

Introduction

Julio Gonzalo¹ and Carol Peters²

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

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 led 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 led 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 Depeschagentur, 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.