CLEF - IST-2000-31002
Cross-Language Evaluation Forum

User Needs: Deliverable 1.1.1

Deliverable Type:Report
Number: D1.1.1
Nature: Final

Task WP1

Name of responsible: Julio Gonzalo (UNED)

Authors: Julio Gonzalo, Felisa Verdejo, Anselmo Peñas (UNED)
Carol Peters (IEI-CNR)

Contributors: Khalid Choukri (ELDA/ELRA); Michael Kluck (IZ-Bonn)

Contact info: julio@lsi.uned.es

Abstract

The results of a user needs survey are given. The survey was conducted in order acquire input for the design of the tasks for the evaluation campaigns planned by the Cross-Language Evaluation Forum (CLEF). Two types of users were considered: cross-language technology developers and cross-language technology deployers. Preliminary results have been used in the definition of the CLEF 2002 evaluation campaign; the final results will provide a basis when deciding on the tasks to be offered in CLEF 2003.

Keyword List:

cross-language information retrieval, system evaluation, system developers, user needs

Executive Summary

The results of a user needs survey are given. The survey was conducted in order to acquire input for the design of the tasks for the evaluation campaigns planned by the Cross-Language Evaluation Forum (CLEF). Two types of users have been considered in the survey: cross-language technology developers and cross-language technology deployers, and two separate questionnaires were designed.

The first group was mainly identified in cross-language or multilingual system developers who had previously participated in CLEF-style evaluation campaigns or groups who had indicated interest in the CLEF evaluation activities. This group was already well aware of the objectives and potential of system evaluation campaigns and thus provided considerable useful input. The main recommendations made with reference to future CLEF campaigns can be summed up in the following list:

- Include more languages in the multilingual document collection.
- Increase the size of the document collection.
- Include more languages in the topic[1] set.
- Increase the number of topics offered.
- Improve the pool quality.
- Provide the possibility to test on different text types (e.g. structured data)
- Provide more task variety
- Provide standard resources to permit objective comparison of individual system components.

A more detailed analysis of these suggestions and others is give in Section 2 of the deliverable. Preliminary results from this group have been used in the definition of the CLEF 2002 evaluation campaign; the final results will provide a basis when deciding on the tasks to be offered in CLEF 2003.

It was more difficult to identify successfully the second user type. The project technical annex referred to end-users. However, this is a much more difficult group to define. Cross-Language systems are still very much in the experimental stage. Thus, the typical search engine end-user really does not have a clear idea of the type of functionality that such systems will offer when they arrive on the market. It was thus decided to survey a group of cross-language technology deployers, i.e. groups that provide multilingual services or content of some type, in order to understand their main needs with respect to the provision of cross-language functionality to their user communities. In this way, it was hoped to acquire additional information with respect to user requirements from multilingual system evaluation activities.

The results of this second survey are reported in Section 3 of the deliverable; they are far harder to classify and reflect the large distance still existing between the development community and the application communities in the cross-language/multilingual system development domain. In fact, our main conclusion from the (not very satisfactory) results of this second questionnaire is that there is a great need for dissemination with regard to the state-of-the-art and the future potential of cross-language information retrieval systems for both information providers and information seekers. Hopefully, the activities of CLEF will also provide a strong contribution in this direction.

[1] In this deliverable we use the terminology adopted in CLEF. By topics we denote structured statements of information needs from which queries are extracted by the participating systems.

Document Evolution

Version	Date	Status	Notes
Preliminary Report	20 December 01	Intermediate – consigned to Commission	Internal; Prepared to provide input for definition of tasks for 2002 evaluation campaign
1.1	15 February 02	Pre-final	Internal; circulated to consortium partners
1.2	20 February 02	Final	Ready

Table of Contents

1. Introduction

The objective of this user needs survey has been to acquire input for the design of the tasks for the evaluation campaigns planned by the Cross-Language Evaluation Forum (CLEF). Two types of users were considered: cross-language technology developers and cross-language technology deployers. Preliminary results have been used in the definition of the CLEF 2002 evaluation campaign; the final results will provide a basis when deciding on the tasks to be offered in CLEF 2003.

1.1 Target Groups

The Technical Annex states that the user survey should be conducted on two user types: identified as *system developers* and *end-users* (interviewing approx. 20 representatives of each target group). The first step was thus to identify clearly these two groups and to design appropriate questionnaires.

System developers: The CLEF activity has not started from scratch but is building on a previous experience conducted within the DELOS Network of Excellence for Digital Libraries[2]: the CLEF 2000 and 2001 evaluation campaigns. The CLEF consortium had thus already establish strong relationships with a highly representative system development community, operating in a multilingual and/or cross-language context, either as participants in previous CLEF campaigns, or as groups that had requested information on CLEF activities. It was thus decided to focus on this community for the first questionnaire. It was felt that these groups would have strong opinions as to the design of the next CLEF campaigns and could provide immediately useful input.

End-Users: This was a much more difficult group to define. Cross-language systems are still very much in the experimental stage. Thus, the typical search engine end-user really does not have a clear idea of the type of functionality that such systems can offer when they arrive on the market. The uninformed user tends to confuse a cross-language search mechanism with an online machine translation service (such as, for example, that offered by Altavista or Google) and may not have very high expectations. It was thus decided to survey a group of cross-language technology deployers, i.e. groups that provide multilingual services or content of some type, in order to identify their main needs with respect to the provision of cross-language functionality to their user community. In this way, it was hoped to acquire additional information with respect to user requirements from multilingual system evaluation activities.

1.2 Survey Methodology

It is a well-known fact that it is very difficult to obtain a good response to surveys conducted entirely remotely. It was thus decided to focus on a previously selected sample which could be expected to be motivated to reply as they were (either actually or potentially) strongly interested in the CLEF activity. The questionnaires were designed to be simple to answer, and where necessary were followed up by a personal contact (generally by telephone, or via specific e-mails, few face-to-face interviews were conducted). The questionnaires used are given in the Appendix and the results obtained are discussed in following sections.

