Lecture Notes in Computer Science 6241

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
TU Dortmund University, Germany

Madhu Sudan
Microsoft Research, Cambridge, MA, USA

Demetri Terzopoulos
University of California, Los Angeles, CA, USA

Doug Tygar
University of California, Berkeley, CA, USA

Gerhard Weikum
Max Planck Institute for Informatics, Saarbruecken, Germany

Carol Peters Giorgio Maria Di Nunzio
Mikko Kurimo Thomas Mandl Djamel Mostefa
Anselmo Peñas Giovanna Roda (Eds.)

Multilingual Information Access Evaluation I

Text Retrieval Experiments

10th Workshop of the Cross-Language Evaluation Forum, CLEF 2009
Corfu, Greece, September 30 - October 2, 2009
Revised Selected Papers

Springer

Volume Editors

Carol Peters
ISTI-CNR, Area Ricerca CNR
56124 Pisa, Italy
E-mail: carol.peters@isti.cnr.it

Giorgio Maria Di Nunzio
University of Padua
35131 Padova, Italy
E-mail: dinunzio@dei.unipd.it

Mikko Kurimo
Aalto University
00076 Aalto, Finland
E-mail: mikko.kurimo@tkk.fi

Thomas Mandl
University of Hildesheim
31141 Hildesheim, Germany
E-mail: mandl@uni-hildesheim.de

Djamel Mostefa
ELDA/ELRA, 75013 Paris, France
E-mail: mostefa@elda.org

Anselmo Peñas
LSI-UNED, 28040 Madrid, Spain
E-mail: anselmo@lsi.uned.es

Giovanna Roda
Matrixware, 1060 Vienna, Austria
E-mail: giovanna.roda@gmail.com

Managing Editors

Pamela Forner and
Danilo Giampiccolo
CELCT, Trento, Italy
Email: {forner; giampiccolo}@celct.it

Library of Congress Control Number: 2010934130

CR Subject Classification (1998): I.2.7, H.3, H.4, H.2, H.5, I.7

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

ISSN 0302-9743
ISBN-10 3-642-15753-X Springer Berlin Heidelberg New York
ISBN-13 978-3-642-15753-0 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.com

© Springer-Verlag Berlin Heidelberg 2010
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180

Preface

The tenth campaign of the Cross Language Evaluation Forum (CLEF) for European languages was held from January to September 2009. There were eight main evaluation tracks in CLEF 2009 plus a pilot task. The aim, as usual, was to test the performance of a wide range of multilingual information access (MLIA) systems or system components. This year, about 150 groups, mainly but not only from academia, registered to participate in the campaign. Most of the groups were from Europe but there was also a good contingent from North America and Asia.

The results were presented at a two-and-a-half day workshop held in Corfu, Greece, September 30 to October 2, 2009, in conjunction with the European Conference on Digital Libraries. The workshop, attended by 160 researchers and system developers, provided the opportunity for all the groups that had participated in the evaluation campaign to get together, compare approaches and exchange ideas.

The schedule was divided between plenary track overviews, and parallel, poster and breakout sessions presenting the CLEF 2009 experiments and discussing ideas for the future. There were several invited talks. Noriko Kando, National Institute of Informatics, Tokyo, reported on the evolution of NTCIR (NTCIR is an evaluation initiative focussed on testing information access technologies for Asian languages), and Jaap Kamps of the University of Amsterdam presented the main outcomes of a SIGIR workshop on the "Future of IR Evaluation." In the final session, Donna Harman, US National Institute of Standards and Technology, summed up what she felt were the main achievements of CLEF over these ten years of activity. The presentations given at the CLEF workshop can be found on the CLEF website at www.clef-campaign.org.

The workshop was preceded by two related events. On September 29, a one-day Workshop on Visual Information Retrieval Evaluation was held. This workshop was sponsored by the THESEUS program and co-organized by the Fraunhofer Institute for Digital Media Technology. The participants discussed the results of the ImageCLEF initiative and identified new challenging image retrieval and analysis tasks for future evaluations. The MorphoChallenge 2009 meeting on "Unsupervised Morpheme Analysis" was held on the morning of September 30. The objective of this year's challenge was to design a statistical machine learning algorithm for morpheme discovery. MorphoChallenge is part of the EU Network of Excellence PASCAL Programme.

