

Introduction

Carol Peters
Istituto di Scienza e Tecnologie dell'Informazione (ISTI-CNR), Pisa, Italy
carol@iei.pi.cnr.it

These Working Notes contain descriptions of the experiments conducted within CLEF 2002 – the third in a series of annual system evaluation campaigns organised by the Cross-Language Evaluation Forum¹. The results of the experiments will be presented and discussed in the CLEF 2002 Workshop, 19-20 September, Rome, Italy. The final papers - revised and extended as a result of the discussions at the Workshop - together with a comparative analysis of the results will appear in the CLEF 2002 Proceedings. These will be published by Springer in their Lecture Notes for Computer Science series.

CLEF organises a series of evaluation tracks designed to test different aspects of mono- and cross-language information retrieval system development. The intention is to encourage systems to move from monolingual searching to the implementation of a full multilingual retrieval service. The main features of the 2002 campaign are briefly outlined here below in order to provide the necessary background to the experiments reported in this volume.

Tasks

CLEF 2002 offered five separate tracks evaluating the performance of systems for:

- multilingual information retrieval
- bilingual information retrieval
- monolingual (non-English) information retrieval
- information retrieval for scientific texts
- interactive cross-language information retrieval

Multilingual/Bilingual/Monolingual Information Retrieval: The main track in CLEF is the multilingual one. Using a chosen query language, the goal is to retrieve relevant documents for all languages in a collection, rather than just a given pair, listing the results in a merged, ranked list. Similarly to last year, the CLEF 2002 collection for this track contained English, German, French, Italian and Spanish documents. In the bilingual track, target document collections in Dutch, Finnish, French, German, Italian, Spanish or Swedish could be searched. First-time CLEF participants only could choose to search the English document collection. CLEF 2002 also offered tasks for Dutch, Finnish, French, German, Italian, Spanish and Swedish monolingual system testing.

For each of these tasks, the participating systems constructed their queries (automatically or manually) from a common set of 50 structured topics.

Mono- and Cross-Language Information Retrieval for Scientific Texts: The rationale in this track is to study information retrieval on other types of collections, serving a different kind of information need. This year's track offered mono- and cross-language tasks for two different collections: **AMARYLLIS**: a multi-disciplinary collection of bibliographic documents in French; **GIRT**: a structured database containing German social science documents. 25 topic statements were provided for both tasks.

Interactive CLIR: The interactive track was first introduced as an experiment in CLEF 2001 – in order to provide the opportunity for groups to test user satisfaction issues. CLEF 2001 focussed on the document selection problem. The success of this experiment led to the interactive task being offered as a main track in CLEF 2002. The aim this year was to explore interactive formulation of cross-language queries and/or cross-language document selection in ways that assist users unfamiliar with the target language.

¹ From October 2001, CLEF is supported by the European Commission under the IST programme (IST-2000-31002). The consortium members are: ISTI-CNR, Pisa; IZ Sozialwissenschaften, Bonn; ELRA/ELDA, Paris; Eurospider, Zurich; UNED, Madrid; NIST, USA. For more information, see: <http://www.clef-campaign.org>.

Test Collections

The CLEF test collections are formed of sets of documents in different European languages but with common features (e.g., same genre and time period, comparable content); sets of topics rendered in a number of languages; relevance judgements determining the set of relevant documents for each topic and collection.

Multilingual Corpus: The main document collection for CLEF 2002 was larger than that used in 2001, containing over 1,000,000 documents in eight languages: Dutch, English, Finnish, French, German, Italian, Spanish and Swedish – two more than last year². It contained both newswires and national newspapers.

Two distinct scientific collections were also available: the GIRT database of about 80,000 German social science documents, which has controlled vocabularies for English-German and German-Russian, and the Amaryllis multidisciplinary database of approximately 150,000 French bibliographic documents and a controlled vocabulary in English and French.

Topics: Participating groups derived their queries in their preferred language from a common set of topics, created to simulate user information needs. Each topic consists of three parts: a brief title statement; a one-sentence description; a more complex narrative specifying the relevance assessment criteria. Topic sets were produced by native speakers in the eight document languages and additionally for Russian, Portuguese and Chinese. A Japanese topic set has also been made available now and can be used by participants for further experiments. As in CLEF 2001, a condition was that, for each task attempted, a mandatory run using the title and description fields had to be submitted. The objective is to facilitate comparison between the results of different systems.

Relevance Judgements: Relevance assessment was distributed over eight different sites and performed in all cases by native speakers. The number of documents in large test collections such as CLEF makes it impractical to judge every document for relevance. Instead approximate recall techniques are calculated using pooling techniques. The results submitted by the participating groups were used to form a pool of documents for each topic and for each language by collecting the highly ranked documents from all submissions. The results were then analysed and run statistics produced and distributed.