[2] DELOS, see http://www.ercim.org/delos

1.3 Preliminary Report

The preliminary report was made necessary because of a delay in the project starting date of two months with respect to the expected date. The project Technical Annex states that the results of the survey conducted in WP1 will be used as input in the definition of the first evaluation campaign. However, the dates of the CLEF evaluation campaigns are not flexible. They are organised on an annual basis and the schedule has been calculated in order to be compatible with the two other internationally known cross-language system evaluation campaigns organised by the National Institute of Standards and Technology (TREC[3] campaign for English/French – Arabic) in the USA and by the National Institute for Informatics (NTCIR[4] cross-language evaluation activity for Asian languages) in Japan. As many research institutions are interested in testing their prototype systems in CLEF, the CLEF schedule also tries to respect the academic calendar. In order to respect these constraints, the CLEF annual campaign schedule is planned as follows:

- Calls for Participation – January
- Data Release - February
- Topic Release - April
- Submission of Runs by Participants - June
- Release of Relevance Assessments and Individual Results – August
- Submission of Paper for Working Notes – September
- Workshop – September (in conjunction with the European Conference for Digital Libraries)

The CLEF 2002 campaign had thus to begin in January with the release of the Calls for Participation in which the different evaluation tracks are clearly described. However, this date was in conflict with the date established for the completion of the survey on user needs. For this reason, it was decided to produced a preliminary report based on a survey made of the CLEF2001 participants. This was an already well-identified group of system developers who could be expected to have a strong interest in shaping future campaigns according to their needs. The input from this group would be used in the preparation of the 2002 evaluation campaign – whereas the final results from the entire survey (including additional system developers and end users) would provide input for the preparation of the CLEF 2003.

The preliminary survey was completed in December 2001. The results have been incorporated into this document which reports on the results of the entire survey. The Deliverable is thus structured as follows: Section 2 provides a description of the system developers part of the survey, extended with respect to the preliminary report; Section 3 discusses the results obtained from the technology deployers, i.e. cross-language information service and content providers. The final section sums up the overall results of the survey and discusses to what extent they can be used as input when studying the design of future evaluation campaigns.

[3] TREC, see http://trec.nist.gov.
[4] NTCIR, see http://research.nii.ac.jp/ntcir/

2. Technology Developers

The target group identified here consisted mainly of previous participants in CLEF campaigns plus people who had indicated interest in CLEF activities or the intention to participate in the future.

2.1 The Questionnaire

The purpose of the questionnaire used for this group was to try to understand their needs and to receive suggestions on how to improve future CLEF campaigns to better meet these needs. Three types of questions were included with the aim of:

- establishing the profile and interests of the responding group
- assessing potential availability of tools or resources.
- eliciting suggestions for future CLEF campaigns.

The questionnaire is given in the Appendix. It was distributed to CLEF 2001 participants and all those requesting information on CLEF 2002.

We finally obtained 35 responses from different groups which is a significant percentage of the R&D groups active in the field . Table 1 lists the profiles of the groups that completed the questionnaire. Probably the most salient feature is that a large part of these groups are research groups and just 8 are from industry (see Figure *2*). Only 5 of the research groups are actually porting their research in CLIR into commercial applications, indicating that the market for CLIR research is still not mature.

The surveyed groups currently handle a total of 22 different languages, mainly European but including some Asian languages. The most widely used languages are those already represented in the CLEF 2001 document collection: English, French, German, Spanish, Italian and Dutch. Immediately following are Finnish, Swedish and Chinese; the first two thus appeared good candidates for further incorporations into the CLEF test suite. Figure 1 shows the most widely used languages. 10 languages that are only handled by one group have not included, namely: Norwegian, Swiss-German, Arabic, Catalan, Danish, Korean, Galician, Polish, Turkish and Greek.

Table 1: Profiles of the Groups that replied to Questionnaire1

Organization Type*	CLIR-related projects	URL(s)	Languages Handled**	Commercial applications	Publicly available resources
+Dutch Industry (spin-off company)	Aqua Browser, Liquid solutions (Medialab)	http://www.medialab.nl	NL,EN,DE	Intranet websites to disclose large corporate datasets.	-
+Non-profit research inst.	Druid, Olive, Twenty-One	http://dis.tpd.tno.nl	EN,FR,DE,NL	www.irion.nl spin-off company commercializing Twenty-One system.	Yes. parallel texts and tools for Dutch. Already made available to CLEF participants via CLEF Web site
+Non-profit Italian research inst.	MUNST	http://munst.itc.it	EN,IT	-	-
+US university	Russian GIRT	http://www.sims.berkeley.edu	EN,FR,DE,IT,ES	-	sims.berkeley.edu /muleum2.html
+US university	HAIRCUT	-	EN,FR,DE,IT,NL, ES,JA,ZH,AR	-	-
+Dutch university	Derive, Computing with meaning	http://www.science.uva.nl/~mdu/Projects{Derive,CoMe}	DE,NL,IT,EN	-	
+Non-profit Swedish research inst.	Clarity DUMAS SPL	http://www.sics.se/humle/clir	EN,SW,DE,FR,FI	-	Possibly, see web page
+US university	MuchMore	http://www.cs.cmu.edu/~yiming	DE,EN	-	Not yet
+Swiss university		http://www.unine.ch/info/clef	EN,FR,IT,DE,ES	-	Yes. See web page. Stop lists and stemmers in FR,IT,EN,DE,ES Most freq. words FR;IT,EN,DE,ES
+Finnish university	UTACLIR	http://www.info.uta.fi	FI,SW,DE,EN (FR,IT,ES to be included)	Negotiations might take place	Yes. Parallel web pages. Maybe statistical trans. models, sentence aligner.
+Canadian university	RALI	http://www-rali.iro.umontreal.ca	EN,FR,IT,DE	-	Yes. parallel web pages made available to CLEF participants
+Spanish university	SINAI	-	ES,EN	-	
+Spanish university	IR-n system	http://www.dlsi.ua.es/~llopis/IR-n	ES,EN	-	
+Thai university	Dictionary-based Thai CLIR	http://www.sci.ku.ac.th	TH	-	
+Spanish university	Hermes, ETB	http://sensei.lsi.uned.es/NLP	ES,EN,CA,IT,FR	-	EuroWordNet via ELRA
+Spanish university		http://milano.usal.es/karpanta	ES,EN	-	Teaching search engine
+Canadian Industry	Hummingbird SearchServer™	http://www.hummingbird.com	EN, DE, FR, IT, ES, NL,PO, FI, SW, DA,JA,KO, Norwegian-Bokmal, Norwegian-Nynorsk, Simplified-Chinese, Traditional-Chinese	electronic publishing, e-commerce, customer care, online technical support and others	Not currently. Some analysis utilities available at a future time
+Spanish university	UDC	http://coleweb.dc.fi.udc.es	ES, GL (Galician)	-	tokeniser, pretagger, tagger. Contact grana@udc.es
Dutch Research Inst.	MIA at CWI CNLP	http://www.cwi.nl/WTCW/MIA http://www.cnlp.org	NL,EN	-	Muscat stemmers. Contact arjen@acm.org

Table 1: Profiles of the groups that replied to Questionnaire (cont.)

