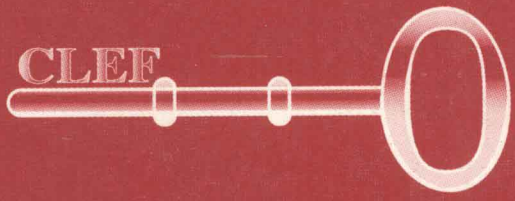
Carol Peters
Julio Gonzalo
Martin Braschler
Michael Kluck (Eds.)

LNCS 3237

Comparative Evaluation of Multilingual Information Access Systems

4th Workshop of the
Cross-Language Evaluation Forum, CLEF 2003
Trondheim, Norway, August 2003, Revised Papers



Springer

Carol Peters Julio Gonzalo
Martin Braschler Michael Kluck (Eds.)

Comparative Evaluation of Multilingual Information Access Systems

4th Workshop of the
Cross-Language Evaluation Forum, CLEF 2003
Trondheim, Norway, August 21-22, 2003
Revised Papers

Springer

Volume Editors

Carol Peters
Istituto di Scienza e Tecnologie dell'Informazione
Consiglio Nazionale delle Ricerche (ISTI-CNR)
Via G. Moruzzi 1, 56124 Pisa, Italy
E-mail: carol.peters@isti.cnr.it

Julio Gonzalo
Universidad Nacional de Educación a Distancia (UNED)
Departamento de Lenguajes y Sistemas Informáticos
c/ Juan del Rosal, 16, 28040 Madrid, Spain
E-mail: julio@lsi.uned.es

Martin Braschler
Eurospider Information Technology AG
Schaffhauser Str. 18, 8006 Zürich, Switzerland
E-mail: martin.braschler@eurospider.com

Michael Kluck
Informationszentrum Sozialwissenschaften der
Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V. (IZ)
Lennéstr. 30, 53113 Bonn, Germany
E-mail: kluck@bonn.iz-soz.de

Library of Congress Control Number: 2004115726

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

ISSN 0302-9743
ISBN 3-540-24017-9 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

springeronline.com

© Springer-Verlag Berlin Heidelberg 2004
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11342052 06/3142 5 4 3 2 1 0

Preface

The fourth campaign of the Cross-language Evaluation Forum (CLEF) for European languages was held from January to August 2003. Participation in this campaign showed a slight rise in the number of participants from the previous year, with 42 groups submitting results for one or more of the different tracks (compared with 37 in 2002), but a steep rise in the number of experiments attempted. A distinctive feature of CLEF 2003 was the number of new tracks and tasks that were offered as pilot experiments. The aim was to try out new ideas and to encourage the development of new evaluation methodologies, suited to the emerging requirements of both system developers and users with respect to today's digital collections and to encourage work on many European languages rather than just those most widely used. CLEF is thus gradually pushing its participants towards the ultimate goal: the development of truly multilingual systems capable of processing collections in diverse media.

The campaign culminated in a two-day workshop held in Trondheim, Norway, 21–22 August, immediately following the 7th European Conference on Digital Libraries (ECDL 2003), and attended by more than 70 researchers and system developers. The objective of the workshop was to bring together the groups that had participated in the CLEF 2003 campaign so that they could report on the results of their experiments. Attendance at the workshop was thus limited to participants in the campaign plus several invited guests with recognized expertise in the multilingual information access field. This volume contains thoroughly revised and expanded versions of the preliminary papers presented at the workshop accompanied by detailed analyses of the results in the various track overview papers and thus provides an exhaustive record of the CLEF 2003 campaign.

CLEF 2003 was conducted within the framework of a project of the Information Society Technologies programme of the European Commission (IST-2000-31002). The campaign was organized in collaboration with the US National Institute of Standards and Technology (NIST) and with the support of the DELOS Network of Excellence for Digital Libraries. The support of NIST and DELOS in the running of the evaluation campaign is gratefully acknowledged. We would also like to thank the other members of the Workshop Steering Committee for their assistance in the coordination of this event.

