

# Comparative Evaluation of Cross-language Information Retrieval Systems

Carol Peters

ISTI-CNR, Area di Ricerca CNR, Via Moruzzi, 1  
56124 Pisa, Italy  
carol.peters@isti.cnr.it

**Abstract:** With the increasing importance of the "Global Information Society" and as the world's depositories of online collections proliferate, there is a growing need for systems that enable access to information of interest wherever and however it is stored, regardless of form or language. In recognition of this, five years ago, the DELOS Network for Digital Libraries launched the Cross-Language Evaluation Forum (CLEF), with the objective of promoting multilingual information access by providing the research community with an infrastructure for testing and evaluating systems operating in multilingual contexts and a common platform for the comparison of methodologies and results. In this paper, we outline the various activities initiated by CLEF over the years in order to meet the emerging needs of the application communities, and trace the impact of these activities on advances in multilingual system development.

## 1 Introduction

In recognition of the fact that "multilinguality" is of particular relevance for European digital collections, although a frequently neglected issue in European digital library (DL) projects, in 2000 DELOS decided to sponsor the work of the Cross-Language Evaluation Forum (CLEF) as part of its DL evaluation activity. CLEF was given the mandate to promote cross-language information retrieval (CLIR) system development by providing the research community with an infrastructure for testing and evaluating systems operating in both monolingual and cross-language contexts and by encouraging take-up of the R&D results by the application communities. In order to achieve these objectives, CLEF organises annual system evaluation campaigns, which culminate each year in a workshop where the participants present and discuss their experiments. In this paper, we will describe the main achievements of CLEF so far and will discuss the efforts that are being made to ensure that CLEF continues to meet the emerging needs of system developers and application communities.

## 2 Consolidating Multilingual Document Retrieval

Over the last decade, with the increasing globalization of the information society, the interest in the potential of multilingual<sup>1</sup> information access functionality has grown considerably. However, when the first CLIR system evaluation activity began in 1997 at TREC<sup>2</sup>, very little IR system testing work had been done for languages other than English and almost all existing cross-language systems were designed to handle no more than two languages: searching from query language to target language. Since its beginnings, CLEF<sup>3</sup> has worked hard to change this situation and to promote the development of systems capable of searching over multiple languages, with the focus on European languages. For this reason, each year the CLEF ad-hoc track has proposed a set of core evaluation tasks designed to test monolingual, bilingual (queries in one language against a target collection in another) and multilingual (queries in one language against target collections in many languages) text retrieval systems. The aim has been to encourage groups to work their way up gradually from mono- to multilingual retrieval, providing them with facilities to test and compare search and access techniques over languages and pushing them to investigate the issues involved in simultaneously processing a number of languages with different characteristics. Over the years, the language combinations provided have increased and the tasks offered have grown in complexity. The CLEF 2003 multilingual track included a task which entailed searching a collection in eight languages. In 2000, the main CLEF multilingual corpus consisted of approximately 360,000 newspaper and news agency documents for four languages; by 2004 it included nearly two million documents and ten languages<sup>4</sup>.

The evaluation environment for this track adopted an automatic scoring method, based on the well-known Cranfield methodology [2]. The test collection consists of a set of “topics” describing information needs and a set of documents to be searched to find those documents that satisfy the information needs. Evaluation is then done for each ranking of documents with respect to a topic by the usual computation of recall and precision. The distinguishing feature of CLEF is that it applies this evaluation paradigm in a multilingual setting. This means that the criteria normally adopted to

---

<sup>1</sup> We use “multilingual” to denote both cross-language and multiple language information processing activities.

<sup>2</sup> TREC – the Text REtrieval Conference series, co-sponsored by the National Institute of Standards and Technology (NIST) and the US Department of Defense, was started in 1992 to support research within the information retrieval community by providing the infrastructure necessary for large-scale evaluation of text retrieval methodologies. See <http://trec.nist.gov/>

<sup>3</sup> CLEF was launched when it was decided to move the coordination of the CLIR track at TREC to Europe. CLEF 2000 and 2001 were sponsored by the 5FP DELOS Network of Excellence for Digital Libraries; CLEF 2002 and 2003 were funded as an independent project by the European Commission under the IST programme. CLEF 2004 was again organized as an activity of the DELOS Network of Excellence, under the Sixth Framework Programme of the European Commission.