Organization Type	CLIR-related projects	URL(s)	Languages	Commercial applications	Available resources
+Taiwan university	NTU	http://nlg3.csie.ntu.edu.tw	ZH,EN	-	-
British Industry	IR	public information not available	FR,IT,DE,ES	-	-
British university	No specific project; just research activity in CLIR	not available	EN, NL	-	
+French research institute	e.Court, IRAIA, Mercure-clir	http://laplace.intrasoft-intl.com/e-court http://iraia.diw.de http://www.irit.fr/ACTIVES/EQ_SIG	DE,FR,EN,IT, PL	-	See IRAIA web page
US industry	LISE	public information not available	FR,ES,DE	support on-line legal search services	-
+ Japanese Research inst.	CLIR (NAIST)	http://db-www.aist-nara.ac.jp/~aki-mae/MLKD/index.html.en	FR,EN,JA,ZH	-	-
+British university	CLEF	http://www.ercim.org/publication/ws-proceedings/CLEF2/matoyo.pdf	FR, EN	-	Yes. See: http://dotty.is.city.ac.uk/okapi-pack/okapi-pack.html
British university	GRIMM	not yet available	FR,IT	-	-
US university	Rapidly Retargetable Translingual Retrieval	http://tides.umiacs	-	-	Yes. Term-for-term translation software and iCLEF document selection interface. Contact oard@glue.umd.edu
Italian University	ISIR	http://www.dei.unipd.it/~ims/isir.htm	IT, ES	-	-
French University	LIA	http://www.lia.univ-avignon.fr/themes.html	-	-	-
Industry	LexiQuest	http://www.lexiquest.com	FR, EN, DE, ES, NL	Intranet and Web site applications	-
Industry	Multimedia Solutions	http://www.lug.com	FR, EN	-	-
Industry	VERITY	http://www.verity.com http://www.veritydemo.com	EN, FR, ES, DE, CH, TH, TR, EL, etc.	Intranet, web and e-commerce applications	-
Non-profit Italian Research Inst.	Infoweb	-	IT	-	-
Industry	TROPES	http://www.acetic.fr	FR, EN, ES, DE, PO, IT	-	-

* CLEF 2001 participants are indicated by +

** ISO 639 2-letter language codes are used (see: http://www.loc.gov/standards/iso639-2/)

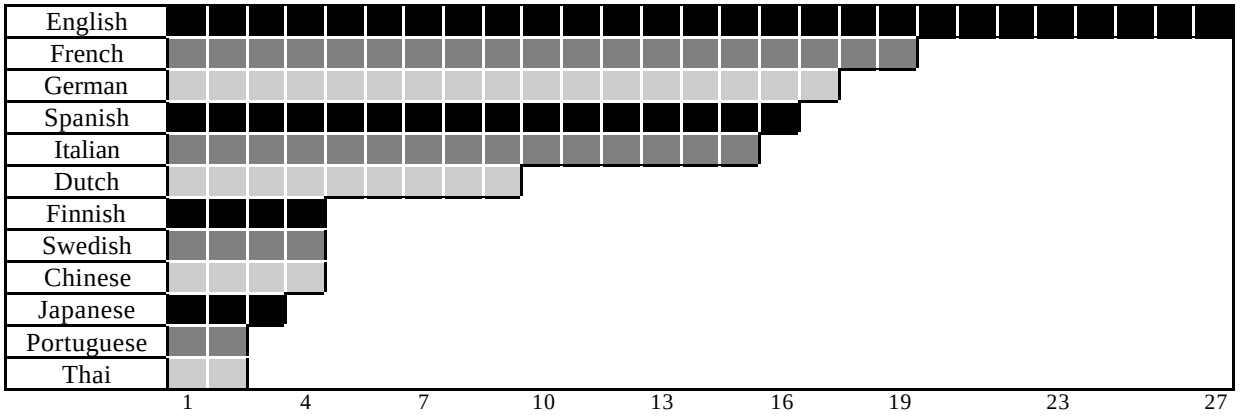

Figure 1. Languages handled by the multilingual systems of respondents to Questionnaire 1

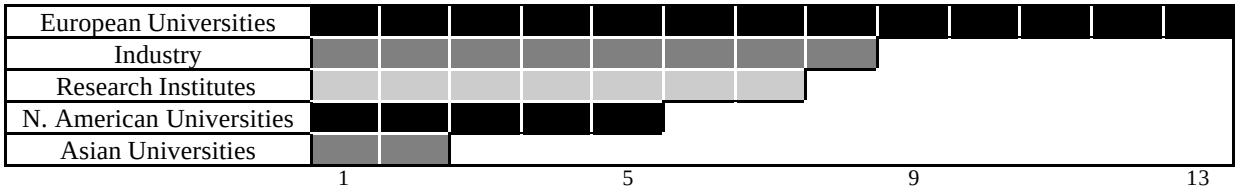

Figure 2. Organization types of respondents to Questionnaire 1

2.2 Results

As stated, the results for this questionnaire were obtained in two stages. In order to produce the preliminary survey, we first contacted last year's CLEF participants. From their replies, it appeared that they were reasonably happy with the distribution of tasks provided by CLEF 2001. The main requests made were the following:

- Include more languages in the multilingual document collection.
- Increase the size of the document collection.
- Include more languages in the topic set.
- Increase the number of topics offered.
- Offer a more diversified set of tasks.
- Improve the pool quality.
- Provide the possibility to test on different text types (e.g. structured data).
- Provide a standard set of resources so that groups can compare the effectiveness of their retrieval system under the same conditions.
- Provide more assistance for newcomers (including improve the Web site).

These requests were used as input when making the final decisions for CLEF 2002[5].

Interestingly, the extension of this questionnaire to additional groups did not make any substantial difference to the results of the preliminary survey – it mainly reinforced the opinions already given. In particular, respondents were interested in the provision of the languages they were currently working with in the document collections and/or the topic sets.