June 2004

Carol Peters
Julio Gonzalo
Martin Braschler
Michael Kluck

CLEF 2003 Workshop Steering Committee

Martin Braschler, Eurospider Information Technology, Switzerland
Khalid Choukri, Evaluations and Language Resources Distribution Agency, Paris, France
Marcello Federico, Centro per la Ricerca Scientifica e Tecnologica, Istituto Trentino di Cultura, Italy
Julio Gonzalo Arroyo, Departamento de Lenguajes y Sistemas Informáticos, Universidad Nacional de Educación a Distancia, Madrid, Spain
Donna Harman, National Institute of Standards and Technology, USA
Gareth Jones, University of Exeter, UK
Noriko Kando, National Institute of Informatics, Japan
Michael Kluck, IZ Sozialwissenschaften, Bonn, Germany
Bernardo Magnini, Centro per la Ricerca Scientifica e Tecnologica, Istituto Trentino di Cultura, Italy
Douglas W. Oard, University of Maryland, USA
Carol Peters, ISTI-CNR, Pisa, Italy
Mark Sanderson, University of Sheffield, UK
Peter Schäuble, Eurospider Information Technology, Switzerland
Ellen Voorhees, National Institute of Standards and Technology, USA

此为试读,需要完整PDF请访问: www.ertongbook.com

Table of Contents

Introduction
Julio Gonzalo, Carol Peters 1

CLEF 2003 Methodology and Metrics
Martin Braschler, Carol Peters 7

Analysis of the Reliability of the Multilingual Topic Set for the Cross Language Evaluation Forum
Thomas Mandl, Christa Womser-Hacker 21

Evaluation of Information Access Technologies at the NTCIR Workshop
Noriko Kando 29

Part I. Ad-hoc Text Retrieval Tracks

CLEF 2003 – Overview of Results
Martin Braschler 44

Mainly Cross-Language Experiments

Report on CLEF-2003 Multilingual Tracks
Jacques Savoy 64

The Impact of Word Normalization Methods and Merging Strategies on Multilingual IR
Eija Airio, Heikki Keskustalo, Turid Hedlund, Ari Pirkola 74

JHU/APL Experiments in Tokenization and Non-word Translation
Paul McNamee, James Mayfield 85

Report on CLEF-2003 Experiments: Two Ways of Extracting Multilingual Resources from Corpora
Nicola Cancedda, Hervé Déjean, Éric Gaussier, Jean-Michel Renders, Alexei Vinokourov 98

Combining Query Translation and Document Translation in Cross-Language Retrieval
Aitao Chen, Fredric C. Gey 108

Cross-Language Experiments with the IR-n System
Fernando Llopis, Rafael Muñoz 122

Multilingual Information Retrieval Using Open, Transparent Resources in CLEF 2003
Monica Rogati, Yiming Yang 133

ITC-irst at CLEF 2003: Monolingual, Bilingual and Multilingual Information Retrieval
Nicola Bertoldi, Marcello Federico 140

Language-Dependent and Language-Independent Approaches to Cross-Lingual Text Retrieval
Jaap Kamps, Christof Monz, Maarten de Rijke, Börkur Sigurbjörnsson 152

Multilingual Retrieval Experiments with MIMOR at the University of Hildesheim
René Hackl, Ralph Kölle, Thomas Mandl, Alexandra Ploedt, Jan-Hendrik Scheufen, Christa Womser-Hacker 166

Concept-Based Searching and Merging for Multilingual Information Retrieval: First Experiments at CLEF 2003
Romaric Besançon, Gaël de Chalendar, Olivier Ferret, Christian Fluhr, Olivier Mesnard, Hubert Naets 174

CLEF 2003 Experiments at UB: Automatically Generated Phrases and Relevance Feedback for Improving CLIR
Miguel E. Ruiz 185

SINAI at CLEF 2003: Decompounding and Merging
Fernando Martínez-Santiago, Arturo Montejo-Ráez, Luis Alfonso Ureña-López, M. Carlos Díaz-Galiano 192

Merging Results by Predicted Retrieval Effectiveness
Wen-Cheng Lin, Hsin-Hsi Chen 202

MIRACLE Approaches to Multilingual Information Retrieval: A Baseline for Future Research
José L. Martínez, Julio Villena, Jorge Fombella, Ana G. Serrano, Paloma Martínez, José M. Goñi, José C. González 210