<sup>4</sup> The CLEF comparable corpus currently contains news documents for the same time period (1994-95) in ten languages: Dutch, English, Finnish, French, German, Italian, Portuguese, Russian, Spanish, and Swedish; we are now working on adding some central European languages: Bulgarian, Czech and Hungarian.

create a test collection, consisting of suitable documents, sample queries and relevance assessments, have been adapted to satisfy the particular requirements of the multilingual context. All language dependent tasks such as topic creation and relevance judgments are performed in a distributed setting by native speakers. Rules are established and a tight central coordination is maintained in order to ensure consistency and coherency of topic and relevance judgment sets over the different collections, languages and tracks.

The results of this track in terms of participation and of the different approaches and techniques tested have been impressive. Regular CLEF participants, i.e. groups that have participated several years running, have shown improvements in performance and flexibility in advancing to more complex tasks. Much work has been done on fine-tuning for individual languages, while other efforts have concentrated on developing language-independent strategies. The issues involved in cross-language text retrieval have been investigated in depth and are now well understood. Recent work at CLEF has shown that what can be called a “blueprint” for a successful multilingual system is now emerging, although the merging of results over collections can still be considered as a largely unsolved problem. At the same time, a clear improvement in system performance has been demonstrated for bilingual information retrieval. At the CLIR track in TREC in 1977, bilingual retrieval averaged at about 50% of monolingual retrieval on the same collection; the figures for both CLEF 2003 and 2004 show the best bilingual system performance at more than 80% of monolingual retrieval. A detailed discussion of these results can be found in [1].

### **3 Towards Multilingual Information Access**

As described in the previous section, in the beginning CLEF focused mainly on testing performance of off-line text retrieval systems, where good system performance is equated with good retrieval effectiveness (in terms of returning lists of documents). However, it became obvious that this was only one part of the CLIR problem. A multilingual system evaluation activity that meets the needs of the application communities must also provide facilities to investigate many other aspects including:

- not just retrieval on news documents but also on other genres which may have their own specific access and retrieval problems;
- not just system performance but also wider usability issues that affect the users’ ability to recognize relevant information and refine search results even when documents are written in an unfamiliar language;
- not just document retrieval, but also targeted information location and extraction;
- not just text but also mixed-media data, e.g. collections containing images or spoken documents.

With these requirements in mind, CLEF has gradually extended its range of interests and increased the number of evaluation tracks offered. Over the years, the focus in CLEF has shifted from straight document retrieval and has diversified to include different kinds of text retrieval across languages (from documents to exact answers) and retrieval on different kinds of media (not just text but image and speech as well).

By designing and offering new and challenging tasks, our intention has been to stimulate advances in system development. In this section, we describe what has been done so far and our plans for the next years. As will be seen, in order to be able to run so many diverse evaluation tracks (six in 2004, eight tracks are planned for 2005), the expertise and the efforts of a large number of research groups have been mobilized.

### 3.1 Cross-Language System Evaluation for Different Genres

The main CLEF collection is the multilingual comparable corpus of news documents described above. However, news media have characteristics which may not hold true for other genres: wide use of proper nouns (names and places), association of date stamps, particular style of writing and a rapid evolution of general-purpose vocabulary. Certain features may facilitate access and retrieval, others may hinder it. [3]. For this reason, CLEF has also included a mono- and cross-language domain-specific retrieval track each year, mainly based on the GIRT corpus of structured social science data which has an associated social science thesaurus in German-English and German-Russian<sup>5</sup>. Since the first CLEF campaign in 2000 the GIRT corpus has been enlarged several times, and is now presented as a pseudo-parallel corpus<sup>6</sup>, with about 150,000 identical documents in German and English. In this way, we have provided the opportunity for developers to test their CLIR systems on domain-specific data, with domain related vocabulary. In 2005 we intend to enlarge the existing social science collection with Russian sociology data. However, a severe problem with offering domain-specific collections in a CLEF-type evaluation campaign is that you need domain-specific experts willing to perform the relevance assessment task. Cross-lingual patent retrieval is already offered as an evaluation task by the Asian initiative, NTCIR<sup>7</sup>, and CLEF is now considering offering a similar task; contacts are now under way with the European patent office. Although there is much industrial interest in such a tasks but its realization depends heavily on the availability and willingness of experts to do the relevance assessments.