The CLEF 2008 and 2009 campaigns were organized by TrebleCLEF, a Coordination Action of the Seventh Framework Programme. TrebleCLEF has built on the results achieved by CLEF, supporting the development of expertise in the multidisciplinary research area of multilingual information access and promoting a dissemination action in the relevant application communities. As part of its activities, the project has released a set of Best Practice recommendations in the areas of MLIA System Development and Search Assistance, Test Collection Creation, and Language Processing Technologies. The results of TrebleCLEF can be accessed at www.trebleclef.eu.

This is the first time that the CLEF proceedings are published in two volumes reporting the results of the Text Retrieval Experiments and the Multimedia Experiments, separately. This decision was made necessary by the large participation in CLEF 2009 and our desire to provide an exhaustive overview of all the various activities. This volume reports research and experiments on various types of textual document collections. It is divided into six main sections presenting the results of the following tracks: Multilingual Document Retrieval (Ad-Hoc), Multiple Language Question Answering (QA@CLEF), Multilingual Information Filtering (INFILE@CLEF), Intellectual Property (CLEF-IP) and Log File Analysis (LogCLEF), plus the activities of the Morpho-Challenge program. The companion volume contains the results of the remaining three tracks running on multimedia data: Interactive Cross-Language Retrieval (iCLEF), Cross-Language Image Retrieval (ImageCLEF), and Cross-Language Video Retrieval (VideoCLEF). The table of contents is included in this volume. The papers are mostly extended and revised versions of the initial working notes distributed at the workshop. All papers were subjected to a reviewing procedure. The final volumes were prepared with assistance of the Center for the Evaluation of Language and Communication Technologies (CELCT), Trento, Italy, under the coordination of Danilo Giampiccolo and Pamela Forner. The support of CELCT is gratefully acknowledged. We should also like to thank all the additional reviewers for their careful refereeing.

April 2010

Carol Peters
Giorgio Maria Di Nunzio
Mikko Kurimo
Thomas Mandl
Djamel Mostefa
Anselmo Peñas
Giovanna Roda

Reviewers

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

- Eneko Agirre, University of the Basque Country, Spain
- Abolfazl AleAhmad, University of Tehran, Iran
- Iñaki Alegria, University of the Basque Country, Spain
- Giambattista Amati, Fondazione Ugo Bordoni, Italy
- Ebru Arisoy, Bogazici University, Turkey
- Victoria Arranz, ELDA, France
- Sören Auer, University of Leipzig, Germany
- Helmut Berger, Matrixware, Austria
- Delphine Bernhard, LIMSI-CNRS, France
- Romaric Besançon, CEA-LIST, France
- Alessio Bosca, Celi s.r.l., Italy
- Gosse Bouma, Rijksuniversiteit Groningen, The Netherlands
- Burcu Can, University of York, UK
- Stéphane Chaudiron, Université de Lille 3, France
- Tolga Ciloglu, Middle East Technical University, Turkey
- Cagri Coltekin, University of Groningen, The Netherlands
- Luis Fernando Costa, SINTEF ICT, Portugal
- Michael Dittenbach, Matrixware, Austria
- Nicola Ferro, University of Padua, Italy
- Corina Forâscu, Al.I. Cuza University of Iasi, Romania
- M. Rami Ghorab, Trinity College, Ireland
- Ingo Glöckner, FernUniversität in Hagen, Germany
- Erik Graf, University of Glasgow, UK
- Harald Hammarström, Chalmers University, Sweden
- Olivier Hamon, ELDA, France
- Allan Hanbury, Information Retrieval Facility, Austria
- Sven Hartrumpf, FernUniversität in Hagen - IICS, Germany
- Jim Jansen, The Pennsylvania State University, USA
- Kalervo Jarvelin, University of Tampere, Finland
- Dietrich Klakow, Saarland University, Germany
- Oskar Kohonen, Aalto University, Finland
- Katrin Lamm, University of Hildesheim, Germany
- Ray Larson, University of California at Berkeley, USA
- Johannes Leveling, Dublin City University, Ireland
- Constantine Lignos, University of Pennsylvania, USA