Experiments to Evaluate Probabilistic Models for Automatic Stemmer Generation and Query Word Translation
Giorgio M. Di Nunzio, Nicola Ferro, Massimo Melucci, Nicola Orio 220

Clairvoyance CLEF-2003 Experiments
Yan Qu, Gregory Grefenstette, David A. Evans 236

Simple Translations of Monolingual Queries Expanded through an Association Thesaurus. X-IOTA IR system for CLIPS Bilingual Experiments
Gilles Sérasset, Jean-Pierre Chevallet 242

Two-Stage Refinement of Query Translation in a Pivot Language Approach to Cross-Lingual Information Retrieval: An Experiment at CLEF 2003
Kazuaki Kishida, Noriko Kando 253

Regular Sound Changes for Cross-Language Information Retrieval
Michael P. Oakes, Souvik Banerjee 263

Exeter at CLEF 2003: Experiments with Machine Translation for Monolingual, Bilingual and Multilingual Retrieval
Adenike M. Lam-Adesina, Gareth J. F. Jones 271

Monolingual Experiments

Lexical and Algorithmic Stemming Compared for 9 European Languages with Hummingbird SearchServer™ at CLEF 2003
Stephen Tomlinson 286

Océ at CLEF 2003
Roel Brand, Marvin Brünner, Samuel Driessen, Pascha Iljin, Jakob Klok 301

Comparing Weighting Models for Monolingual Information Retrieval
Gianni Amati, Claudio Carpineto, Giovanni Romano 310

Pruning Texts with NLP and Expanding Queries with an Ontology: TagSearch
Gil Francopoulo 319

Report on CLEF-2003 Monolingual Tracks: Fusion of Probabilistic Models for Effective Monolingual Retrieval
Jacques Savoy 322

Selective Compound Splitting of Swedish Queries for Boolean Combinations of Truncated Terms
Rickard Cöster, Magnus Sahlgren, Jussi Karlgren 337

COLE Experiments at CLEF 2003 in the Spanish Monolingual Track
Jesús Vilares, Miguel A. Alonso, Francisco J. Ribadas 345

Experiments with Self Organizing Maps in CLEF 2003
Javier Fernández, Ricardo Mones, Irene Díaz, José Ranilla, Elías F. Combarro 358

Ricoh at CLEF 2003
Yuichi Kojima, Hideo Itoh, Hiroko Mano, Yasushi Ogawa 367

MediaLab at CLEF 2003: Using Keyword Disambiguation
Peter van der Weerd 373

Part II. Domain-Specific Document Retrieval

The GIRT Data in the Evaluation of CLIR Systems - from 1997 Until 2003
Michael Kluck 376

A Data-Compression Approach to the Monolingual GIRT Task: An Agnostic Point of View
Daniela Alderuccio, Luciana Bordoni, Vittorio Loreto 391

UC Berkeley at CLEF-2003 – Russian Language Experiments and Domain-Specific Retrieval
Vivien Petras, Natalia Perelman, Fredric Gey 401

University of Hagen at CLEF 2003: Natural Language Access to the GIRT4 Data
Johannes Leveling 412

Part III. Interactive Cross-Language Retrieval

The CLEF 2003 Interactive Track
Douglas W. Oard, Julio Gonzalo 425

iCLEF 2003 at Maryland: Translation Selection and Document Selection
Bonnie Dorr, Daqing He, Jun Luo, Douglas W. Oard, Richard Schwartz, Jianqiang Wang, David Zajic 435

UNED at iCLEF 2003: Searching Cross-Language Summaries
Fernando López-Ostenero, Julio Gonzalo, Felisa Verdejo 450

Comparing Syntactic-Semantic Patterns and Passages in Interactive Cross Language Information Access
Borja Navarro, Fernando Llopis, Miguel Ángel Varó 462

Continued Experiments on Cross-Language Relevance Assessment
Jussi Karlgren, Preben Hansen 468

Part IV. Cross-Language Question Answering

The Multiple Language Question Answering Track at CLEF 2003
Bernardo Magnini, Simone Romagnoli, Alessandro Vallin, Jesús Herrera, Anselmo Peñas, Víctor Peinado, Felisa Verdejo, Maarten de Rijke 471