[5] The CLEF 2002 agenda can be seen at http://www.clef-campaign.org

This would mean the inclusion of European languages such as Polish, Turkish, Catalan and Galician. Korean and Arabic were also listed. These languages are, however, covered by the current NTCIR and TREC tracks for cross-language system evaluation and will not be considered for CLEF at the moment. There was also interest in the addition of new tasks covering different aspects of IR: multilingual web retrieval, text-categorization, question-answering and domain-specific retrieval tasks were the most requested. Several groups mentioned the need to be able to test controlled –vocabulary and metadata retrieval as well as free text. Other tasks mentioned included image caption retrieval, phrase browsing, multimedia retrieval.

Issues mentioned by single questionnaires as of importance in evaluation are scalability, dynamic indexing and merging of rankings. There were also some suggestions as to how the current evaluation methodology adopted by CLEF could be improved. Suggestions included providing more relevance levels, offering a query by query analysis, and a baseline translation (monolingual English). The need for the evaluation of aspects of user satisfaction instead of focussing entirely on system performance was also evidenced.

Figure 3 shows the most common recommendations received. In 2.3 we provide a more detailed response analysis and in 2.4 we discuss to what extent we have been able to use the input from the system developers survey in the design of CLEF 2002 and what suggestions it may be possible to adopt for CLEF 2003.

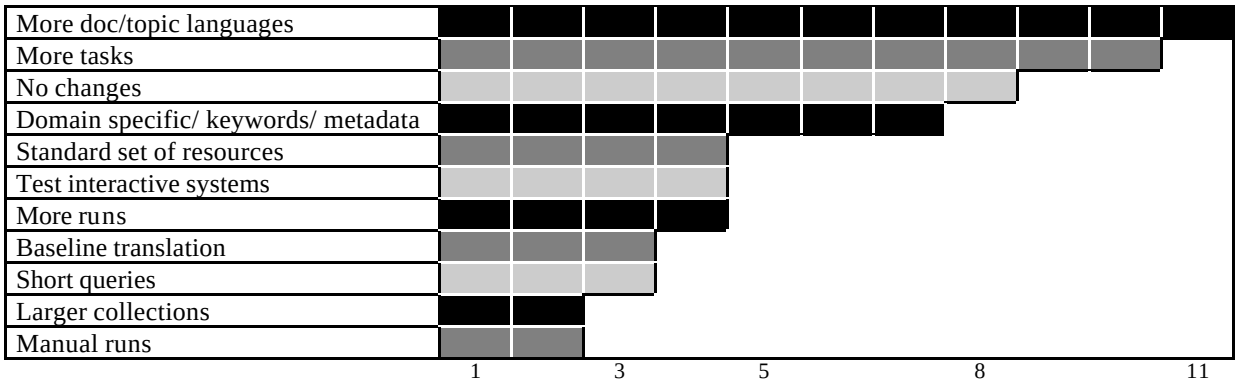

Figure 3. Most frequent suggestions in Questionnaire 1

2.3 Response Analysis

In general, the responding groups seemed quite satisfied with the current organisation of the CLEF campaigns. No major requests were made by a significant number of respondents. However, a number of interesting suggestions were received. We have organised the main observations according to different aspects of the CLEF activities:

2.3.1 Tasks

Overall, the groups seemed to be happy with the three main tracks offered for monolingual, bilingual and multilingual system evaluation in CLEF 2001. In particular, recommendations were received to continue with the experimental interactive task and to study ways in which it could be extended..

Another important request was for a domain-specific task, which is more clearly differentiated from the other tasks. Participants wanted the opportunity to test their systems on structured data as well as the free-text supplied by the newspaper corpora. It was requested that if the GIRT collection[6] continues to be offered, the use of the keywords provided should be allowed. Other participants requested tasks using different domains with an even more specialized vocabulary/structure, such as patent retrieval.

Apart from this, the most frequent requests were for: a) a task with fixed resources for all participants, so that different techniques can be better compared; b) a categorization task; c) multilingual web information retrieval; and d) a multilingual question answering task.

Other suggestions include multilingual phrase browsing, multilingual summarization, multilingual topic detection, interactive query expansion, retrieval of image captions, speech-based retrieval, metadata retrieval, and routing.

2.3.2 Test Collection

The CLEF2001 test collection consists of a multilingual document collection in six languages (Dutch, English, French, German, Italian, Spanish) and topic (query) sets in 12 languages (also Finnish, Swedish, Russian, Thai, Japanese and Chinese). Many isolated suggestions were made:

- Increase number of query languages. Portuguese (4) was the most popular, Danish, Polish, Arabic, Chinese and Turkish. The continuation of topic sets in Finnish (2), Russian, Swedish was also urged.
- Include Dutch in the multilingual task (currently the multilingual task is on a five-language collection).
- Separate results in the bilingual task according to the target language.
- Provide the multilingual task with as many European languages as possible.
- Craft experiments using the tags in the Dutch collection.
- Increase the size of the collection.

2.3.3 Queries

The queries are extracted by the participating systems from sets of structured statements of user needs, denominated “topics”. Each topic consists of three fields: a brief title statement; a one-sentence description; a more complex narrative specifying the relevance assessment criteria. The English version of a typical topic from CLEF 2000 is shown below:

> **Title**: Drugs in Holland
> **Description**: What is the drugs policy in the Netherlands?
> **Narrative**: Relevant documents report regulations and decisions made by the Dutch government regarding the sale and consumption of hard and soft drugs.

The motivation is to provide query “input” for all kinds of IR systems, ranging from simple keyword-based procedures to more sophisticated systems supporting morphological analyses, parsing, query expansion and so on. Participating systems can submit runs that use any combination of these fields in their query construction.

[6] The GIRT collection is a structured database of social science documents.

Four suggestions were made regarding the topic creation:

- permit short queries without "constraints".
- include a base comparison on title runs only (more realistic).
- provide more local topics to better address the results merging issue.
- provide more topics to obtain better discriminating power

The first two requests need some comment to be comprehensive. There is much interest in the development of cross-language web search engines, and it is well known that the typical query to such engines is very short (approx. 2 terms only). This makes system testing using the title run very interesting to many groups. However, due to the high costs of performing manual assessments of the results submitted by participating systems, relevance assessment is performed only on the basis of the narrative. Looking at the example topic shown above, this means that documents found by systems using queries constructed just on the title field will be judged as irrelevant if they only talk about drugs in Holland without referring to the governmental regulation in the Netherlands concerning drug consumption and sale. Furthermore, finding documents that simply talk about “drugs” and “holland” would not be testing fully a system’s search capabilities. It has to be remembered that CLEF and similar system evaluation activities are based on contrastive analysis of results, and not on absolute values, and thus such apparent inconsistencies are of little relevance. However, many participants or would-be participants do not understand this point. See [2] for an important discussion on this aspect of system evaluation.

