Preface

The fifth campaign of the Cross-Language Evaluation Forum (CLEF) for European languages was held from January to September 2004. Participation in the CLEF campaigns has increased each year and CLEF 2004 was no exception: 55 groups submitted results for one or more of the different tracks compared with 42 groups in the previous year. CLEF 2004 also marked a breaking point with respect to previous campaigns. The focus was no longer mainly concentrated on multilingual document retrieval as in previous years but was diversified to include different kinds of text retrieval across languages (e.g., exact answers in the question-answering track) and retrieval on different kinds of media (i.e., not just plain text but collections containing image and speech as well). In addition, increasing attention was given to issues that regard system usability and user satisfaction with tasks to measure the effectiveness of interactive systems or system components being included in both the cross-language question answering and image retrieval tasks with the collaboration of the coordinators of the interactive track.

The campaign culminated in a two-and-a-half-day workshop held in Bath, UK, 15–17 September, immediately following the 8th European Conference on Digital Libraries. The workshop was attended by nearly 100 researchers and system developers. In addition to presentations by participants in the campaign, talks included reports on the activities of the NTCIR evaluation initiative for Asian languages, and on industrial experience in building cross-language applications. The final session consisted of a panel in which the members attempted to analyze the current organization of the CLEF campaigns in depth, discussing whether CLEF is working on the right problems, choosing its investments wisely, and giving sufficient attention to the user perspective. Suggestions for the CLEF 2005 campaign included multilingual Web retrieval and a cross-language Geographic Information Retrieval track.

CLEF 2004 was conducted as an activity of the DELOS Network of Excellence on Digital Libraries, within the framework of the Information Society Technologies programme of the European Commission. These post-campaign proceedings were prepared with the assistance of the Center for the Evaluation of Language and Communication Technologies (CELCCT), Trento, Italy. The support of DELOS and CELCT is gratefully acknowledged. We should also like to thank the other members of the CLEF Steering Committee for their assistance in the coordination of this event.

April 2005

Carol Peters
Paul Clough
Julio Gonzalo
Gareth J.F. Jones
Michael Kluck
Bernardo Magnini
CLEF 2004 Coordination

CLEF is coordinated by the Istituto di Scienza e Tecnologie dell’Informazione, Consiglio Nazionale delle Ricerche, ISTI-CNR, Pisa.

The following institutions contributed to the organization of the different tracks of the CLEF 2004 campaign:

- Center for the Evaluation of Language and Communication Technologies (CELC), Trento, Italy
- Centro per la Ricerca Scientifica e Tecnologica, Istituto Trentino di Cultura, Trento, Italy
- College of Information Studies and Institute for Advanced Computer Studies, University of Maryland, USA
- Department of Computer Science and Information Systems, University of Limerick, Ireland
- Department of Information Studies, University of Sheffield, UK
- Department of Information Studies, University of Tampere, Finland
- Eurospider Information Technology AG, Zürich, Switzerland
- Evaluations and Language Resources Distribution Agency, ELDA, Paris, France
- German Research Centre for Artificial Intelligence, DFKI, Saarbrücken, Germany
- Information and Language Processing Systems, University of Amsterdam, Netherlands
- Informationszentrum Sozialwissenschaften, Bonn, Germany
- Lenguajes y Sistemas Informáticos, Universidad Nacional de Educación a Distancia, Madrid, Spain
- Linguateca, Sintef, Oslo, Norway
- Linguistic Modelling Laboratory, Bulgarian Academy of Sciences, Sofia, Bulgaria
- National Institute of Standards and Technology, Gaithersburg MD, USA
- School of Computing, Dublin City University, Ireland
- University and University Hospitals of Geneva, Switzerland
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?
*Carol Peters* .................................................. 1

## Part I. Ad Hoc Text Retrieval Tracks

CLEF 2004: Ad Hoc Track Overview and Results Analysis
*Martin Braschler, Giorgio M. Di Nunzio, Nicola Ferro, Carol Peters* .................................................. 10