Creating the DISEQuA Corpus: A Test Set for Multilingual Question Answering
Bernardo Magnini, Simone Romagnoli, Alessandro Vallin, Jesús Herrera, Anselmo Peñas, Víctor Peinado, Felisa Verdejo, Maarten de Rijke 487

Bridging Languages for Question Answering: DIOGENE at CLEF 2003
Matteo Negri, Hristo Tanev, Bernardo Magnini 501

Cross-Language Question Answering at the USC Information Sciences Institute
Abdessamad Echihabi, Douglas W. Oard, Daniel Marcu, Ulf Hermjakob 514

How Frogs Built the Berlin Wall. A Detailed Error Analysis of a Question Answering System for Dutch
Valentin Jijkoun, Gilad Mishne, Maarten de Rijke 523

Cross-Lingual QA: A Modular Baseline in CLEF 2003
Lucian Vlad Lita, Monica Rogati, Jaime Carbonell 535

Question Answering in Spanish
José L. Vicedo, Ruben Izquierdo, Fernando Llopis, Rafael Muñoz 541

Quantum, a French/English Cross-Language Question Answering System
Luc Plamondon, George Foster 549

A Cross–Language Question/Answering–System for German and English
Günter Neumann, Bogdan Sacaleanu 559

Cross-Language French-English Question Answering Using the DLT System at CLEF 2003
Richard F. E. Sutcliffe, Igal Gabbay, Aoife O'Gorman 572

Part V. Cross-Language Image Retrieval

The CLEF 2003 Cross Language Image Retrieval Track
Paul Clough, Mark Sanderson 581

Assessing Translation Quality for Cross Language Image Retrieval
Paul Clough, Mark Sanderson 594

Foreign Name Backward Transliteration in Chinese-English Cross-Language Image Retrieval
Wen-Cheng Lin, Changhua Yang, Hsin-Hsi Chen 611

Image Retrieval: the MIRACLE Approach
Julio Villena, José L. Martínez, Jorge Fombella, Ana G. Serrano, Alberto Ruiz, Paloma Martínez, José M. Goñi, José C. González 621

Scene of Crime Information System: Playing at St. Andrews
Bogdan Vrusias, Mariam Tariq, Lee Gillam 631

Part VI. Cross-Language Spoken Document Retrieval

The CLEF 2003 Cross-Language Spoken Document Retrieval Track
Marcello Federico, Gareth J. F. Jones 646

Exeter at CLEF 2003: Cross-Language Spoken Document Retrieval Experiments
Gareth J. F. Jones, Adenike Lam-Adesina 653

N-grams for Translation and Retrieval in CL-SDR
Paul McNamee, James Mayfield 658

Spoken Document Retrieval Experiments with IR-n System
Fernando Llopis, Patricio Martínez-Barco 664

ITC-irst at CLEF 2003: Cross-Language Spoken Document Retrieval
Nicola Bertoldi, Marcello Federico 672

Appendix 677

Author Index 701

此为试读,需要完整PDF请访问: www.ertongbook.com

Introduction

Julio Gonzalo[1] and Carol Peters[2]

[1] UNED, c/Juan del Rosal, 16, 28040 Madrid, Spain
julio@lsi.uned.es
[2] ISTI-CNR, Area di Ricerca, 56124 Pisa, Italy
carol.peters@isti.cnr.it

This volume reports the results of the fourth campaign of the Cross Language Evaluation Forum (CLEF). This campaign consisted of a number of comparative evaluations for multilingual systems (and system components) performing document retrieval (on open-domain or domain-specific collections, and involving both batch and interactive systems), and for cross-language question answering, image and spoken document retrieval.

42 research groups from industry and academia, from 14 different countries around the globe, participated in this comparative research initiative. As in previous years, the participating groups consisted of a nice mix of new-comers (14) and veteran groups (28) coming back for a second, third or even fourth time. Another important trend that was again noticeable was the progression of many of the returning groups to a more complex task, from monolingual to bilingual, from bilingual to multilingual, from pure text retrieval tasks to tasks that involve searching collections in multimedia.