- Mihai Lupu, Information Retrieval Facility, Austria
- Thomas Mandl, University of Hildesheim, Germany
- Diego Molla, Macquarie University, Australia
- Christian Monson, Oregon Health & Science University, USA
- Nicolas Moreau, ELDA, France
- Michael Oakes, University of Sunderland, UK
- Constantin Orasan, University of Wolverhampton, UK
- Vivien Petras, Humboldt University, Germany
- Florina Piroi, Information Retrieval Facility, Austria
- Horacio Rodríguez, Polytechnic University of Catalonia, Spain
- Paolo Rosso, Polytechnic University of Valencia, Spain
- Erik Tjong Kim Sang, University of Groningen, Netherlands
- Murat Saraclar, Bogazici University, Turkey
- Julia Maria Schulz, University of Hildesheim, Germany
- Sebastian Spiegler, University of Bristol, UK
- John Tait, Information Retrieval Facility, Austria
- Jordi Turmo, Polytechnic University of Catalonia, Spain
- Jose Luis Vicedo, University of Alicante, Spain
- Sami Virpioja, Aalto University, Finland
- Christa Womser-Hacker, University of Hildesheim, Germany
- Alex Yeh, The MITRE Corporation, USA
- Daniel Zeman, Charles University, Czech Republic
- Veronika Zenz, Matrixware, Austria

CLEF 2009 Coordination

CLEF 2000–2009 was coordinated by the Istituto di Scienza e Tecnologie dell' Informazione, Consiglio Nazionale delle Ricerche, Pisa. The following institutions contributed to the organization of the different tracks of the 2009 campaign:

- Adaptive Informatics Research Centre, Helsinki University of Technology, Finland
- Berlin School of Library and Information Science, Humboldt University, Germany
- Business Information Systems, University of Applied Sciences Western Switzerland, Sierre, Switzerland
- CEA LIST, France
- Center for Autonomous Systems, Royal Institute of Technology, Sweden
- Center for Evaluation of Language and Communication Technologies, Italy
- Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
- Computer Science Department, University of the Basque Country, Spain
- Computer Vision and Multimedia Lab, University of Geneva, Switzerland
- Database Research Group, University of Tehran, Iran
- Department of Computer Science & Information Systems, University of Limerick, Ireland
- 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 Clinical Epidemiology, Oregon Health and Science University, USA
- Department of Medical Informatics, Aachen University of Technology, Germany
- Evaluations and Language Resources Distribution Agency Sarl, Paris, France
- Fraunhofer Institute for Digital Media Technology (IDMT), Germany
- GERiiCO, Université de Lille, France
- Idiap Research Institute, Switzerland
- Information Retrieval Facility (IRF), Austria
- Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI), Orsay, France
- Leiden Institute of Advanced Computer Science, Leiden University, The Netherlands
- Lenguajes y Sistemas Informáticos, Universidad Nacional de Educación a Distancia, Madrid, Spain
- Linguateca, SINTEF ICT, Norway
- Linguistic Modelling Laboratory, Bulgarian Academy of Sciences, Bulgaria
- Matrixware Information Services, Austria

- Mediamatics, Delft University of Technology, The Netherlands
- Medical Informatics Service, University Hospitals and University of Geneva, Switzerland
- MITRE Corporation, USA
- National Institute of Standards and Technology, Gaithersburg MD, USA
- NLE Lab., Universidad Politènica de Valencia, Spain
- Research Institute for Artificial Intelligence, Romanian Academy, Romania
- Romanian Institute for Computer Science, Romania
- Royal Institute of Technology (KTH), Stockholm, Sweden
- School of Computing, Dublin City University, Ireland
- Swedish Institute of Computer Science, Sweden
- University of Applied Sciences Western Switzerland (HES-SO), Switzerland

CLEF 2009 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, ETH, Switzerland
- Giorgio Di Nunzio, University of Padua, Italy
- David A. Evans, Clairvoyance Corporation, USA
- Marcello Federico, Fondazione Bruno Kessler, Trento, Italy
- Nicola Ferro, University of Padua, Italy
- Christian Fluhr, Cadege, 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, The Netherlands
- Noriko Kando, National Institute of Informatics, Tokyo, Japan
- Jussi Karlgren, Swedish Institute of Computer Science, Sweden
- Michael Kluck, German Institute for International and Security Affairs, Berlin, Germany
- Natalia Loukachevitch, Moscow State University, Russia
- Bernardo Magnini, Fondazione Bruno Kessler, Trento, Italy
- Paul McNamee, Johns Hopkins University, USA
- Henning Müller, University of Applies Sciences Western Switzerland, Sierre and University of Geneva, Switzerland
- Douglas W. Oard, University of Maryland, USA
- Anselmo Peñas, LSI-UNED, Madrid, Spain
- Vivien Petras, Humboldt University Berlin, Germany
- Maarten de Rijke, University of Amsterdam, The Netherlands
- Diana Santos, Linguateca, Sintef, Oslo, Norway
- 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, 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 – Part I

