

# Introduction

Carol Peters

IEI- CNR, Area di Ricerca, 56124 Pisa, Italy  
carol@iei.pi.cnr.it

The objective of the Cross-Language Evaluation Forum (CLEF) is to develop and maintain an infrastructure for the testing and evaluation of information retrieval systems operating on European languages, and to create test-suites of reusable data that can be employed by system developers for benchmarking purposes. The second CLEF evaluation campaign was held from January to September 2001 and ended with a workshop held in Darmstadt, Germany, 3-4 September, in which the participants in the campaign reported their experiments and discussed their results.

These Proceedings consist of the revised, extended versions of the preliminary papers presented by the participants at the Workshop. In many cases, the participating groups not only describe and analyse their first results but report additional experiments made subsequent to the workshop. The volume consists of two parts and an appendix. The first part provides an exhaustive overview of the CLEF 2001 experiments whereas the second describes the framework against which these experiments were held. Readers who have never participated in CLEF or in similar evaluation campaigns may well prefer to begin with the second part in order to acquire the necessary background information on the organisation of this type of campaign, before entering into the details of the different cross-language and monolingual retrieval experiments. The appendix presents the results of all the participating groups for each track and task, run by run.

## 1 CLEF 2001 Experiments

Part I of this volume contains papers from the individual participating groups and provides a complete record of the CLEF 2001 experiments. The first paper by Martin Braschler introduces the experiments by giving a description of the various tracks and tasks and a summary of the main results. The remainder of Part I has been divided into three sections, reflecting to some extent the organisation of the Workshop.

In the first section, we have grouped all those papers that describe cross-language system testing activities: both multilingual, bilingual and domain-specific experiments are included here. The name of the section, “Mainly Cross-language” is determined by the fact that many of the authors also mention strategies implemented in monolingual runs; however, the main focus is on the cross-language aspect of their work. Twenty-two groups submitted runs for this kind of cross-language task, twenty of these groups have contributed a paper to these Proceedings.

Nine groups preferred to remain with the monolingual track at CLEF 2001; most of these were newcomers to CLEF activities. We expect many of them to move onto a

cross-language task in the future. The work of seven of these groups is reported in the Monolingual section. CLEF 2001 provided the possibility to test monolingual systems on French, German, Dutch, Spanish and Italian collections. For most of these languages, CLEF offers the first opportunity for objective system evaluation under controlled conditions.

The final section of Part I is dedicated to the activities of another kind of cross-language task: interactive retrieval. The aim of the main cross-language and monolingual tracks was to measure system performance in terms of document rankings. However, this is not the only issue that interests the user. User satisfaction with an IR system is based on a number of factors, depending on the functionality of the particular system. An Interactive Track focusing on the interactive selection of documents that had been automatically translated from a language that the searcher would otherwise have been unable to read was experimented by a small group of three participants in CLEF 2001. The first paper in this section by Douglas Oard and Julio Gonzalo discusses the experimental design of this track, summarizes the results from the evaluation, and concludes with observations that can inform the design of subsequent evaluation activities for cross-language interactive systems. The other three papers describe the individual experiments conducted by the three groups.

Overall, the papers of Part I provide an important record of the state-of-the-art in the multilingual information access area, and of the main emerging research trends.

## 2 Evaluation Issues

Part II of the Proceedings consists of several papers which provide information concerning system evaluation activities not only for CLEF but also for the other two major international initiatives for system evaluation: the Text Retrieval Conferences (TREC) and the NACSIS Test Collection for Information Retrieval (NTCIR), both of which also organize tracks for cross-language system evaluation. Evaluation conferences such as TREC, NTCIR, and CLEF are all based on the evaluation methodology introduced in the Cranfield experiments. The first paper in this section by Ellen Voorhees reviews the fundamental assumptions and appropriate uses of the Cranfield paradigm, in particular as they apply in the context of today's evaluation conferences. The second paper by Noriko Kando introduces the NTCIR Workshop held in Japan for Asian languages, giving a brief history of the Workshop, and describing the tasks, test collections and evaluation methodology adopted.

The third paper in Part II is very much a CLEF-specific paper. Christa Womser-Hacker describes the topic generation process in a multilingual context. Topics are structured statements representing information needs from which the systems derive their queries and as such are an essential part of the test collection created for the campaign. Womser-Hacker discusses the various issues that have to be considered – linguistic and pragmatic – when constructing a standard set of topics in many different languages in a distributed framework.