Mainly Cross-Language Experiments

Selection and Merging Strategies for Multilingual Information Retrieval
*Jacques Savoy, Pierre-Yves Berger* ................................. 27

Using Surface-Syntactic Parser and Deviation from Randomness
*Jean-Pierre Chevallet, Gilles Sérasset* ............................. 38

Cross-Language Retrieval Using HAIRCUT at CLEF 2004
*Paul McNamee, James Mayfield* .................................. 50

Experiments on Statistical Approaches to Compensate for Limited Linguistic Resources
*Giorgio M. Di Nunzio, Nicola Ferro, Nicola Orio* ............... 60

Application of Variable Length *N*-Gram Vectors to Monolingual and Bilingual Information Retrieval
*Daniel Gayo-Avello, Darío Álvarez-Gutiérrez, José Gayo-Avello* .... 73

Integrating New Languages in a Multilingual Search System Based on a Deep Linguistic Analysis
*Romaric Besançon, Olivier Ferret, Christian Fluhr* ............. 83

IR-n r2: Using Normalized Passages
*Fernando Llopis, Rafael Muñoz, Rafael M. Terol, Elisa Noguera* ..... 90

Using COTS Search Engines and Custom Query Strategies at CLEF
*David Nadeau, Mario Jarmasz, Caroline Barrière, George Foster, Claude St-Jacques* ................................. 100
# Table of Contents

Report on Thomson Legal and Regulatory Experiments at CLEF-2004  
*Isabelle Moulinier, Ken Williams* ........................................ 110

Effective Translation, Tokenization and Combination for Cross-Lingual  
Retrieval  
*Jaap Kamps, Sisay Fissaha Adafre, Maarten de Rijke* .................... 123

Two-Stage Refinement of Transitive Query Translation with English  
Disambiguation for Cross-Language Information Retrieval: An  
Experiment at CLEF 2004  
*Kazuaki Kishida, Noriko Kando, Kuang-Hua Chen* ....................... 135

Dictionary-Based Amharic – English Information Retrieval  
*Atelach Alemu Argaw, Lars Asker, Rickard Cöster, Jussi Karlgren* .................................................. 143

Dynamic Lexica for Query Translation  
*Jussi Karlgren, Magnus Sahlgren, Timo Järvinen, Rickard Cöster* .... 150

SINAI at CLEF 2004: Using Machine Translation Resources with a  
Mixed 2-Step RSV Merging Algorithm  
*Fernando Martínez-Santiago, Miguel A. García-Cumbreras, Manuel C. Díaz-Galiano, L. Alfonso Ureña* ....................... 156

Mono- and Crosslingual Retrieval Experiments at the University of  
Hildesheim  
*René Hackl, Thomas Mandl, Christa Womser-Hacker* ..................... 165

University of Chicago at CLEF2004: Cross-Language Text and Spoken  
Document Retrieval  
*Gina-Anne Levow, Irina Matveeva* ......................................... 170

UB at CLEF2004: Cross Language Information Retrieval Using  
Statistical Language Models  
*Miguel E. Ruiz, Munirathnam Srikanth* .................................. 180

MIRACLE’s Hybrid Approach to Bilingual and Monolingual  
Information Retrieval  
*José M. Goñi-Menoyo, José C. González, José Luis Martínez-Fernández, Julio Villena* ..................... 188

Searching a Russian Document Collection Using English, Chinese and  
Japanese Queries  
*Fredric C. Gey* ..................................................................... 200
Dublin City University at CLEF 2004: Experiments in Monolingual, Bilingual and Multilingual Retrieval
  Gareth J.F. Jones, Michael Burke, John Judge, Anna Khasin, Adenike Lam-Adesina, Joachim Wagner ......................... 207

Monolingual Experiments

Finnish, Portuguese and Russian Retrieval with Hummingbird SearchServer™ at CLEF 2004
  Stephen Tomlinson .................................................. 221