What Happened in CLEF 2009 1
Carol Peters

I: Multilingual Textual Document Retrieval (AdHoc)

CLEF 2009 Ad Hoc Track Overview: TEL and Persian Tasks 13
Nicola Ferro and Carol Peters

CLEF 2009 Ad Hoc Track Overview: Robust-WSD Task 36
Eneko Agirre, Giorgio Maria Di Nunzio, Thomas Mandl, and Arantxa Otegi

AdHoc-TEL

Evaluating Cross-Language Explicit Semantic Analysis and Cross Querying 50
Maik Anderka, Nedim Lipka, and Benno Stein

Document Expansion, Query Translation and Language Modeling for Ad-Hoc IR 58
Johannes Leveling, Dong Zhou, Gareth J.F. Jones, and Vincent Wade

Smoothing Methods and Cross-Language Document Re-ranking 62
Dong Zhou and Vincent Wade

Cross-Language Information Retrieval Using Meta-language Index Construction and Structural Queries 70
Amir Hossein Jadidinejad and Fariborz Mahmoudi

Sampling Precision to Depth 10000 at CLEF 2009 78
Stephen Tomlinson

Multilingual Query Expansion for CLEF Adhoc-TEL 86
Ray R. Larson

Experiments with N-Gram Prefixes on a Multinomial Language Model versus Lucene's Off-the-Shelf Ranking Scheme and Rocchio Query Expansion (TEL@CLEF Monolingual Task) 90
Jorge Machado, Bruno Martins, and José Borbinha

AdHoc-Persian

Evaluation of Perstem: A Simple and Efficient Stemming Algorithm for Persian 98
Amir Hossein Jadidinejad, Fariborz Mahmoudi, and Jon Dehdari

Ad Hoc Retrieval with the Persian Language 102
Ljiljana Dolamic and Jacques Savoy

Ad Hoc Information Retrieval for Persian 110
AmirHossein Habibian, Abolfazl AleAhmad, and Azadeh Shakery

AdHoc-Robust

Combining Probabilistic and Translation-Based Models for Information Retrieval Based on Word Sense Annotations 120
Elisabeth Wolf, Delphine Bernhard, and Iryna Gurevych

Indexing with WordNet Synonyms May Improve Retrieval Results 128
Davide Buscaldi and Paolo Rosso

UFRGS@CLEF2009: Retrieval by Numbers 135
Thyago Bohrer Borges and Viviane P. Moreira

Evaluation of Axiomatic Approaches to Crosslanguage Retrieval 142
Roman Kern, Andreas Juffinger, and Michael Granitzer

UNIBA-SENSE @ CLEF 2009: Robust WSD Task 150
Pierpaolo Basile, Annalina Caputo, and Giovanni Semeraro

Using WordNet Relations and Semantic Classes in Information Retrieval Tasks 158
Javi Fernández, Rubén Izquierdo, and José M. Gómez

Using Semantic Relatedness and Word Sense Disambiguation for (CL)IR 166
Eneko Agirre, Arantxa Otegi, and Hugo Zaragoza

II: Multiple Language Question Answering (QA@CLEF)

Overview of ResPubliQA 2009: Question Answering Evaluation over European Legislation 174
Anselmo Peñas, Pamela Forner, Richard Sutcliffe, Álvaro Rodrigo, Corina Forčscu, Iñaki Alegria, Danilo Giampiccolo, Nicolas Moreau, and Petya Osenova

Overview of QAST 2009 197
Jordi Turmo, Pere R. Comas, Sophie Rosset, Olivier Galibert, Nicolas Moreau, Djamel Mostefa, Paolo Rosso, and Davide Buscaldi

GikiCLEF: Expectations and Lessons Learned 212
Diana Santos and Luís Miguel Cabral

ResPubliQA

NLEL-MAAT at ResPubliQA 223
Santiago Correa, Davide Buscaldi, and Paolo Rosso

Question Answering on English and Romanian Languages 229
Adrian Iftene, Diana Trandabăţ, Alex Moruz, Ionuţ Pistol, Maria Husarciuc, and Dan Cristea

Studying Syntactic Analysis in a QA System: FIDJI @ ResPubliQA'09 237
Xavier Tannier and Véronique Moriceau

Approaching Question Answering by Means of Paragraph Validation . . . 245
Álvaro Rodrigo, Joaquín Pérez-Iglesias, Anselmo Peñas, Guillermo Garrido, and Lourdes Araujo