The final paper in this section gives a complete overview of the organization of the CLEF 2001 evaluation campaign, describing the tasks and the test collection and

explaining the techniques and measures used for result calculation and analysis. Readers who have no experience of CLEF are probably well-advised to look at this paper before beginning to read the details of the different experiments.

### 3 The Future

CLEF 2002 is now well under way. The CLEF 2000 and 2001 campaigns were sponsored by the DELOS Network of Excellence for Digital libraries; from October 2001, CLEF is funded mainly by the European Commission under the IST programme (IST-2000-31002). CLEF 2002 participants are able to build on and profit from the documented results of the first two campaigns. In particular, they have access to the test-suites that have been constructed as a result of the first two CLEF campaigns. It is our intention, in the future, to make this valuable data accessible to a wider research community.

Previous to the launching of the 2002 campaign, we conducted a survey in order to acquire input for the design of the tasks to be offered. Two types of users were considered: cross-language technology developers and cross-language technology deployers. The main recommendations made can be summed up in the following list:

- Increase the size and the number of languages in the multilingual test collection (both with respect to documents and topics);
- Provide the possibility to test on different text types (e.g. structured data);
- Provide more task variety (question-answering, web-style queries, text categorization);
- Study ways to test retrieval with multimedia data;
- Provide standard resources to permit objective comparison of individual system components (e.g. groups using a common retrieval system can compare the effect of their individual translation mechanisms);
- Focus more on user satisfaction issues (e.g. query formulation, results presentation).

As far as possible, the findings of this survey have been integrated into the definition of the CLEF 2002 campaign. Points that could not be taken up immediately will be considered for the future. As a first step, the size of the newspaper/newsagency collections and the number of languages covered have been increased. Language coverage in CLEF depends on two factors: the demand from potential participants and the existence of sufficient resources to handle the requirements of new languages. Our goal is to be able to cover not only the most widely used European languages but also some representative samples of less common languages, including members from each major group: e.g. Germanic, Romance, Slavic, and Ugro-Finnic languages. CLEF 2002 is seeing the addition of Finnish and Swedish to the multilingual corpus; hopefully, 2003 will see the inclusion of a Russian collection. Other languages will be considered in future years.

With respect to the demand for different types of texts and evaluation tasks, CLEF 2002 has seen the addition of the Amaryllis corpus of French bibliographic documents to the multilingual collection of scientific documents. We now have a specific track dedicated to testing systems operating on different types of domain-specific collec-

tions. For the future, we are considering the possibility of setting up a track for text categorization in multiple languages.

In order to meet the demand regarding end-user related issues, the interactive track has been extended in 2002 and will be testing both user-assisted query translation and also document selection.

Finally, as a first move towards handling multimedia, we are examining the feasibility of organising a spoken CLIR track in which systems would have to process and match spoken queries in more than one language against a spoken document collection. An experiment in this direction is being held this year within the framework of the DELOS Network of Excellence for Digital Libraries. The results will be presented at the annual CLEF Workshop in September 2002.

In conclusion, the results of the survey make it very clear that CLIR search functionality is perceived as just one component in a far more complex system cycle which goes from query formulation to results assimilation. In future years, we hope to go further in the extension of CLEF evaluation tasks, moving gradually from a focus on cross-language text retrieval and the measuring of document rankings to the provision of a comprehensive set of tasks covering all major aspects of multilingual, multimedia system performance with particular attention to the needs of the end-user.

More information on the activities of CLEF can be found on our Web site: <http://www.clef-campaign.org/>.

**Acknowledgements.** Many people and organizations must be thanked for their help in the running of CLEF 2001. First of all I should like to thank the other members of the CLEF Coordinating Group for all their efforts aimed at making both the campaign and workshop a great success. I should also like to express my gratitude to the ECDL 2001 Conference organisers for their assistance in the organisation of the CLEF Workshop.

It is not easy to set up an infrastructure that handles a large number of languages. The main topic sets (DE, EN, FR, IT, NL, SP) plus Russian were prepared by the project partners. Here, I should like to thank the following organisations that voluntarily engaged translators to provide topic sets in Chinese, Finnish, Japanese, Swedish and Thai, 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;
- Human Computer Interaction and Language Engineering Laboratory, Swedish Institute of Computer Science (SICS), 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
- Kasetsart University, Thailand, for the Thai topics

I 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.

Last, but certainly not least, I should like to express my gratitude to Francesca Borri, IEI-CNR, for all her hard work in the editing and preparation of the final version of the texts. Much of the success of both the CLEF campaigns and the CLEF Workshops is a result of Francesca's valuable collaboration.