

Foreword

The Cross-Language Evaluation Forum (CLEF)¹ aims at promoting research and development in Cross-Language Information Retrieval (CLIR) by (i) providing an infrastructure for the testing and evaluation of information retrieval systems operating on European languages, and (ii) creating test-suites of reusable data which can be employed by system developers for benchmarking purposes. These objectives are being achieved through the organisation of a series of annual system evaluation campaigns. The Working Notes report the preliminary results of CLEF 2001 – the second campaign in the series. The results will be presented and discussed in the CLEF 2001 Workshop, 3-4 September, Darmstadt, Germany. The main features of this year's campaign are briefly outlined here below.

Tasks

Similarly to last year, CLEF 2001 offered four main evaluation tracks:

- multilingual information retrieval
- bilingual information retrieval
- monolingual (non-English) information retrieval
- domain-specific system evaluation.

For each task, participating systems constructed queries (automatically or manually) from a common set of statements of information needs (topics) and searched for relevant documents in the collections provided. Results were submitted in a ranked list in decreasing order of estimated relevance.

In addition, this year we also included an experimental track testing interactive cross-language systems.

Test Collection

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

Multilingual Corpus: The 2001 document collection is considerably larger than that used in 2000, containing approximately 1,000,000 documents in six languages instead of four – the new languages being Spanish and Dutch. It contains both newswires and national newspapers. The collection used for the multilingual task contained documents in five of those languages (English, French, German, Italian and Spanish) – one more than last year. There were two target collections for the bilingual track. Participants could query sets of either English or Dutch newspaper documents using their preferred topic language. Dutch was included this year not only to meet the demands of the considerable number of Dutch participants (the largest group) but also because it provided a challenge for those who wanted to test the adaptability of their systems to a new language.

The monolingual retrieval task was offered for five languages: Dutch, French, German, Italian and Spanish, instead of three as in CLEF2000. We exclude English from this track as it is already so well covered by the Text REtrieval Conference (TREC) series organised in the United States. The domain-specific task was very similar to that of last year. Again we used the GIRT database of social science documents, which has controlled vocabularies for English-German and German-Russian. The novelty this year was that Russian was included as a query language in addition to the English and German of previous years. The interactive track used data (documents and results) from the CLEF 2000 campaign.

Topics: The participating groups derived their queries in their preferred language from a set of topics created to simulate user information needs. Following the TREC² philosophy, each topic consists of three parts: a brief title statement; a one-sentence description; a more complex narrative specifying the relevance assessment criteria. 50 topics were developed on the basis of the contents of the multilingual collection and topic sets were produced in all six document languages. Additional topic sets were then created for Finnish, Swedish, Russian, Japanese, Chinese and Thai. Participants could thus choose to formulate their queries in any one of nine

¹ CLEF is conducted as an EU-US collaboration. The US partner is the National Institute for Standards and Technology (NIST), Gaithersburg, MD. The CLEF2000 and 2001 campaigns have been sponsored by the DELOS Network of Excellence for Digital Libraries (<http://www.ercim.org/delos>). From October 2001, CLEF will be run as an independent project of the European Commission (IST-2000-31002). For more information, see: <http://www.clef-campaign.org>.

European or three Asian languages. A condition in this year's CLEF was that, for each task attempted, a mandatory run using the title and description fields had to be submitted. The objective was to facilitate comparison between the results of different systems.

Relevance Judgments: Relevance assessment was distributed over six different sites and performed in all cases by native speakers. The results were then analysed and run statistics produced and distributed.

Participants

Participation in CLEF 2001 was up approximately 50% from last year, with more than 40 groups registering to participate in one or more of the main tasks. Many were participants from last year but there were also a good number of newcomers. In the end, 31 groups actually submitted results: 8 from N.America; 19 from Europe, and 4 from Asia – compared with 20 groups for CLEF2000. A total of 193 runs were received; runs were submitted for all tasks (multilingual, bilingual, monolingual and domain-specific) and for all topic languages. Twenty-one groups tried a cross-language task, while ten preferred to remain with the monolingual track. Only eight groups were brave enough to attempt the multilingual track (processing a document collection in five languages is certainly a challenging task) and of these just two were CLEF newcomers. An additional three groups tackled the experimental interactive task.

Working Notes and Workshop

The Working Notes provide a first description of the different experiments run by the participating groups. The Appendix gives a summary of the characteristics of all runs together with overview graphs for the different tasks and individual statistics for each run. Other papers in this volume include a report on the NTCIR Workshop series for Asian languages system evaluation and presentations on cross-language evaluation at TREC-9 and the NIST perspective on the implications of information retrieval system evaluation. 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 2001 Proceedings. These will be published by Springer in their Lecture Notes for Computer Science series.

The aim of the Workshop is to give all the groups that have participated in the CLEF 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. We very much hope that this event will prove an interesting, worthwhile and enjoyable experience to all those who participate.

Carol Peters, 1 September 2001

The Workshop Steering Committee

Martin Braschler, Eurospider, Switzerland
Julio Gonzalo Arroyo, UNED, Madrid, Spain
Donna Harman, National Institute for Standards and Technology, USA
Djoerd Hiemstra, University of Twente, The Netherlands
Noriko Kando, National Institute of Informatics, Japan
Michael Kluck, IZ Sozialwissenschaften, Bonn, Germany
Carol Peters, IEI-CNR, Pisa, Italy
Peter Schäuble, Eurospider, Switzerland
Ellen Voorhees, National Institute for Standards and Technology, USA
Christa Womser-Hacker, University of Hildesheim, Germany

² CLEF is a continuation and extension of the track for cross-language information retrieval, which was included in TREC from 1997-1999.

Acknowledgements

We have many people and organisations to thank for their help in the running of CLEF 2001.

First of all, we should like to express our gratitude to the ECDL 2001 Conference organisers for their assistance in the organisation of the CLEF Workshop.

With the single exception of the Thai experiment, the topic sets were prepared by independent groups, i.e. by groups not participating in the system evaluation tasks. The main topic sets (DE, EN, FR, IT, NL, SP) plus Russian were prepared by the project partners. Here, we should like to thank the following organisations that voluntarily engaged translators to provide topic sets in Chinese, Finnish, Japanese and Swedish, working on the basis of the set of source topics:

- Department of Information Studies (University of Tampere, Finland) which engaged the [UTA Language Centre](#) for the Finnish topics;
- SICS Human Computer Interaction and Language Engineering Laboratory for the Swedish topics.
- National Institute of Informatics (NII), Tokyo for the Japanese topics
- Natural Language Processing Lab, Department of Computer Science and Information Engineering, National Taiwan University for the Chinese topics

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.
- 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.
- Schweizerische Depeschagentur, Switzerland, for the French, German and Italian Swiss news agency data.

Without their help, this evaluation activity would be impossible.