Information Retrieval Baselines for the ResPubliQA Task 253
Joaquín Pérez-Iglesias, Guillermo Garrido, Álvaro Rodrigo, Lourdes Araujo, and Anselmo Peñas

A Trainable Multi-factored QA System 257
Radu Ion, Dan Ştefănescu, Alexandru Ceauşu, Dan Tufiş, Elena Irimia, and Verginica Barbu Mititelu

Extending a Logic-Based Question Answering System for Administrative Texts 265
Ingo Glöckner and Björn Pelzer

Elhuyar-IXA: Semantic Relatedness and Cross-Lingual Passage Retrieval 273
Eneko Agirre, Olatz Ansa, Xabier Arregi, Maddalen Lopez de Lacalle, Arantxa Otegi, Xabier Saralegi, and Hugo Zaragoza

Are Passages Enough? The MIRACLE Team Participation in QA@CLEF2009 281
María Teresa Vicente-Díez, César de Pablo-Sánchez, Paloma Martínez, Julián Moreno Schneider, and Marta Garrote Salazar

QAST

The LIMSI Participation in the QAst 2009 Track: Experimenting on Answer Scoring 289
Guillaume Bernard, Sophie Rosset, Olivier Galibert, Gilles Adda, and Eric Bilinski

Robust Question Answering for Speech Transcripts: UPC Experience in QAst 2009 297
Pere R. Comas and Jordi Turmo

GikiCLEF

Where in the Wikipedia Is That Answer? The XLDB at the GikiCLEF 2009 Task 305
Nuno Cardoso, David Batista, Francisco J. Lopez-Pellicer, and Mário J. Silva

Recursive Question Decomposition for Answering Complex Geographic Questions 310
Sven Hartrumpf and Johannes Leveling

GikiCLEF Topics and Wikipedia Articles: Did They Blend? 318
Nuno Cardoso

TALP at GikiCLEF 2009 322
Daniel Ferrés and Horacio Rodríguez

Semantic QA for Encyclopaedic Questions: *EQUAL* in GikiCLEF 326
Iustin Dornescu

Interactive Probabilistic Search for GikiCLEF 334
Ray R. Larson

III: Multilingual Information Filtering (INFILE)

Information Filtering Evaluation: Overview of CLEF 2009 INFILE Track 342
Romaric Besançon, Stéphane Chaudiron, Djamel Mostefa, Ismaïl Timimi, Khalid Choukri, and Meriama Laïb

Batch Document Filtering Using Nearest Neighbor Algorithm 354
Ali Mustafa Qamar, Eric Gaussier, and Nathalie Denos

UAIC: Participation in INFILE@CLEF Task 362
Cristian-Alexandru Drăguşanu, Alecsandru Grigoriu, and Adrian Iftene

Multilingual Information Filtering by Human Plausible Reasoning 366
Asma Damankesh, Farhad Oroumchian, and Khaled Shaalan

Hossur'Tech's Participation in CLEF 2009 INFILE Interactive Filtering 374
John Anton Chrisostom Ronald, Aurélie Rossi, and Christian Fluhr

Experiments with Google News for Filtering Newswire Articles 381
Arturo Montejo-Ráez, José M. Perea-Ortega, Manuel Carlos Díaz-Galiano, and L. Alfonso Ureña-López

IV: Intellectual Property (CLEF-IP)

CLEF-IP 2009: Retrieval Experiments in the Intellectual Property Domain 385
Giovanna Roda, John Tait, Florina Piroi, and Veronika Zenz

Exploring Structured Documents and Query Formulation Techniques for Patent Retrieval 410
Walid Magdy, Johannes Leveling, and Gareth J.F. Jones

Formulating Good Queries for Prior Art Search 418
José Carlos Toucedo and David E. Losada

UAIC: Participation in CLEF-IP Track 426
Adrian Iftene, Ovidiu Ionescu, and George-Răzvan Oancea

PATATRAS: Retrieval Model Combination and Regression Models for Prior Art Search 430
Patrice Lopez and Laurent Romary

NLEL-MAAT at CLEF-IP 438
Santiago Correa, Davide Buscaldi, and Paolo Rosso

Simple Pre and Post Processing Strategies for Patent Searching in CLEF Intellectual Property Track 2009 444
Julien Gobeill, Emilie Pasche, Douglas Teodoro, and Patrick Ruch

Prior Art Search Using International Patent Classification Codes and All-Claims-Queries 452
Benjamin Herbert, György Szarvas, and Iryna Gurevych