2.3.4 Pooling

The number of documents in large test collections such as CLEF makes it impractical to judge every document for relevance, as would theoretically be required to calculate recall-based evaluation measures (relevance assessment is expensive as mentioned above). Instead, approximate recall figures are calculated by using pooling techniques. The results submitted by the participating groups are used to form a "pool" of documents for each topic by collecting the highly ranked documents from all the submissions. The assumption is that if a sufficient number of diverse systems contribute results to a pool, it is likely that a large percentage of all relevant documents will be included. All documents not included in the pool remain unjudged and are therefore assumed to be irrelevant.

A main concern with such a pooling strategy is that if the number of not detected relevant documents is above a certain (low) threshold, the resulting test collection will be of limited future use for non-participants, since their systems did not contribute to the pool. A grossly incomplete pool would unfairly penalize such systems when calculating precision and recall measures.

With respect to this issue, some suggestions already discussed during CLEF 2001 were made on the questionnaire:

- Allow English monolingual to improve pool, or use different bilingual target language.
- Allow using all textual fields to improve pool.
- Allow more runs per language to allow more “official experiments”.
- Allow manual runs to improve pool quality.

2.3.5 Assessment

In general, there were no comments on the assessment procedure, except for one respondant that asked for different levels of relevance, instead of the binary relevant/not relevant distinction used in CLEF 2001. This type of assessment is currently adopted by NTCIR [2].

2.3.6 Management

The previous editions of the CLEF campaigns were organised to a large extent on a voluntary basis. Two of the suggestions made on the questionnaire reflected this fact:

- Improve the structure of the Web site.
- Provide more assistance to help new participants.

It was also suggested from academic participants that the workshop is a bit too late (conflict with classes).

2.3.7 Resources

There is a general claim for standard resources, already mentioned in "new tasks". People want standard resources to have more meaningful comparison of approaches, i.e. to make it possible to compare system performance without "interference" from the particular transfer/translation component used. Other suggestions were for the provision of a baseline translation to English for bilingual task and the design of a standard test for lexical coverage.

The sharing of tools and resources is not common behaviour, but there are a few interesting cases of resource-sharing, e.g. the parallel corpora provided by RALI/U. Montreal in CLEF 2000 and CLEF 2001. One of the most competitive groups in CLEF 2001 is also offering stop lists, stemmers and frequent word lists in the five languages of the multilingual task, which can be a very valuable resource for future campaigns.

The question "Do you envisage to provide part of your archives for evaluation purposes?" was not answered affirmatively by any of the participants. Apparently CLEF participants or would-be participants are not archive handlers; obviously encouraging closer relations between DL projects and CLEF could be of benefit here.

2.4 Input for CLEF 2002 and CLEF2003

As has been stated the first results obtained from the system developers questionnaire were used to help make the final definition of the tasks for CLEF 2002. Others will be borne in mind for CLEF 2003. Each comment and suggestion from the respondents has been considered carefully, however, clearly it will never be possible to implement all of them. The reasons are diverse. Some suggestions were only made by one group and must thus be considered as reflecting a minority opinion (this is not necessarily a reason for refusing them, but certainly they need careful assessment); other suggestions are either not realistic or would just be too resource consuming to be practical; a third set of suggestions did not fit in with the general philosophy of the CLEF-style organisation and as such are rejected.

Where possible, the main recommendations listed above have been incorporated into CLEF2002 as follows:

- More languages in the multilingual document collection.
    - *A Finnish newspaper collection 1994-95 has been included.*

- Include more languages in the topic set.
    - *The twelve topic languages offered in CLEF 2001 will be offered again; additional languages will be provided on demand – as long as independent competent groups accept the task of providing reliable translations following CLEF topic production criteria (already promised for Portuguese – one of the languages requested in the questionnaires).*

- Include different text types (e.g. structured data).
    - *CLEF 2002 has now added a track for scientific document collections: this provides monolingual and cross-language tasks for searching the GIRT German database for social sciences with thesaurus in German, English and Russian (already used in CLEF 20001) and the Amaryllis multi-disciplinary scientific database of approximately 150,000 French bibliographic documents with English/French controlled vocabulary (new).*

- Increase the size of the document collection.
    - *Existing collections remain constant but new collections have been added (see above).*

- Increase the number of topics offered.
    - *This is not possible at the moment; topic creation and relevance assessment is very resource consuming; however, the fact that the main multilingual, bilingual and monolingual collections remain the same as last year means that at the end of the CLEF 2002 campaign, the test collection CLEF2001/2002 will include a topic set of 100 items for twelve languages.*

- Improve the pool quality.
    - *This has been a concern in particular with respect to the pool depth used (i.e. the number of documents per run added to the separate language pools); pool construction is a very delicate task (see [1]); a series of tests have however shown that the CLEF2000 pool was very stable [3] and this result has since been sustained by the results reported for the NTCIR pool [4]. The CLEF2001 pool is now being studied. Our final decision for pool depth for CLEF 2002 will depend to a large extent on the result of this study. We will consider also adding some monolingual runs to the CLEF2002 language pools to guarantee their coverage.*[7]

- Offer a more diversified set of tasks.
    - *CLEF 2002 will offer an interactive track, focusing on alternatives to help users to formulate their queries, to refine them and to select the relevant documents in other languages; this track is additional to the four tracks listed in the Technical Annex.*

[7] The pooling question for this type of evaluation activity is complex and a detailed discussion is out of the scope of this deliverable. The interested reader is referred to the literature cited.