The results of the experiments of the CLEF 2003 campaign were first presented at the Workshop, held in Trondheim, Norway, 21-22 August. These proceedings contain thoroughly revised and expanded versions of the preliminary system reports published in the CLEF 2003 Working Notes and distributed at the Workshop. Many of the papers also include descriptions of additional experiments and results as groups often further optimize their systems or try out new ideas as a consequence of the discussions during the workshop.

From Cross-Language Text Retrieval to Multilingual Information Access

From its initial campaign in 2000, the Cross Language Evaluation Forum (CLEF) has fostered the creation of a research and development community around the Cross Language Information Retrieval (CLIR) domain, broadly understood as that sector of multidisciplinary research interested in the challenge of retrieving information across language boundaries. Up until 2003, the main focus of CLEF was the study of the multilingual text retrieval problem, defined as a fully automatic process in which a query (a statement of user needs) in one language is used to retrieve a single ranked set of documents from a text collection in a number of languages. Over the years, CLEF has considered and evaluated variants and components of this challenge: bilingual information retrieval (where a query

C. Peters et al. (Eds.): CLEF 2003, LNCS 3237, pp. 1–6, 2004.
© Springer-Verlag Berlin Heidelberg 2004

in one language searches in a document collection in a different language) and monolingual information retrieval (in European languages other than English). The strategy adopted has been to offer a series of progressively more complex tasks on an increasingly large and varied multilingual corpus, in order to stimulate the development of fully multilingual retrieval systems capable of handling many languages at the same time and of easily adapting to meet the demands of new languages.

However, from the very beginnings, it was clear to the organisers of CLEF that retrieving documents across language boundaries is only a component of usable multilingual search applications, albeit a core one. The real challenge is broader: helping users to search, browse, recognize and use information (rather than documents) from (possibly interlinked) sets of multilingual, multimedia information objects. Relevant topics include not only multilingual document retrieval, but also image retrieval (with image captions in different languages), cross-language speech retrieval, multilingual information extraction and question answering systems, the interactive aspects of multilingual retrieval, etc. Researchers often refer to this broader problem as "Multilingual Information Access (MLIA)". We thus decided that a goal of CLEF should be to stimulate system research and development in this wider direction.

CLEF 2003 has thus made an initial - and promising - move into the evaluation of multilingual information access systems with the introduction of a number of pilot experiments aimed at investigating different aspects of the CLIR/MLIA paradigm. The paper by Braschler and Peters, this volume, gives an overview of the organisation of CLEF 2003 with brief descriptions of all the different tracks offered and the test collections provided. Here below we attempt to summarise the range of Multilingual Information Access problems addressed in the 2003 campaign:

Ad hoc Open-Domain and Domain-Specific Text Retrieval: The so-called ad hoc tracks (testing monolingual, bilingual and multilingual text retrieval systems) and the domain-specific retrieval track remained the core CLEF tracks in the 2003 campaign. The goal is to offer participants the chance to test and tune systems handling many different languages and searching across languages, to investigate the problems involved, and to experiment with new approaches. As in previous editions, a lot of work was done on different kinds of text indexing, on experimenting with various types of translation resources and combinations of them and, in the multilingual tasks, on the best methods for merging the documents found in collections in different languages into a single result set. The overview paper by Braschler, this volume, discusses the trends observed in these tracks in 2003 in much more detail.

Multilingual Question Answering (QA@CLEF): An exploratory track introduced in CLEF 2003 aimed at testing question answering systems finding exact answers to open-domain questions in target collections in languages other than English (the monolingual QA task) or in languages other than that of the question (the cross-language QA task). The participation in the track was

limited to just eight groups as this was the very first time that cross-language QA system evaluation had been offered to the community and research in this area was only in its initial stages. However, the track attracted a lot of attention, and is already helping to boost research in the topic. Evidence of this is the fact that eighteen groups have registered for this track in CLEF 2004. The results of this initial QA track have shown that the cross-language QA problem is significantly harder than its monolingual counterpart, and perhaps more difficult to solve accurately than cross-language document retrieval, because the "bag of translations" approach is far too crude to locate specific pieces of information within a text.