UTA and SICS at CLEF-IP'09 460
Antti Järvelin, Anni Järvelin, and Preben Hansen

Searching CLEF-IP by Strategy 468
W. Alink, Roberto Cornacchia, and Arjen P. de Vries

UniNE at CLEF-IP 2009 476
Claire Fautsch and Jacques Savoy

Automatically Generating Queries for Prior Art Search 480
Erik Graf, Leif Azzopardi, and Keith van Rijsbergen

Patent Retrieval Experiments in the Context of the CLEF IP Track 2009 491
Daniela Becks, Christa Womser-Hacker, Thomas Mandl, and Ralph Kölle

Prior Art Retrieval Using the Claims Section as a Bag of Words 497
Suzan Verberne and Eva D'hondt

UniGE Experiments on Prior Art Search in the Field of Patents 502
Jacques Guyot, Gilles Falquet, and Karim Benzineb

V: Logfile Analysis (LogCLEF)

LogCLEF 2009: The CLEF 2009 Multilingual Logfile Analysis Track Overview 508
Thomas Mandl, Maristella Agosti, Giorgio Maria Di Nunzio, Alexander Yeh, Inderjeet Mani, Christine Doran, and Julia Maria Schulz

Identifying Common User Behaviour in Multilingual Search Logs 518
M. Rami Ghorab, Johannes Leveling, Dong Zhou, Gareth J.F. Jones, and Vincent Wade

A Search Engine Based on Query Logs, and Search Log Analysis by Automatic Language Identification 526
Michael Oakes and Yan Xu

Identifying Geographical Entities in Users' Queries 534
Adrian Iftene

Search Path Visualization and Session Performance Evaluation with Log Files 538
Katrin Lamm, Thomas Mandl, and Ralph Koelle

User Logs as a Means to Enrich and Refine Translation Dictionaries 544
Alessio Bosca and Luca Dini

VI: Grid Experiments (GRID@CLEF)

CLEF 2009: Grid@CLEF Pilot Track Overview 552
Nicola Ferro and Donna Harman

Decomposing Text Processing for Retrieval: Cheshire Tries GRID@CLEF 566
Ray R. Larson

Putting It All Together: The Xtrieval Framework at Grid@CLEF 2009 570
Jens Kürsten and Maximilian Eibl

VII: Morphochallenge

Overview and Results of Morpho Challenge 2009 578
Mikko Kurimo, Sami Virpioja, Ville T. Turunen, Graeme W. Blackwood, and William Byrne

MorphoNet: Exploring the Use of Community Structure for Unsupervised Morpheme Analysis 598
Delphine Bernhard

Unsupervised Morpheme Analysis with Allomorfessor 609
Sami Virpioja, Oskar Kohonen, and Krista Lagus

Unsupervised Morphological Analysis by Formal Analogy 617
Jean-François Lavallée and Philippe Langlais

Unsupervised Word Decomposition with the Promodes Algorithm 625
Sebastian Spiegler, Bruno Golénia, and Peter Flach

Unsupervised Morpheme Discovery with Ungrade 633
Bruno Golénia, Sebastian Spiegler, and Peter Flach

Clustering Morphological Paradigms Using Syntactic Categories 641
Burcu Can and Suresh Manandhar

Simulating Morphological Analyzers with Stochastic Taggers for Confidence Estimation 649
Christian Monson, Kristy Hollingshead, and Brian Roark

A Rule-Based Acquisition Model Adapted for Morphological Analysis 658
Constantine Lignos, Erwin Chan, Mitchell P. Marcus, and Charles Yang

Morphological Analysis by Multiple Sequence Alignment 666
Tzvetan Tchoukalov, Christian Monson, and Brian Roark

Author Index 675

Table of Contents – Part II

What Happened in CLEF 2009 1
Carol Peters

I: Interactive Cross-Language Retrieval (iCLEF)

Overview of iCLEF 2009: Exploring Search Behaviour in a Multilingual Folksonomy Environment 13
Julio Gonzalo, Víctor Peinado, Paul Clough, and Jussi Karlgren

Analysis of Multilingual Image Search Logs: Users' Behavior and Search Strategies 21
Víctor Peinado, Fernando López-Ostenero, and Julio Gonzalo

User Behaviour and Lexical Ambiguity in Cross-Language Image Retrieval 29
Borja Navarro-Colorado, Marcel Puchol-Blasco, Rafael M. Terol, Sonia Vázquez, and Elena Lloret