Another document type will be included in CLEF 2005 with the introduction of a web task. The World Wide Web is, of course, a natural setting for cross-language retrieval activities and many issues for which people consult the Web are essentially issues that cross over language boundaries, e.g. news, education, travel, leisure. As web pages normally contain information in mixed media, searchers not familiar with the language of a page are often able to extract useful information, such as names, maps, timetables, etc. However, current web search engines do not offer truly multilingual search functionality. We believe CLEF could provide a real service to the research community by providing a multilingual Web test collection and testing

---

<sup>5</sup> A German-English Social Science Thesaurus and the German-Russian wordlist are provided and made available in machine readable formats by IZ-Bonn, Germany. UC Berkeley made an XML version available.

<sup>6</sup> The GIRT4 corpus is called pseudo-parallel because the original documents are in German and the English part consists of translations of these German documents into English; the English part is actually considerably smaller than the German.

<sup>7</sup> NTCIR (NII-NACSIS Test Collection for IR Systems) organized by the Institute of Informatics, Tokyo: <http://research.nii.ac.jp/ntcir/>. There is an ongoing collaboration and sharing of ideas between NTCIR and CLEF.

systems with respect to different issues, eg results presentation, user interaction. Unfortunately, the building of a Web corpus for multilingual access is not a trivial task; the basic snapshot data of the Web and the evaluation metrics will have to be defined very carefully [4, 5]. A multilingual Web corpus should include carefully calculated representative samples for less commonly used languages plus random samples for the most dominant ones. Questions of intellectual property rights must also be considered.

The first step in 2005 will be to start work on building such a collection. A group from the University of Amsterdam is responsible for developing a pilot WebCLEF track. They intend to build on insights gained while building the W10G and the .gov collections at TREC [6]. Their aim for the first year is to have a collection of between one and two million pages with a minimum of 50,000 documents per language for at least ten languages; html, pdf and txt type documents will be included. Queries will be generated in many languages and will be created in two ways: partly artificially by participants and partly by extraction from log files.

There will also be a pilot experiment in CLEF 2005 in cross-language geographic information retrieval (GeoCLEF). This track will provide a framework in which to evaluate GIR systems for search tasks involving both spatial and multilingual aspects and is the first evaluation exercise of this type to be attempted.

### **3.2 Considering the Users' Perspective**

Of course, the user is at the centre of any information retrieval system and the system functionality must be designed with the user requirements in the forefront. This is especially true when users are searching for relevant information contained in documents written in languages that are unfamiliar to them. In this case, they may need assistance in recognizing which of a ranked list of documents are relevant for their purposes. They also may need help in refining their original query taking these results into account, and they need to understand how information contained in documents in an unfamiliar language can be exploited. These problems are crucial questions for the users of a CLIR system. The default assumptions for a cross-language search engine are that a) commercial Machine Translation (MT) systems can be used to translate documents into the native language of the user, and b) document selection and query refinement can be done using such translations. In order to challenge such (untested) assumptions, the University of Maryland, USA, and UNED, Spain, decided to organize an interactive track – *iCLEF* – within CLEF. The goal was comparative studies of user interaction issues in CLIR.

The first pilot interactive track at CLEF, known as *iCLEF*, was held in 2001 [7], and focused on document selection questions, i.e. approaches to facilitate fast and accurate relevance judgments for documents that the user could not read without assistance. The *iCLEF* experimental design combined insights from the interactive TREC track with the distributed multilingual assessment process of CLEF. The experience led to non-trivial empirical conclusions; for instance, full MT performed better than word-by-word translation, but cross-language pseudo-summaries (allowing faster relevance judgments without loss of precision) outperformed full MT. The success of this experiment led to the *iCLEF* track being included as a regular event in CLEF.

*i*CLEF tracks in CLEF 2002 and 2003 studied support mechanisms for interactive query formulation and refinement [8, 9], and included experiments such as user-assisted term translation (via inverse dictionaries, translation definitions, etc), which was shown to outperform automatic query translation, and user-assisted query reformulation by selection of relevant noun-phrases, which was shown to outperform user-assisted term translation.

In the 2004 track, five participating teams used a common evaluation design to assess the ability of interactive systems of their own design to support the task of finding specific answers to narrowly focused questions in collections of documents in a language different from that of the language in which the questions were expressed. This task was the interactive counterpoint to the fully automatic cross-language question-answering task in CLEF 2003 and 2004 [10]. Overall, *i*CLEF experiences so far have involved more than one thousand interactive cross-language searches in several languages (English, Spanish, Finnish, German and Swedish), constituting the largest set of empirical data about multilingual information access from the user perspective known to us.