- Provide a standard set of resources so that groups can compare the effectiveness of their retrieval system under the same conditions.
    - *This is an interesting point; we will attempt to make some provision for this, enlisting the assistance of CLEF 2002 participants. When activated, this will be one of the points treated in the Web-based discussion area.*

- Provide more assistance for newcomers (including improve the Web site).
    - *The Web site has been totally redesigned; more information is provided and a comprehensive set of Guidelines for participants is under preparation (a preliminary version is already available). A discussion area will be included.*

- Include more languages in the multilingual task – in particular more than one group requested the addition of Dutch
    - *This request was discussed but it has been decided to maintain the multilingual task on the five-language(EN,DE,FR,IT,SP) collection for CLEF 2002.*

- Provide separate bilingual results for each language in the multilingual document collection
    - *After much discussion, this request has been accepted. In CLEF 2002, bilingual tasks will be possible for the French, German, Italian, Spanish and Dutch collection. English as a target collection will only be possible for newcomers to the CLEF activity (in order to avoid overweighing the English document pool). Finnish has been excluded from the bilingual task for this year. There are pros and cons to this extension of the bilingual track. There is a risk that there will not be a relevant number of runs for all languages, and thus results will be of little significance.*

- Provide and assess Web-style (i.e. very short) queries.
    - *Although it is easy to understand the motives for this request, it is not realistic for the reasons listed previously. This point will be another item for discussion on the Web site.*

The agenda for CLEF 2003 still has to be defined. Care will be taken to respect as far as possible the recommendations of this questionnaire, including some that we could not include in CLEF 2002.. In this respect, we hope to be able to:

- Further extend the document collection,
    - *in particular we would be interested in adding a collection for a Slavic language if possible.*

- Extend the topic set with new languages, on demand.

- Add tasks for new types of data.
    - *CLEF will be following with interest and supporting an experiment for the evaluation of systems for cross-language retrieval of spoken documents conducted by DELOS this year; if successful, we hope to incorporate an activity of this type into CLEF 2003.*
    - *we have also received a proposal for the setting up of a text categorization evaluation in multiple languages; this proposal will be considered, but may be too resource consuming under the current budget for CLEF.*

3. Technology Deployers

As stated in the Introduction, the second target group for our survey consisted of cross-language technology deployers, i.e. providers of multilingual information contents and services. A priori we identified the following user classes as potential deployers of cross-language search capabilities:

- EU projects related to Digital Libraries and multilingual contents. This was considered as a primary group of deployers for CLEF, for many reasons, including their EU affiliation (hence their handling of European languages) and the multilingual nature of most contents .
- News agencies, which are already providers of data for CLEF test suites.
- Information services, including scientific publications, juridical information, educational resources, business information, etc.
- E-commerce companies and related EU projects.
- Multinational corporative information servers

3.1 The Questionnaire

We attempted to contact representatives from companies and projects in these areas, asking them to compile a specific questionnaire. When the first contacts were successful, they were generally followed up by an interview in which we identified their company/project profile, the typology of their clients/users and tried to establish how cross-language capabilities could be of benefit to them.

The harvesting of responses stopped just in time to produce this deliverable. We obtained 14 responses with at least one representative member of each of the classes above. In particular, we managed to acquire information from a representative set of EU Digital Library projects. These were the respondents:

- Digital Libraries and multilingual content projects: AMICITIA, ARION, COVAX, RENARDUS, ECHO, ETB, SCHOLNET, COLLATE.
- (multilingual) News Agencies: Agencia EFE (the major Spanish news agency)
- Information Services: Ouest France (a documentation center for journalists), Max Planck society (organization of German research institutes), INIST/CNRS (government provider of scientific and technical information in France).
- E-commerce: Leroy-Merlin company, MKBEEM EU project.
- Corporative information servers: EADS.

3.2 Results

Some general conclusions can be drawn from the overall set of answers:

- There is not sufficient awareness of CLIR state-of-the-art research among the content and technology providers that constitute the natural target group for CLIR technology.
- The needs of potential users are not so much focused on the precision/recall excellence of the CLIR search engines (as measured in the main CLEF tracks), but on other usability issues such as presentation of results (e.g. cross-language summaries of foreign-language documents), document selection and description, or query formulation. This suggests that the new interactive track (not foreseen in the TA) might help to bridge the gap with potential CLIR users.

- Non-textual material such as video or images is common content, and could benefit more from CLIR capabilities. It seems that a speech track (using automatic speech transcriptions) or a track with small documents (such as document summaries, image captions or metadata) could also bring the CLEF activity closer to the user needs.

In the remainder of this section, we will analyse each of the user classes in more detail.

3.3 Response Analysis

3.3.1 Digital Libraries Projects and Multilingual Contents

Eight EU projects participated in the survey. A summary of their answers to the questionnaire seen in *Table 2.* From these answers, together with the results of the follow-up contacts seen in *Table 2.* From these answers, together with the results of the follow-up contacts with representatives from these projects, we can draw some interesting conclusions.

Table 2: Digital Libraries projects

Project	Users	Type of info	Domain	Languages	CL search need	CL strategy	Problems
AMICITIA	Broadcast professionals: Journalists, achivists, documentalists	Film Video clips	News materials	EN, NL, GE + FR, ES	Query translation (QT), Translation of document description (DDT)	Multilingual thesaurus	Results presentation.
ARION	Scientists, local authorities	Scientific data sets	Coastal zone info	EN + GR, IT, NO	QT, fully automatic	(future) public domain MT	Query formulation.
COVAX	Researchers Librarians General public	Text	Energy, environment, Spanish literature, Navy museum.	IT, EN, ES, CA, SW, GE	QT, DDT, document translation (DT), browsing, fully automatic.		Query formulation.
RENARDUS	Academic (professors, students)	Text and others	Education	EN, GE, FI, SW, NL, DA + FR, NO	QT, DDT, browsing, user interface	Cross-language browsing via DDC	Translation of thesaurus of different kinds.
ECHO	Archivists, educational, academics, historians, film	Film	News 1900-1960	IT, NL, FR, GE	QT, DDT, automatic + interactive	Search metadata w. controlled vocabulary + full text search	Results presentation, document description
ETB	Professors, students	Text, web pages	Schools, education	GE, EN, FR, IT, ES.	QT, DT, DDT, browsing	QT by multilingual thesaurus + statistical Correlation text with thesaurus terms	Query formulation, document description.
SCHOLNET	Academic community	Text, film	Computer science	EN, FR, SW, GE, IT, GR + European Union lang.	QT, fully automatic	Multilingual thesaurus + pseudo-relevance feedback (and English summaries)	Results presentation.
COLLATE	Students, archivists, journalists, film industry, film scientists	Text, film, image.	Film, historic docs., movies	EN + GE, CZ, others	QT, DDT, browsing, automatic + interactive		Query formulation, document description, selection, discrimination.