Users' Image Seeking Behavior in a Multilingual Tag Environment 37
Miguel E. Ruiz and Pok Chin

II: Cross-Language Retrieval in Image Collections (ImageCLEF)

Diversity in Photo Retrieval: Overview of the ImageCLEFPhoto Task 2009 45
Monica Lestari Paramita, Mark Sanderson, and Paul Clough

Overview of the WikipediaMM Task at ImageCLEF 2009 60
Theodora Tsikrika and Jana Kludas

Overview of the CLEF 2009 Medical Image Retrieval Track 72
Henning Müller, Jayashree Kalpathy–Cramer, Ivan Eggel, Steven Bedrick, Saïd Radhouani, Brian Bakke, Charles E. Kahn Jr., and William Hersh

Overview of the CLEF 2009 Medical Image Annotation Track 85
Tatiana Tommasi, Barbara Caputo, Petra Welter, Mark Oliver Güld, and Thomas M. Deserno

Overview of the CLEF 2009 Large-Scale Visual Concept Detection and Annotation Task 94
Stefanie Nowak and Peter Dunker

Overview of the CLEF 2009 Robot Vision Track 110
Andrzej Pronobis, Li Xing, and Barbara Caputo

ImageCLEFPhoto

Diversity Promotion: Is Reordering Top-Ranked Documents Sufficient? 120
Sergio Navarro, Rafael Muñoz, and Fernando Llopis

Comparison of Several Combinations of Multimodal and Diversity Seeking Methods for Multimedia Retrieval 124
Julien Ah-Pine, Stephane Clinchant, and Gabriela Csurka

University of Glasgow at ImageCLEFPhoto 2009: Optimising Similarity and Diversity in Image Retrieval 133
Teerapong Leelanupab, Guido Zuccon, Anuj Goyal, Martin Halvey, P. Punitha, and Joemon M. Jose

Multimedia Retrieval by Means of Merge of Results from Textual and Content Based Retrieval Subsystems 142
Ana García-Serrano, Xaro Benavent, Ruben Granados, Esther de Ves, and José Miguel Goñi

Image Query Expansion Using Semantic Selectional Restrictions 150
Osama El Demerdash, Sabine Bergler, and Leila Kosseim

Clustering for Text and Image-Based Photo Retrieval at CLEF 2009 . . . 157
Qian Zhu and Diana Inkpen

ImageCLEFwiki

Combining Text/Image in WikipediaMM Task 2009 164
Christophe Moulin, Cécile Barat, Cédric Lemaître, Mathias Géry, Christophe Ducottet, and Christine Largeron

Document Expansion for Text-Based Image Retrieval at CLEF 2009 172
Jinming Min, Peter Wilkins, Johannes Leveling, and Gareth J.F. Jones

Multimodal Image Retrieval over a Large Database 177
Débora Myoupo, Adrian Popescu, Hervé Le Borgne, and Pierre-Alain Moëllic

Using WordNet in Multimedia Information Retrieval 185
Manuel Carlos Díaz-Galiano, María Teresa Martín-Valdivia, L. Alfonso Ureña-López, and José Manuel Perea-Ortega

ImageCLEFmed

Medical Image Retrieval: ISSR at CLEF 2009 189
Waleed Arafa and Ragia Ibrahim

An Integrated Approach for Medical Image Retrieval through Combining Textual and Visual Features 195
Zheng Ye, Xiangji Huang, Qinmin Hu, and Hongfei Lin

Analysis Combination and Pseudo Relevance Feedback in Conceptual Language Model: LIRIS Participation at ImageCLEFMed 203
Loïc Maisonnasse, Farah Harrathi, Catherine Roussey, and Sylvie Calabretto

The MedGIFT Group at ImageCLEF 2009 211
Xin Zhou, Ivan Eggel, and Henning Müller

An Extended Vector Space Model for Content-Based Image Retrieval . . . 219
Tolga Berber and Adil Alpkocak

Using Media Fusion and Domain Dimensions to Improve Precision in Medical Image Retrieval 223
Saïd Radhouani, Jayashree Kalpathy-Cramer, Steven Bedrick, Brian Bakke, and William Hersh

ImageCLEFmed Annotation

ImageCLEF 2009 Medical Image Annotation Task: PCTs for Hierarchical Multi-Label Classification 231
Ivica Dimitrovski, Dragi Kocev, Suzana Loskovska, and Sašo Džeroski