The connection with the question-answering (QA) track (see next section) will be continued in 2005 where *i*CLEF will study the problem of cross-language question answering from a user-inclusive perspective. The challenge is twofold: from the point of view of QA as a machine task, interaction with the user may help a QA engine to retrieve better answers; from the point of view of QA as a user task, a search assistant may help the user to locate the answer more easily, more quickly and more accurately. The participating groups will share a common experiment design in which users explore different aspects of the above research challenges. *i*CLEF also intends to organize an interactive image retrieval task in coordination with the ImageCLEF track (see section 3.4 below).

### **3.3 Multilingual Information Extraction**

Multilinguality has been identified as one of the major challenges for QA systems [11]. However, little is yet known about cross-language QA, i.e. when the question is not written in the same language as the documents. It is certainly a harder problem than CLIR: while, for document retrieval purposes, finding appropriate candidate translations for the query words can lead to good performance, questions demand a careful linguistic analysis (in the source language) which is not trivially translated into the target language; and a (possibly noisy) machine-translated question may produce errors in the linguistic analysis.

In order to study these issues, in 2003, three research groups from Italy, Spain and the Netherlands (ITC-irst, UNED and ILLC) organized a pilot Cross-Language QA track under the auspices of CLEF. This track, known as QA@CLEF, evaluated monolingual QA systems in Dutch, Italian and Spanish, and cross-language QA systems searching an English target collection with Spanish, Italian, Dutch, French or German questions [12]. It was a pilot experiment, just 200 simple fact-based questions were provided and participants were allowed to provide up to three responses per question, either exact or 50-byte answer strings. Although only limited comparisons between systems could be made (there was only one participating system

per monolingual task) and just 6 groups participated in the bilingual experiment, the experience was considered a success, generating the first cross-language QA test suite and attracting a lot of interest from CLEF participants.

In 2004, the QA track was one of the most popular tracks at CLEF with 18 groups submitting mono- and/or cross-language runs using any of nine possible source languages against target collections in seven languages. New types of questions (how-questions and definition type questions) were given as input to the participating systems, while just one exact answer per question was allowed as output. A disadvantage of the wide range of language combinations means that for the cross-language tasks there were no real comparable results. However, much is being learnt about language-dependent factors for monolingual QA systems and the issues to be considered when working on cross-language QA. It is interesting to note that the average performance achieved by the monolingual systems in 2004 compared very favourably with the results generally obtained at TREC for English QA systems [13].

Following the positive outcome of the 2003 and 2004 QA@CLEF evaluation campaigns, in CLEF 2005, monolingual (non English) and cross-language QA systems will again be tested. All the combinations between nine or more source languages and eight target languages (Bulgarian, Dutch, English, French, German, Italian, Portuguese and Spanish) will be explored. The track will include a main task, where factoid and definition questions are given as input, and pilot tasks, that will explore other facets of QA, proposing both context oriented and more difficult types of questions. In addition, as reported above, experiments will be conducted in collaboration with the *i*CLEF track.

### **3.4 Cross-Language Retrieval in Mixed Media Collections**

Of course, text is not the only media in which information is stored. Today's digital collections frequently tend to be in multi-media, text being accompanied or replaced by image, speech, or video. In recent years, CLEF has begun to examine the issues involved in multilingual information access in collections of mixed media.

**Cross-Language Spoken Document Retrieval.** CLEF began to pay attention to the issues involved in cross-language spoken document retrieval in 2002 when a pilot experiment was organized within the DELOS Network of Excellence by two groups (ITC-irst and University of Exeter) and the results were reported at the CLEF workshop that year [14]. The experiment was continued on a slightly larger scale as a regular track in CLEF 2003 with four participating groups. The track ran with very limited resources, using data from the TREC 8 and 9 English monolingual SDR tracks, made available by NIST. The results are closer to a benchmark rather than a real evaluation. The TREC collections were extended to a CL-SDR task by manually translating the short topics into five European languages: Dutch, French, German, Italian and Spanish. In 2003 the track aimed at evaluating CLIR systems on noisy automatic transcripts of spoken documents with known story boundaries. The results of the experiments showed that, as expected, bilingual performance was lower for all participants than the comparative English monolingual run. However, the degree of degraded performance was shown to depend on the translation resources used. It was

also shown that it can be effective to use different indexing units for monolingual and bilingual retrieval on the data set [15]. The track was repeated in CLEF 2004 but with only two groups finally submitting runs and no conclusive results.