Participants

Participation in CLEF 2002 shows a slight rise in number of groups with respect to last year, with 37 groups submitting results for one or more of the main tracks: 8 from N.America; 28 from Europe, and just 1 from Asia – compared with 34 groups for CLEF 2001. However, close to 300 system runs were received; this is a 50% increase over last year and is a direct consequence of the new tasks/collections introduced. Runs were submitted for all tasks (multilingual, bilingual, monolingual, GIRT and Amaryllis) and for all topic languages. 22 groups tried a cross-language task and eleven of these submitted results for the multilingual track; 13 submitted runs for the monolingual track only. Five groups tackled iCLEF, compared with three in the experimental interactive task of 2001. Participating groups consist of a nice mix of first-comers and groups coming back for a second or third time. It is interesting to note that while eight of the twelve newcomers submitted runs to monolingual tracks only, a number of the groups that had participated before attempted a more complex task in CLEF 2002. Thus, we had groups that had submitted monolingual runs last year also participating in the bilingual task this year, and groups that had only tried bilingual last year moving on to the multilingual track this year.

Working Notes and Workshop

The Working Notes provide a first description of the different experiments run by the participating groups. The volume is divided into four parts: Cross-language and More, Monolingual Experiments, Mainly Domain-Specific IR and Interactive CLEF. Papers have been collocated in the section considered most appropriate, even though many papers describe more than one type of experiment. The Appendix gives a list of the participants and a summary of the characteristics of all runs for the main tracks together with overview graphs for the different tasks and individual statistics for each run. It contains one page for each result set submitted by a participating group.

The aim of the Workshop is to give all the groups that have participated in the CLEF 2002 evaluation campaign the opportunity to get together in order to compare approaches and to exchange ideas. It will also provide the opportunity for an open discussion on the organisation and scheduling of future CLEF evaluation campaigns.

² The new languages were Finnish and Swedish.

The work of the groups participating in this year's campaign will be presented in paper and poster sessions. Additional talks will include descriptions of the results of the other two major international cross-language evaluation initiatives, a summary of the conclusions of a workshop held at SIGIR2002 that aimed at producing a roadmap for research in the cross-language information retrieval field, and proposals for future evaluation tasks. The presentations at the Workshop will be posted on the CLEF website.

We very much hope that the Workshop will prove an interesting, worthwhile and enjoyable experience to all those who participate.

The Workshop Steering Committee

Martin Braschler, Eurospider, Switzerland
Khalid Choukri, ELDA/ELRA, Paris, France
Julio Gonzalo Arroyo, UNED, Madrid, Spain
Donna Harman, National Institute of Standards and Technology, USA
Noriko Kando, National Institute of Informatics, Japan
Michael Kluck, IZ Sozialwissenschaften, Bonn, Germany
Patrick Kremer, INstitute de Information Scientifique et Technique, Vandoeuvre, France
Carol Peters, ISTI-CNR, Pisa, Italy
Peter Schäuble, Eurospider, Switzerland
Laurent Schmitt INstitute de Information Scientifique et Technique, Vandoeuvre, France
Ellen Voorhees, National Institute of Standards and Technology, USA

Local Organisation

Francesca Borri, ISTI-CNR, Pisa, Italy

Acknowledgements

First of all, we should like to express our gratitude to the ECDL 2002 Conference organisers for their assistance in the preparation of the CLEF Workshop, and in particular to Tarina Ayazi for all her hard work.

It would be impossible to run CLEF without considerable assistance from many groups, working mainly on a voluntarily basis. We have many people and organisations to thank for their help in the running of CLEF 2002. Here below we list some of them:

Associated Members of the CLEF Consortium

- INIST - INstitute de Information Scientifique et Technique, Vandoeuvre, France – responsible for the Amaryllis task
- 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
- Universitatet 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

In addition, we should like to thank the following people/organisations for preparing topics in Chinese, Japanese and Portuguese.

- Natural Language Processing Lab, Department of Computer Science and Information Engineering, National Taiwan University
- National Institute of Informatics (NII), Tokyo
- José Borbinha and Eulalia Carvalho of the National Library of Portugal

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

- The Los Angeles Times, for the English data collection;
- Le Monde S.A. and ELDA: European 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.
- INIST - INstitute de Information Scientifique et Technique, Vandoeuvre, France, for the Amaryllis data.
- 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
- Schweizerische Depeschagentur, Switzerland, for the French, German and Italian Swiss news agency data.

Without their help, this evaluation activity would be impossible.