Data Fusion for Effective European Monolingual Information Retrieval
  Jacques Savoy ............................................................ 233

The XLDB Group at CLEF 2004
  Nuno Cardoso, Mário J. Silva, Miguel Costa .................... 245

The University of Glasgow at CLEF 2004: French Monolingual Information Retrieval with Terrier
  Christina Lioma, Ben He, Vassilis Plachouras, Iadh Ounis ........ 253

Part II. Domain-Specific Document Retrieval

The Domain-Specific Track in CLEF 2004: Overview of the Results and Remarks on the Assessment Process
  Michael Kluck ............................................................ 260

University of Hagen at CLEF 2004: Indexing and Translating Concepts for the GIRT Task
  Johannes Leveling, Sven Hartrumpf ............................... 271

IRIT at CLEF 2004: The English GIRT Task
  Mustapha Baziz, Mohand Boughanem, Nathalie Aussenac-Gilles .... 283

Ricoh at CLEF 2004
  Yuichi Kojima ........................................................... 292

GIRT and the Use of Subject Metadata for Retrieval
  Vivien Petras ............................................................ 298
Part III. Interactive Cross-Language Information Retrieval

iCLEF 2004 Track Overview: Pilot Experiments in Interactive Cross-Language Question Answering
Julio Gonzalo, Douglas W. Oard ................................................................. 310

Interactive Cross-Language Question Answering: Searching Passages Versus Searching Documents
Fernando López-Ostenero, Julio Gonzalo, Víctor Peinado,
Felisa Verdejo ................................................................. 323

Improving Interaction with the User in Cross-Language Question Answering Through Relevant Domains and Syntactic Semantic Patterns
Borja Navarro, Lorenza Moreno, Sonia Vázquez, Fernando Llopis,
Andrés Montoyo, Miguel Ángel Varó .................................................. 334

Cooperation, Bookmarking, and Thesaurus in Interactive Bilingual Question Answering
Preben Hansen, Jussi Karlgren, Magnus Sahlgren ............................. 343

Summarization Design for Interactive Cross-Language Question Answering
Daqing He, Jianqiang Wang, Jun Luo, Douglas W. Oard .................. 348

Interactive and Bilingual Question Answering Using Term Suggestion and Passage Retrieval
Carlos G. Figuerola, Angel F. Zazo, José L. Alonso Berrocal,
Emilio Rodríguez Vázquez de Aldana .................................................. 363

Part IV. Multiple Language Question Answering

Overview of the CLEF 2004 Multilingual Question Answering Track
Bernardo Magnini, Alessandro Vallin, Christelle Ayache,
Gregor Erbach, Anselmo Peñas, Maarten de Rijke, Paulo Rocha,
Kiril Simov, Richard F.E. Sutcliffe .................................................. 371

A Question Answering System for French
Laura Perret ................................................................. 392

Cross-Language French-English Question Answering Using the DLT System at CLEF 2004
Richard F.E. Sutcliffe, Igal Gabbay, Michael Mulcahy,
Aoife O’Gorman ................................................................. 404
Experiments on Robust NL Question Interpretation and Multi-layered Document Annotation for a Cross-Language Question/Answering System
  Günter Neumann, Bogdan Sacaleanu ........................................... 411

Making Stone Soup: Evaluating a Recall-Oriented Multi-stream Question Answering System for Dutch
  David Ahn, Valentin Jijkoun, Karin Müller, Maarten de Rijke,
  Stefan Schlobach, Gilad Mishne ................................................. 423

The DIOGENE Question Answering System at CLEF-2004
  Hristo Tanev, Matteo Negri, Bernardo Magnini,
  Milen Kouylekov ........................................................................ 435

Cross-Lingual Question Answering Using Off-the-Shelf Machine Translation
  Kisuh Ahn, Beatrice Alex, Johan Bos, Tiphaine Dalmas,
  Jochen L. Leidner, Matthew B. Smillie .......................................... 446