It was felt that one of the reasons for the lack of enthusiasm for the track in 2004 was the fact that the resources were so limited and the test collection not particularly appropriate for tasks focused on multilinguality. It is expected that there will be much more interest next year as a very new collection with immense potential for research is being introduced in the CLEF 2005 speech track. Both mono- and cross-language retrieval will be assessed on the Malach collection of spontaneous conversational speech from the archives of the Shoah foundation. The collection that will be used in 2005 is in English and will consist of 625 hours in 10,000 segments in ASR form plus ~5 keywords and a 3 sentence summary. A thesaurus, an in-domain expansion collection, and a word lattice should also be available. 25 topics in six languages (Czech, English, French, German, Russian and Spanish) will be prepared. Training topics will also be available. In future years, parts of the Malach collection in Russian and in Czech will be made available. This track is being coordinated by Dublin City University and the University of Maryland.

**Cross-Language Retrieval in Image Collections.** A cross-language image retrieval task, known as ImageCLEF, was first introduced into CLEF in 2003 [16]. Research in cross-language image retrieval is not just of academic interest, there is also a particularly strong commercial potential. Retrieval from an image collection is quite different than from a text collection. For example, the way in which a query is formulated, the methods used for retrieval (e.g. based on low-level features derived from an image, or based on associated textual information such as a caption), the types of query, how relevance is assessed, the involvement of the user during the search process, and fundamental cognitive differences between the interpretation of visual versus textual media are all issues that need careful consideration. Within CLEF, the problem is further complicated by user queries being expressed in a language different to that of the document collection. This requires crossing the language barrier by translating the collection, the queries, or both into the same language. ImageCLEF aims to provide the necessary collection(s) and framework in which to analyze the link between image and text and promote the discovery of alternate methods for cross-language image retrieval. The goal is thus to provide a test-bed that can be used to evaluate different retrieval methods and to analyze user behaviour during the search process, e.g. query formulation in both cross-language and visual environments, iterative searching and query reformulation.

The first ImageCLEF was set up as a pilot experiment: just four groups participated in an ad-hoc retrieval task. Participants were free to use either content-based or text-based retrieval methods, relevance feedback and any translation method. The search requests were first produced in English and then prepared in Dutch, French, German, Italian and Spanish and consisted of both visual examples and text descriptions of the user need. The image collection of approximately 30,000 historical photographs of Scotland complete with short captions was made available by St Andrews University Library. Although all groups chose to use the information derived from captions only, there was much interest in the potential of this exercise at the CLEF 2003 workshop and it was thus decided to expand it in CLEF 2004, adding another collection: the Casimage medical images of the University Hospitals Geneva.

Eighteen groups participated in ImageCLEF 2004. Three tasks were offered: ad-hoc type retrieval on the St Andrews and the Casimage collections and a user-centred evaluation task. Topics were offered in twelve languages and participating groups could use text-based and/or content-based retrieval methods. In general, it was found that for the two retrieval tasks across very different domains a combination of visual and textual features provided a retrieval effectiveness that was higher than retrieval based on text or visual features alone. It was also shown that, like cross-language retrieval on text collections, query expansion and relevance feedback improved performance. The high participation showed that there was a need for this type of evaluation event and the results proved that further research into the combination of visual and textual retrieval techniques should lead to optimal results on image collections containing some kind of caption or annotation text in the future [17].

The success of 2004 has led to the ImageCLEF track being further extended for CLEF 2005. Four tasks are offered: bilingual ad-hoc retrieval (collection in English, queries in several languages); interactive search; medical image retrieval (collection in English and French, images as queries); an automatic annotation task for medical images (a fully categorized collection with categories available in English and German). The tasks offer different and challenging retrieval problems for cross-language image retrieval. Three test collections will be made available: the St Andrews University historical photographic collection, the Casimage medical collection, enlarged for this year's CLEF by other teaching files, and the IRMA medical image collection for automatic image annotation created jointly by University Hospitals Geneva and the University of Aachen. The track will be coordinated jointly by University of Sheffield, University Hospitals Geneva and Oregon Health and Science University.

## 4 Conclusions

In this paper, we have shown how the activity of the Cross-Language Evaluation Forum has expanded and widened its range of activities over the years. The main objective of the CLEF initiative has been to stimulate and advance multilingual information retrieval system development. Our challenge has been to encourage the building of systems that will allow real people to find the information they need in languages that they have not mastered and in collections of diverse types and diverse media, and then to measure how well representative users are able to use these systems. For this reason, we have encouraged R&D work on multilingual system development in many different directions and have concentrated on developing appropriate evaluation infrastructures and methodologies in order to assess the results obtained in an objective setting.