Dense Simple Features for Fast and Accurate Medical X-Ray Annotation 239
Uri Avni, Hayit Greenspan, and Jacob Goldberger

Automated X-Ray Image Annotation: Single versus Ensemble of Support Vector Machines 247
Devrim Unay, Octavian Soldea, Sureyya Ozogur-Akyuz, Mujdat Cetin, and Aytul Ercil

ImageCLEF Annotation and Robot Vision

Topological Localization of Mobile Robots Using Probabilistic Support Vector Classification 255
Yan Gao and Yiqun Li

The University of Amsterdam's Concept Detection System at ImageCLEF 2009 261
Koen E.A. van de Sande, Theo Gevers, and Arnold W.M. Smeulders

Enhancing Recognition of Visual Concepts with Primitive Color Histograms via Non-sparse Multiple Kernel Learning 269
Alexander Binder and Motoaki Kawanabe

Using SIFT Method for Global Topological Localization for Indoor Environments 277
Emanuela Boroş, George Roşca, and Adrian Iftene

UAIC at ImageCLEF 2009 Photo Annotation Task 283
Adrian Iftene, Loredana Vamanu, and Cosmina Croitoru

Learning Global and Regional Features for Photo Annotation 287
Jiquan Ngiam and Hanlin Goh

Improving Image Annotation in Imbalanced Classification Problems with Ranking SVM 291
Ali Fakeri-Tabrizi, Sabrina Tollari, Nicolas Usunier, and Patrick Gallinari

University of Glasgow at ImageCLEF 2009 Robot Vision Task: A Rule Based Approach 295
Yue Feng, Martin Halvey, and Joemon M. Jose

A Fast Visual Word Frequency - Inverse Image Frequency for Detector of Rare Concepts 299
Emilie Dumont, Hervé Glotin, Sébastien Paris, and Zhong-Qiu Zhao

Exploring the Semantics behind a Collection to Improve Automated Image Annotation 307
Ainhoa Llorente, Enrico Motta, and Stefan Rüger

Multi-cue Discriminative Place Recognition 315
Li Xing and Andrzej Pronobis

MRIM-LIG at ImageCLEF 2009: Robotvision, Image Annotation and Retrieval Tasks 324
Trong-Ton Pham, Loïc Maisonnasse, Philippe Mulhem, Jean-Pierre Chevallet, Georges Quénot, and Rami Al Batal

ImageCLEF Mixed

The ImageCLEF Management System 332
Ivan Eggel and Henning Müller

Interest Point and Segmentation-Based Photo Annotation 340
Bálint Daróczy, István Petrás, András A. Benczúr, Zsolt Fekete, Dávid Nemeskey, Dávid Siklósi, and Zsuzsa Weiner

University of Jaén at ImageCLEF 2009: Medical and Photo Tasks 348
Miguel A. García-Cumbreras, Manuel Carlos Díaz-Galiano, María Teresa Martín-Valdivia, Arturo Montejo-Raez, and L. Alfonso Ureña-López

III: Cross-Language Retrieval in Video Collections (VideoCLEF)

Overview of VideoCLEF 2009: New Perspectives on Speech-Based Multimedia Content Enrichment 354
Martha Larson, Eamonn Newman, and Gareth J.F. Jones

Methods for Classifying Videos by Subject and Detecting Narrative Peak Points 369
Tudor-Alexandru Dobrilă, Mihail-Ciprian Diaconaşu, Irina-Diana Lungu, and Adrian Iftene

Using Support Vector Machines as Learning Algorithm for Video Categorization 373
José Manuel Perea-Ortega, Arturo Montejo-Ráez, María Teresa Martín-Valdivia, and L. Alfonso Ureña-López

Video Classification as IR Task: Experiments and Observations 377
Jens Kürsten and Maximilian Eibl

Exploiting Speech Recognition Transcripts for Narrative Peak Detection in Short-Form Documentaries 385
Martha Larson, Bart Jochems, Ewine Smits, and Roeland Ordelman

Identification of Narrative Peaks in Video Clips: Text Features Perform Best 393
Joep J.M. Kierkels, Mohammad Soleymani, and Thierry Pun

A Cocktail Approach to the VideoCLEF'09 Linking Task 401
Stephan Raaijmakers, Corné Versloot, and Joost de Wit

When to Cross Over? Cross-Language Linking Using Wikipedia for VideoCLEF 2009 409
Ágnes Gyarmati and Gareth J.F. Jones

Author Index 413