The answers to the questionnaire have shown that:

- There are many digital libraries whose contents are mainly multilingual, and offer some kind of search as a main component of their related projects. Almost all of them have contemplated the possibility of adding a cross-language search mechanism. However, such mechanisms are usually very simple and based on a limited multilingual thesaurus built (or adapted) for the project. In general, there is little awareness of CLIR research. And vice versa, there is little concern about multilingual thesauri in mainstream CLIR research.
- Video, film and/or images are a part (or even the main part) in 5 out of 8 projects in the survey. This is a strong indication in favour of considering multimedia cross-language retrieval as an interesting task to attract the DL research community.
- In most cases, the search is conducted on metadata: rich document, film, image descriptions. Therefore, the texts to be searched (document summaries, video and film descriptions, image captions) are much shorter than those in the current CLEF test suite (newspaper and news agency material). Some respondents also mentioned the presence of mixed-language documents.
- The presentation of retrieved (foreign-language) information is as important as the retrieval process itself.
- Three user profiles seem most popular, and could be considered for domain-specific tracks: 1) researchers/scientific contents, 2) teachers-students/educational contents, and 3) broadcast professionals/news and historic documents.

3.3.2 News Agencies: Agencia EFE

Agencia EFE is the main Spanish news agency, creating and distributing multilingual, multimedia news material. They produce textual material mainly in Spanish, but also in Catalan, English, Arabic and Portuguese. In the case of Catalan and English, the news is mostly translations of the Spanish originals. In the case of Arabic and Portuguese, the news is originally produced in these languages by EFE delegations in Granada and Brasilia.

Agencia EFE is also carrier for photographs and news produced by other agencies. For instance, EFE handles photographs produced by EPA (European Press Photo Agency), which have short English descriptions of their content. EFE changes or/and translates such descriptions into Spanish, ending up in a bilingual photo library.
The textual material produced by EFE also includes press reviews, biographies, and other background material.

There are, at least, two kinds of users that should be considered in relation with a news agency such as EFE:

1. The staff who write, organize and prepare the textual and photographic material to be distributed.
2. The customers who access the material for their particular purposes.

News Production and Handling

In news production and processing, a number of tasks are already seen as problematic without full CLIR capabilities: this is the case of EPA photograph handling at EFE. A multilingual

system able to accept Spanish queries, retrieve English image captions, and translate them with enough accuracy for classification purposes, would facilitate EFE tasks.

In general, a full multilingual search engine that is able to present cross-language summaries of the documents retrieved is seen as a potentially useful tool for news writers, although its incorporation as a common tool would not be immediate, as it is the case with any new tool that involves re-adapting already developed work methodologies.

EFE Clients

There are many different profiles of EFE clients:

- Newspapers and magazines, national and international, such as *El Pais* or the *New York Times.*
- Televisions and radio stations.
- Internet portals, such as *Terra Networks, Ya-sports, Navegalia*, etc.
- Administrative and political entities such as Spanish ministries, city councils, etc.
- International organisms such as the UN or the OID.
- Libraries and Documentation services.
- Law firms.
- Companies of diverse genre such as *Seat, Philip Morris or Alcampo*.

In principle, a complete CLIR engine (which retrieves and also translates material) would enhance the EFE Internet data servers (*EFE data*), broadening the amount of usable information for many of such clients. There is not, however, sufficient current demand for such capabilities that would set CLIR capabilities as a priority for EFE. The impression from a number of conversations with Manuel Fuentes and other people at EFE is that the possibility of having CLIR capabilities raises moderate interest, which could be high enough to devote EFE resources in a mid-term future.

New Evaluation Tasks

The above interview suggests possible two evaluation tasks:

1. CLIR evaluation over short texts (e.g. image captions).
2. Document selection/description with short translations.

As was the case for DL projects, there is not much awareness of CLIR research, but there is a moderate interest that could lead EFE to include cross-language search mechanisms if a flexible system, considering all aspects of a search session (including presentation of the information retrieved via cross-language summaries/descriptions), becomes available on the market. CLIR evaluation over short texts (for image captions) is perceived already as a practical issue.

3.3.3 Information Services

Three information services[8] were surveyed: Ouest France, a documentation centre for journalists in France; INIST/CNRS, a provider of scientific/technical information in France; and Max Planck, a consortium of research institutions in Germany.

[8] Some of the EU projects surveyed are also future information services. We have included in this category only information services currently available for users.

Unfortunately, we could not obtain information from personal interviews, and the questionnaires do not provide many details. The two French institutions seem to work mainly in French, with some interest also in English. They are not planning to add cross-language capabilities in a near future.

The Max Planck institute handles scientific information in all areas, mainly in German and English. They are not currently planning to add a cross-language search capability, but domain-specific access to information is important for them, and they see query translation and browsing as their major needs in a Cross-Language setting.

3.3.4 E-commerce

We obtained a response from a company with e-commerce facilities, Leroy-Merlin, and we studied the MKBEEM e-commerce EU project to deduce whether they might have a need for Cross-Language search. For different reasons, CLIR does not seem to suit their needs.

Leroy Merlin is a company that offers a "do it yourself " information/advise system online, together with their online catalogue. Currently, it is aimed at a French-speaking public. Although they are planning to provide the same services for Spanish speakers, the proposed approach is just to translate their service and their online catalogue; thus, no cross-language search is needed.

On the other hand, the MKBEEM (multilingual knowledge-based European electronic marketplace) project is developing a mediation system which adapts the language and the trading condition of an internet sales point according to its international customer. The user partners are Ellos (FI), SNCF (FR) and FIDAL (FR).

Multilinguality is represented in the project by "automated translation and interpretation of natural language user requests" into domain ontology models in order to derive relevant concept bindings. In principle, this multilingual component is much more ambitious than a traditional CLIR tool, because queries are processed to match domain ontology models, not textual catalogues. So a CLIR capability does not have a direct integration with the system. What the project does need is multilingual query processing, but not of a traditional CLIR nature.

CLIR capabilities seem more useful for meta-providers (such as the Altavista shopping facility) than for individual providers such as SNCF (trains, hotels) or Ellos (clothes). A meta-provider search facility could provide access to similar products and compare prices, regardless of the original language in which they are offered. For the transaction itself (e-commerce), CLIR capabilities do not seem a must.

With respect to meta-providers, an interesting evaluation task would be to find products that suit user needs in a multilingual repository of catalogues of various genres. The task would be halfway between CLIR/question answering/information extraction.