Cross-Language Spoken Document Retrieval (CL-SDR): A second CLEF 2003 pilot experiment regarded cross-language spoken document retrieval. This track explored the problem of retrieving voice recordings, via automatic transcriptions, in languages different from the query language. Again, this problem introduces new challenges in the multilingual context, because some state-of-the-art translation techniques cannot be directly applied on error-prone speech transcriptions. Although this track is currently still concerned with a particular type of text retrieval - written queries in one language are matched somehow against automatically transcribed spoken documents - it represents a first step towards the development of truly multi-media cross-language retrieval systems.

Cross-Language Retrieval in Image Collections (ImageCLEF): Another step in the multimedia direction was the introduction of a cross-language track on an image collection. Cross-language retrieval using image captions in a different language from the query is perhaps the only cross-language information access problem which is directly helpful for users, without further translation aids. Images are generally interpretable whatever the native language of the user. At the same time, this is a particularly challenging task from the point of view of cross-language text retrieval, because image captions are significantly smaller and less redundant than documents. It is thus to be expected that different retrieval strategies could be needed. This pilot track had considerable success, and has lead to a continuation and considerable expansion both of the tasks offered and of the scope in CLEF 2004, which includes content-based retrieval and interactive image retrieval. An idea of the interest aroused by this track is shown by the fact that while just four groups participated in the 2003 experiment, twenty-one different groups have registered for 2004.

Interactive Cross-Language Information Retrieval (iCLEF): From a user point of view, a system that accepts a query in his/her native language and returns a list of foreign-language documents is only one component in cross-language search assistance. For instance, how will the user recognize relevant information in an unfamiliar language? And how can the user best formulate, translate and refine queries? The interactive track addresses these problems. iCLEF has been part of CLEF since its second campaign in 2001. Unlike interactive monolingual retrieval (where it has been traditionally difficult to

measure quantitative differences between alternative approaches), different interactive strategies have been shown to produce quite different results (in terms of precision and recall) in this evaluation framework.

Altogether, the CLEF 2003 tracks constitute the largest comparative evaluation of multilingual information access components and applications ever attempted and the success of the pilot tracks has lead to their confirmation and expansion in CLEF 2004.

Reporting the CLEF 2003 Experiments

This volume is organized into separate sections for each of the main evaluation tasks discussed above. However, it begins with three introductory papers discussing different aspects of the organisation of information retrieval evaluation campaigns. The first two papers focus on the organisation of CLEF: Martin Braschler and Carol Peters describe the overall organization of the CLEF 2003 evaluation campaign, with a particular focus on the cross-language ad hoc and domain-specific retrieval tracks and Thomas Mandl and Christa Womser-Hacker analyse the reliability of the CLEF multilingual topic set for the cross-language document retrieval tasks. Finally, Noriko Kando presents the evaluation experience of the NTCIR workshop, which is the main forum for the evaluation of Information Access technologies in Asian languages.

The rest of the volume is structured as follows. Part I is dedicated to the ad-hoc retrieval tracks and has two sub-sections. The first reports on cross-language work - both multilingual and bilingual - while the second contains those papers describing the specifically monolingual-only experiments. The thirty one papers included cover most of the state-of-the-art approaches to cross-language and monolingual retrieval problems in the multiple language context and also present many new and interesting ideas. This section begins with an overview of the trends observed and the results obtained in these tracks by Martin Braschler. Part II presents the experiments in domain-specific document retrieval on the GIRT collection of structured social science documents. It starts with an overview by Michael Kluck and includes three papers describing monolingual and cross-language experiments on structured document retrieval. Part III describes the user-inclusive experiments in the iCLEF track, starting with an overview by the track organizers (Doug Oard and Julio Gonzalo) and reporting six experiments on query reformulation and/or document selection issues.

The rest of the volume is devoted to other kinds of information access tasks: Part IV consists of ten pioneering papers on multilingual question answering. It begins with two papers providing a track overview and a detailed description of the test corpus that has been constructed by this track from the international group that coordinated these experiments, headed by Bernardo Magnini. Part V includes five papers on cross-language image retrieval, beginning with a track overview by Paul Clough and Mark Sanderson. Finally, Part VI describes in detail the pilot evaluation of cross-language spoken document retrieval systems at CLEF, and includes an overview by the track coordinators, Marcello Federico

and Gareth Jones, followed by four other papers giving detailed descriptions of the experiments.