Bulgarian-English Question Answering: Adaptation of Language Resources
  Petya Osenova, Alexander Simov, Kiril Simov, Hristo Tanev,
  Milen Kouylekov ........................................................................ 458

Answering French Questions in English by Exploiting Results from Several Sources of Information
  Brigitte Grau, Gabriel Illouz, Laura Monceaux, Isabelle Robba,
  Anne Vilnat, Guillaume Bourdil, Faïza Elkateb-Gara, Olivier Ferret,
  Benoît Mathieu ............................................................................ 470

Finnish as Source Language in Bilingual Question Answering
  Lili Aunimo, Reeta Kuuskoski, Juha Makkonen .............................. 482

miraQA: Experiments with Learning Answer Context Patterns from the Web
  César de Pablo-Sánchez, José Luis Martínez-Fernández,
  Paloma Martínez, Julio Villena .................................................... 494

Question Answering for Spanish Supported by Lexical Context Annotation
  Manuel Pérez-Coutiño, Thamar Solorio, Manuel Montes-y-Gómez,
  Aurelio López-López, Luis Villaseñor-Pineda .................................. 502

Question Answering Using Sentence Parsing and Semantic Network Matching
  Sven Hartrumpf ............................................................................ 512
First Evaluation of Esfinge – A Question Answering System for Portuguese

Luís Costa ............................................................... 522

University of Évora in QA@CLEF-2004

Paulo Quaresma, Luís Quintano, Irene Rodrigues, José Saias, 
Pedro Salgueiro .................................................... 534

COLE Experiments at QA@CLEF 2004 Spanish Monolingual Track

Enrique Méndez Díaz, Jesús Vilares Ferro, David Cabrero Souto .... 544

Does English Help Question Answering in Spanish?

José Luis Vicedo, Maximiliano Saiz, Rubén Izquierdo, 
Fernando Llopis ..................................................... 552

The TALP-QA System for Spanish at CLEF 2004: Structural and 
Hierarchical Relaxing of Semantic Constraints

Daniel Ferrés, Samir Kanaan, Alicia Ageno, Edgar González, 
Horacio Rodríguez, Mihai Surdeanu, Jordi Turmo ............... 557

ILC-UniPI Italian QA

Francesca Bertagna, Luminita Chiran, Maria Simi .................. 569

Question Answering Pilot Task at CLEF 2004

Jesús Herrera, Anselmo Peñas, Felisa Verdejo ...................... 581

Evaluation of Complex Temporal Questions in CLEF-QA

Estela Saquete, José Luis Vicedo, Patricio Martínez-Barco, 
Rafael Muñoz, Fernando Llopis .................................. 591

Part V. Cross-Language Retrieval in Image Collections

The CLEF 2004 Cross-Language Image Retrieval Track

Paul Clough, Henning Müller, Mark Sanderson ..................... 597

Caption and Query Translation for Cross-Language Image Retrieval

Paul Clough ............................................................ 614

Pattern-Based Image Retrieval with Constraints and Preferences on 
ImageCLEF 2004

