# Cross-Language Evaluation Forum

# CLEF 2002: PRELIMINARY CALL FOR PARTICIPATION

## Objectives

CLEF promotes research in cross-language system development for European languages by providing an infrastructure for:

- cross-language and monolingual system evaluation, testing and tuning
- comparison of different approaches and discussion of results
- creation of resuable evaluation test-suites for system benchmarking

# CLEF Agenda for 2002

## Multilingual Information Retrieval

The main task in CLEF requires searching a multilingual document collection for relevant documents. Using a selected query language, the goal is to retrieve documents for all languages in the collection, rather than just a given pair, listing the results in a merged, ranked list.
The CLEF 2002 document collection for this task contains English, German, French, Italian and Spanish documents. A common set of topics, from which queries are extracted, will be prepared in a large number of languages. These will include Dutch, English, Finnish, French, German, Italian, Spanish, Swedish, Japanese and Chinese. It will also be possible to submit runs in which the document collection is queried in other languages. In this case, participants will be responsible for the translation of the query into their selected language. The results for such runs will be given separately.

CLEF 2002 also offers a series of additional tracks designed to test different aspects of information retrieval system development.

### Bilingual Information Retrieval

In the cross-language task, any topic language can be used to query target document collections in Dutch, French, German, Italian or Spanish. First-time CLEF participants only can choose to query the English document collection using a European topic language.

### Monolingual (non-English) Information Retrieval

Until recently, most IR system evaluation focused on English. CLEF provides the opportunity for monolingual system testing and tuning, and for building test suites in other European languages. CLEF 2002 offers tracks for Dutch, French, German, Italian and Spanish; we may also be able to include a Finnish collection.

---

The CLEF 2002 evaluation campaign is sponsored by the IST Programme of the European Commission (project no. IST-2000-31002) in collaboration with the US National Institute of Standards and Technology (NIST).

### Domain-specific Mono- and Cross-Language Information Retrieval

Depending on demand, CLEF 2002 will again offer a domain-specific task. This task is based on the GIRT collection which contains nearly 80,000 German social science documents in a structured database. A German/English/Russian thesaurus and English translations of the document titles are available. The rationale for this task is to study CLIR in a vertical domain (i.e. social science).

### Interactive Cross-language Information Retrieval

An interactive track that focused on the document selection problem was experimented with success in CLEF 2001. The design of this year's interactive track will be determined by the coordinators on the basis of input from interested participants.

## Participation

Those wishing to take part in CLEF 2002 are requested to send an e-mail to Carol Peters (carol@iei.pi.cnr.it), indicating in which task(s) they intend to participate. Participants will be requested to sign an agreement restricting the use of the data and regulating publication and dissemination of results.

## Important Dates

| | |
|---|---|
| Data Release | 1 February 2002 |
| Topic Release | 1 April 2002 |
| Receipt of results from participants | 15 June 2002 |
| Release of relevance assessments and individual results | 1 August 2002 |
| Submission of paper for Working Notes | 1 September 2002 |
| Workshop | 19-20 September 2002 |

## Workshop

A two-day Workshop will be held on 19-20 September 2002 in Rome, Italy, immediately following the Sixth European Conference on Digital Libraries (ECDL 2002).
The aim of the Workshop will be to present and discuss the results of the CLEF activity and allow researchers and developers to compare performance between systems using different cross-language strategies.

## CLEF Consortium

CNR-IEI, Pisa, Italy; ELRA/ELDA, Paris, France; Eurospider, Zurich, Switzerland; UNED, Madrid, Spain; NIST, Gaithersburg, MD, USA; IZ Sozialwissenschaften, Bonn, Germany.

## Associated Members

University of Hildesheim, Germany; University of Twente, The Netherlands.

## Guidelines and Contact Information

For further information see: http://www.clef-campaign.org

or contact:

Carol Peters - IEI-CNR
Area della Ricerca CNR, 56124 Pisa, Italy
Tel: +39 050 3152897 - Fax: +39 050 3152810
E-mail: carol@iei.pi.cnr.it