We feel that our main results can be summarized in the following points:

- the building of a strong multidisciplinary research community, focused on improving and expanding multilingual system development;
- documented improvement in cross-language text retrieval system performance;
- the creation of a large set of empirical data about multilingual information access from the user perspective;

- the promotion of in-depth studies in previously uninvestigated areas, such as cross-language speech and image retrieval and multilingual question answering;
- the development of appropriate evaluation methodologies to assess new multilingual information access systems and/or new system components;
- the creation of a set of immensely valuable, reusable multilingual test collections, complete with documentation on how to use them.

Further information on the CLEF activities can be found at <http://clef-campaign.org/>

## References

1. Braschler, M., Peters, C.: Cross-Language Evaluation Forum: Objectives, Results, Achievements. *Information Retrieval*, 7, Nos 1-2 (2004) 7-31.
2. Cleverdon, C.: The Cranfield Tests on Index Language Devices. In: Spärck-Jones, K., Willett, P. (eds.): *Readings in Information Retrieval*, Morgan Kaufmann (1997) 47-59.
3. Gey, F., Kando, N., Peters, C.: Cross-Language Information Retrieval: the Way Ahead. *Information Processing and Management*. 4 (3). Forthcoming.
4. Gurrin, C., Smeaton, A. F. : Improving Evaluation of Web Search Systems. In Sebastiani, F. (ed.) *Advances in Information Retrieval*. Springer LNCS 2633 (2003) 25-40
5. Kando, N.: Evaluation of Information Access Technologies at the NTCIR Workshop. In *CLEF 2003 Proceedings*, Springer LNCS 3237 (2004), in print.
6. Bailey, P., Craswell, N., Hawking, D.: Engineering a multi-purpose test collection for web retrieval experiments. *Information Processing and Management*, 39(6). (2003) 853-871.
7. Oard, D., Gonzalo, J.: The CLEF 2001 Interactive Track. In *Proceedings of CLEF 2001*, Springer LNCS 2406 (2002) 308-319.
8. Gonzalo, J., Oard, D.: The CLEF 2002 Interactive Track. In *CLEF 2002 Proceedings*, Springer LNCS 2785 (2003) 372-382.
9. Oard, D., Gonzalo, J.: The CLEF 2003 Interactive Track. In *CLEF 2003 Proceedings*, Springer LNCS 3237 (2004). In print.
10. Gonzalo, J., Oard, D.: In *CLEF 2004 Working Notes*. See [http://clef.isti.cnr.it/2004/working\\_notes/CLEF2004WN-Contents.html](http://clef.isti.cnr.it/2004/working_notes/CLEF2004WN-Contents.html)
11. Maybury, M.: *Toward a Question Answering Roadmap*, MITRE Tech. Report (2002).
12. Magnini, B., Romagnoli, S., Vallin, A., Herrera, J., Peñas, A., Peinado, V., Verdejo, F., de Rijke, M.: The Multiple Language Question Answering Track at CLEF 2003. In *CLEF 2003 Proceedings*, Springer LNCS 3237 (2004). In print.
13. Magnini, B., Vallin, A., Ayache, C., Erbach, G., Peñas, A. de Rijke, M., Rocha, P., Simov, K., Sutcliffe, R.: Overview of the CLEF 2004 Multilingual Question Answering Track. In *CLEF 2004 Working Notes*. See [http://clef.isti.cnr.it/2004/working\\_notes/CLEF2004WN-Contents.html](http://clef.isti.cnr.it/2004/working_notes/CLEF2004WN-Contents.html)
14. Jones, G.J.F., Federico, M.: CLEF 2002 Cross-Language Spoken Document Retrieval Pilot Track Report. In *Proceedings of CLEF 2002*, Springer LNCS 2785 (2003) 446-457.
15. Federico, M., Jones, G.J.F.: The CLEF 2003 Cross-Language Spoken Document Retrieval Track. In *CLEF 2003 Proceedings*, Springer LNCS 3237 (2004). In print.
16. Clough, P.D., Sanderson, M.: The CLEF 2003 Cross Language Image Retrieval Track. In *CLEF 2003 Proceedings*, Springer LNCS 3237 (2004). In print.
17. Clough, P.D., Sanderson, M., Müller, H.: The CLEF Cross Language Image Retrieval Track (ImageCLEF) 2004. In *CLEF 2004 Working Notes*. See [http://clef.isti.cnr.it/2004/working\\_notes/CLEF2004WN-Contents.html](http://clef.isti.cnr.it/2004/working_notes/CLEF2004WN-Contents.html)