Ben Heuwing, Thomas Mandl, and Christa Womser-Hacker</i>	
GIR Experiments with Forostar	856
<i>Simon Overell, João Magalhães, and Stefan Rüger</i>	

Part VIII: CLEF in Other Evaluations

CLEF at MorphoChallenge

Morpho Challenge Evaluation Using a Linguistic Gold Standard	864
<i>Mikko Kurimo, Mathias Creutz, and Matti Varjokallio</i>	
Simple Morpheme Labelling in Unsupervised Morpheme Analysis	873
<i>Delphine Bernhard</i>	

Unsupervised and Knowledge-Free Morpheme Segmentation and Analysis <i>Stefan Bordag</i>	881
Unsupervised Acquiring of Morphological Paradigms from Tokenized Text <i>Daniel Zeman</i>	892
ParaMor: Finding Paradigms across Morphology <i>Christian Monson, Jaime Carbonell, Alon Lavie, and Lori Levin</i>	900
CLEF at SemEval 2007	
SemEval-2007 Task 01: Evaluating WSD on Cross-Language Information Retrieval <i>Eneko Agirre, Oier Lopez de Lacalle, Bernardo Magnini, Arantxa Otegi, German Rigau, and Piek Vossen</i>	908
Author Index	919