The volume ends with an Appendix listing the results of all runs submitted to the ad-hoc retrieval tasks. For reasons of space, only the most significant figures for each run are included. An exhaustive listing of the results can be found on the main CLEF website in the Working Notes for CLEF2003.

For more information on the activities of CLEF and the agenda for CLEF 2004, visit the CLEF websites:

CLEF main site	`www.clef-campaign.org`
Interactive retrieval (iCLEF)	`nlp.uned.es/iCLEF`
Question Answering (QA@CLEF)	`clef-qa.itc.it`
Image Retrieval (ImageCLEF)	`ir.shef.ac.uk/imageclef2004`
Spoken Document Retrieval (CL-SDR)	`hermes.itc.it/clef-sdr04.html`

Acknowledgments

We have numerous people and organisations to thank for their help in the running of CLEF 2003. First of all, we should like to thank the other members of the CLEF Consortium and, in particular, Martin Braschler, Eurospider, Khalid Choukri, ELRA/ELDA, Donna Harman, NIST, and Michael Kluck, IZ-Bonn, for all their efforts at making both the campaign and the workshop a great success. We also express our immense gratitude to Francesca Borri, ISTI-CNR, who has been responsible for the administrative management of the CLEF project and has worked very hard to ensure a smooth day-by-day organisation. However, it would be impossible to run CLEF without considerable assistance from many other groups, working mainly on a voluntary basis. Here below we list some of them:

Associated Members of the CLEF Consortium

- Department of Information Studies, University of Tampere, Finland - responsible for work on the Finnish collection
- Human Computer Interaction and Language Engineering Laboratory, SICS, Kista, Sweden - responsible for work on the Swedish collection
- University of Twente, Centre for Telematics and Information Technology, The Netherlands - responsible for work on the Dutch collection
- Universität Hildesheim, Institut für Angewandte Sprachwissenschaft - Informationswissenschaft, Germany - responsible for checking and revision of the multilingual topic set
- College of Information Studies and Institute for Advanced Computer Studies, University of Maryland, College Park, MD, USA - co-organisers of iCLEF
- Centro per la Ricerca Scientifica e Tecnologica, Istituto Trentino di Cultura, Italy, main coordinators of the multilingual question answering track and co-organisers of the cross-language spoken document retrieval track
- University of Exeter, co-organisers of the cross-language spoken document retrieval track
- University of Sheffield, coordinators of the cross-language image retrieval track

Furthermore, we should like to thank colleagues from the Natural Language Processing Lab, Department of Computer Science and Information Engineering, National Taiwan University, for preparing topics in Chinese, the National Institute of Informatics, Tokyo, for the post-campaign Japanese topics, and the Moscow State University, Russia, for their assistance in obtaining the Russian collection.

We also gratefully acknowledge the support of all the data providers and copyright holders, and in particular:

- The Los Angeles Times, for the American English data collection;
- Le Monde S.A. and ELDA: Evaluations and Language resources Distribution Agency, for the French data;
- Frankfurter Rundschau, Druck und Verlagshaus Frankfurt am Main; Der Spiegel, Spiegel Verlag, Hamburg, for the German newspaper collections;
- InformationsZentrum Sozialwissenschaften, Bonn, for the GIRT database;
- Hypersystems Srl, Torino and La Stampa, for the Italian newspaper data;
- Agencia EFE S.A. for the Spanish newswire data;
- NRC Handelsblad, Algemeen Dagblad and PCM Landelijke dagbladen/Het Parool for the Dutch newspaper data;
- Aamulehti Oyj for the Finnish newspaper documents;
- Tidningarnas Telegrambyrå for the Swedish newspapers;
- The Herald 1995, SMG Newspapers, for the British English newspaper data;
- Schweizerische Depeschenagentur, Switzerland, for the French, German and Italian Swiss news agency data;
- Russika-Izvestia for the Russian collection;
- St Andrews University Library for the image collection;
- NIST for access to the TREC-8 and TREC-9 SDR transcripts.

Without their help, this evaluation activity would be impossible.

此为试读,需要完整PDF请访问: www.ertongbook.com