Maximiliano Saiz, José Luis Vicedo, Rubén Izquierdo ............. 626
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>How to Visually Retrieve Images from the St. Andrews Collection Using GIFT</td>
<td>633</td>
</tr>
<tr>
<td>Henning Müller, Antoine Geissbühler</td>
<td></td>
</tr>
<tr>
<td>UNED at ImageCLEF 2004: Detecting Named Entities and Noun Phrases for Automatic Query Expansion and Structuring</td>
<td>643</td>
</tr>
<tr>
<td>Víctor Peinado, Javier Artiles, Fernando López-Ostenero, Julio Gonzalo, Felisa Verdejo</td>
<td></td>
</tr>
<tr>
<td>Dublin City University at CLEF 2004: Experiments with the ImageCLEF St. Andrew’s Collection</td>
<td>653</td>
</tr>
<tr>
<td>Gareth J.F. Jones, Declan Groves, Anna Khasin, Adenike Lam-Adesina, Bart Mellebeek, Andy Way</td>
<td></td>
</tr>
<tr>
<td>From Text to Image: Generating Visual Query for Image Retrieval</td>
<td>664</td>
</tr>
<tr>
<td>Wen-Cheng Lin, Yih-Chen Chang, Hsin-Hsi Chen</td>
<td></td>
</tr>
<tr>
<td>Toward Cross-Language and Cross-Media Image Retrieval</td>
<td>676</td>
</tr>
<tr>
<td>Carmen Alvarez, Ahmed Id Oumohmed, Max Mignotte, Jian-Yun Nie</td>
<td></td>
</tr>
<tr>
<td>Thomas Deselaers, Daniel Keysers, Hermann Ney</td>
<td></td>
</tr>
<tr>
<td>MIRACLE Approach to ImageCLEF 2004: Merging Textual and Content-Based Image Retrieval</td>
<td>699</td>
</tr>
<tr>
<td>José Luis Martínez-Fernández, Ana García-Serrano, Julio Villena, Víctor Méndez-Sáenz</td>
<td></td>
</tr>
<tr>
<td>Cross-Media Feedback Strategies: Merging Text and Image Information to Improve Image Retrieval</td>
<td>709</td>
</tr>
<tr>
<td>Romaric Besançon, Patrick Hède, Pierre-Alain Moellic, Christian Fluhr</td>
<td></td>
</tr>
<tr>
<td>ImageCLEF 2004: Combining Image and Multi-lingual Search for Medical Image Retrieval</td>
<td>718</td>
</tr>
<tr>
<td>Henning Müller, Antoine Geissbühler, Patrick Ruch</td>
<td></td>
</tr>
<tr>
<td>Multi-modal Information Retrieval Using FINT</td>
<td>728</td>
</tr>
<tr>
<td>Menno van Zaanen, Guido de Croon</td>
<td></td>
</tr>
<tr>
<td>Medical Image Retrieval Using Texture, Locality and Colour</td>
<td>740</td>
</tr>
<tr>
<td>Peter Howarth, Alexei Yavlinsky, Daniel Heesch, Stefan Rüger</td>
<td></td>
</tr>
</tbody>
</table>
SMIRE: Similar Medical Image Retrieval Engine
Pei-Cheng Cheng, Been-Chian Chien, Hao-Ren Ke, Wei-Pang Yang ........................................ 750

A Probabilistic Approach to Medical Image Retrieval
Koen Lubbers, Arjen P. de Vries, Theo Huibers, Paul van der Vet .......... 761

UB at CLEF2004 Cross Language Medical Image Retrieval
Miguel E. Ruiz, Munirathnam Srikanth ........................................... 773

Content-Based Queries on the CasImage Database Within the IRMA Framework
Christian Thies, Mark Oliver G"uld, Benedikt Fischer, Thomas M. Lehmann ........................................ 781

Comparison and Combination of Textual and Visual Features for Interactive Cross-Language Image Retrieval
Pei-Cheng Cheng, Jen-Yuan Yeh, Hao-Ren Ke, Been-Chian Chien, Wei-Pang Yang ........................................ 793

MSU at ImageCLEF: Cross Language and Interactive Image Retrieval
Vineet Bansal, Chen Zhang, Joyce Y. Chai, Rong Jin ...................... 805

Part VI. Cross-Language Spoken Document Retrieval

CLEF 2004 Cross-Language Spoken Document Retrieval Track
Marcello Federico, Nicola Bertoldi, Gina-Anne Levow, Gareth J.F. Jones ........................................... 816

Part VII. Issues in CLIR and in Evaluation

The Key to the First CLEF with Portuguese: Topics, Questions and Answers in CHAVE
Diana Santos, Paulo Rocha ............................................................. 821

How Do Named Entities Contribute to Retrieval Effectiveness?
Thomas Mandl, Christa Womser-Hacker ........................................ 833

Author Index ............................................................................ 843