3.3.5 Corporative Information Servers: EADS

EADS is a multinational corporation in aeronautics, telecommunications, space and defense. They provide textual information in French, English, German and Spanish. According to the questionnaire, they need cross-language functionalities such as query translation, document translation and document description translation within automatic and interactive search

services. They are considering adding a cross-language tool in the future, and perceive query formulation, document description and results presentation as the main problems to face.

This interview confirms that CLIR capabilities are attractive to manage corporative information servers in large companies.

4. Conclusions

The main general conclusion that can be drawn from a comparison between the results of the survey for our two user groups: cross-language system developers and cross-language service and content providers, is that there is still a considerable gap between the work and interests of the R&D community and the implementation of results by the application community. This is not too unexpected: cross-language system development is still in the experimental stage and a real impact is yet to be felt at the ground level.

For this reason, the results we obtained from the first questionnaire are far more immediately usable in the definition of the tasks for CLEF campaigns. We intend to conduct a similar questionnaire on the CLEF 2002 participants, at the end of this year's campaign to see whether additional recommendations are made that can be acted on in CLEF 2003.

However, the information acquired from the second questionnaire is also important for our activity and invites reflection. Two important points emerge clearly:

- Real-world applications do not just regard textual information and document ranking is not the only factor of relevance.
- There is a surprising lack of perception of the need for cross-language functionality, even in applications that are regularly handling information in multiple languages.

The first point urges us to include tasks which evaluate aspects which regard questions such as user satisfaction, passage retrieval, and results presentation, and to consider media other than text, eg spoken document and/or image caption retrieval. The second suggests that there is a strong need for more dissemination among technology deployers of the state-of-the-art of CLIR systems. Content and service providers should be made aware of the additional functionality that could be offered by their system by the inclusion of tools to handle all aspects of multilingual information access.

It remains to be seen how far these points can be covered within the constraints of the current CLEF project activity.

5. References

1. Voorhees, Ellen (forthcoming). "The Philosophy of Information Retrieval Evaluation" to appear in Peters, C., Braschler, M, Gonzalo, J and Kluck, M. (Ed.). Cross-Language Information Retrieval System Evaluation: Proceedings of the CLEF 2001 Workshop, Lecture Notes in Computer Science, Springer Verlag.
2. Kando, N. (2001). "Overview of the Second NTICIR Workshop". In Kando, N., Aihara, K., Eguchi, K., Kato, H. (Eds.). Proceedings of the Second NTCIR Workshop Meeting on Evaluation of Chinese & Japanese Text Retrieval and Text Summarization, National Institute of Informatics (NII), ISBN 4-924600-89-X.

3. Braschler, M. (2001). CLEF 2000 – Overview of Results. In C. Peters (Ed) Cross-Language Information Retrieval and Evaluation. Workshop of the Cross-Language Evaluation Forum, CLEF 2000, Lisbon, Portugal 2000. *Lecture Notes in Computer Science* 2069, Springer 2001.
4. Kuriyama, K., Yoshioka, M., Kando, N. (2001). The Effect of Cross-Lingual Pooling on Evaluation. In Kando, N., Aihara, K., Eguchi, K., Kato, H. (Eds.). Proceedings of the Second NTCIR Workshop Meeting on Evaluation of Chinese & Japanese Text Retrieval and Text Summarization, National Institute of Informatics (NII), ISBN 4-924600-89-X.

Appendix 1

1. Questionnaire Distributed to Technology Developers

Cross-Language Evaluation Forum

CLEF 2001 - Questionnaire

1. What is the name of your CLIR-related project(s)?

   Please give the URL(s) of the Web site(s).

2. What languages are included?

3. Is your research on CLIR ported to/tested in commercial applications with real users? If so, could you make a brief description of such applications?

4. What aspects of your system are not covered by current CLEF evaluation tasks?

5. What improvements would you suggest for current CLEF tasks?

6. What new tasks/languages/features would you like to see in forthcoming CLEF evaluations?

7. Is any of your resources/software packages publicly available? Please include a brief description and a contact web page or e-mail.

8. Do you envisage to provide part of your archives (textual material?) for evaluation purposes and/or HLT R&D?

   Yes ☐

   No ☐ (could you elaborate ……..)

(The questionnaire shown is the one used for the Preliminary Report and distributed to CLEF 2001 participants. A slightly modified version of this questionnaire was sent to System Developers who had not previously participated in CLEF evaluation campaigns. They were first referred to a full description of the CLEF activity before responding.)

2. Questionnaire Distributed to Technology Deployers

Cross-Language Evaluation Forum

http://www.clef-campaign.org

Multilingual Information Access is a key issue in Digital Libraries and other information services. The issues involved include:

- multiple language access, browsing, display
- cross-language information discovery and retrieval

The objective of the **CLEF** project is to promote research and stimulate system development by providing an appropriate infrastructure for system testing and evaluation. One of the project activities will be to identify real world issues by studying the needs of both system developers and end users. As a provider of multilingual information services and/or products, we should be very grateful if you could spare the few minutes needed to compile the following questionnaire.

Questionnaire

1. What is the name of your project/group/company?

Please give the URL of the Web site and/or any other contact points.

2. Do you provide ... based on search technology/engines?
   ☐ products ☐ services

What kind of users do you have?

3. What type of information do you provide?
   ☐ text ☐ image
   ☐ speech ☐ music
   ☐ film ☐ others, please specify below

- ☐ general/unrestricted information
- ☐ domain specific information, please specify which domain

4. What languages do you currently handle?

What languages do you expect to use in the future?

5. What type of information need do you have?
   - ☐ Informational (e.g. “What is the history of the leaning tower of Pisa?”)
   - ☐ Navigational (e.g. “Show me the *url* of the site ...”)
   - ☐ Transactional (e.g. “Where can I buy a ticket for ...?”)
   - ☐ Other, please specify

6. What kind of cross-language functionality do you need?
   - ☐ Query translation
   - ☐ Document translation
   - ☐ Translation of document descriptions
   - ☐ Browsing
   - ☐ Fully-automatic mechanism
   - ☐ Interactive mechanism
   - ☐ Notification
   - ☐ Other, please specify

Are you considering adding a cross-language tool in the future?

☐ No ☐ Yes

If so, have you decided what method will be employed? Describe below.

7. What are the main problems you are currently facing wrt to search and retrieval in archives containing documents in multiple languages?
   - ☐ Query formulation
   - ☐ Document selection/discrimination
   - ☐ Document description
   - ☐ Results presentation
   - ☐ Others, please specify below

8. Any other comments

9. Please, write your name and e-mail (optional)