



Project no. 215231

TrebleCLEF

Treble-CLEF: Evaluation, Best Practices and Collaboration for Multilingual Information Access
IST: ICT-1-4-1, Digital libraries and technology-enhanced learning

Deliverable 3.2

**Workshop on Best Practices for the Development of Multilingual
Information Access Systems: the User Perspective
(TrebleCLEF User Communities Workshop)
REPORT**

Start Date of Project: 01 January 2008

Duration: 24 Months

Organisation Name of Lead Contractor for this Deliverable: UNED

Version 2 (final), December 2008

Project co-funded by the European Commission within the Seventh Framework programme

Document Information

Deliverable number: 3.2
 Deliverable title: *Workshop on best practices for the development of Multilingual Information Access systems: the user perspective* – TrebleCLEF user communities workshop report
 Due date of deliverable: 31/12/2008
 Actual date of deliverable: 31/12/2008
 Author(s): Julio Gonzalo, Anselmo Peñas, Felisa Verdejo, Carol Peters
 Participant(s): UNED, ISTI-CNR, USFD, UNIPD, ELDA
 Workpackage: 3
 Workpackage title: Best Practices in System Development and User Studies
 Workpackage leader: UNED
 Dissemination Level: Public
 Version: 1.00
 Keywords: User Communities Workshop, Interactive Multilingual Information Access, User-Oriented Multilingual Information Access, Technology Transfer

History of Versions

Version	Date	Status	Author (Partner)	Description/Approval Level
1.00	27/12/08	Near-final	UNED	Circulated to all partners for comments
2.00	31/12/08	Final	UNED	Approved by all partners

Abstract

This document summarises the organization and results of the TrebleCLEF User Communities Workshop entitled “Workshop on Best Practices for the Development of Multilingual Information Access Systems: the User Perspective” held 24-25 June 2008 in Segovia, Spain. The workshop brought together representatives from relevant user communities (Cultural Heritage, European government agencies, news agencies, patent and trademark professionals, enterprise and web search companies, and EU projects) and MLIA researchers. Besides technical presentations, the workshop was tightly focused on a debate from which consensus emerged regarding (i) the features that MLIA systems should have from the users’ perspective; and (ii) strategies to provide MLIA technology with these features and transfer these technologies to society. This document summarises the audience, goals, agenda and results from the workshop. Together with the actual workshop, it forms deliverable 3.2 of the TrebleCLEF project.

Table of Contents

Document Information	1
Abstract	1
Executive Summary	3
1 Introduction	4
2 Audience and Agenda	4
2.1 First Day Presentations	5
2.2 Second Day Working Session.....	9
3 Outcome of the Second Day Discussion	10
3.1 Required features for MLIA Systems from a User's Perspective.....	10
3.2 Strategies to Improve User Awareness in MLIA and Transfer Technologies to Society	11
3.3 Immediate Actions for CLEF.....	12
4 Conclusions	12

Executive Summary

This document summarises the organization and results of the TrebleCLEF User Communities Workshop entitled “Workshop on best practices for the development of Multilingual Information Access Systems: the user perspective” held 24-25 June 2008 in Segovia, Spain. The workshop brought together representatives from relevant user communities (Cultural Heritage, European government agencies, news agencies, patent and trademark professionals, enterprise and web search companies, and EU projects) and MLIA researchers.

Despite the large and active research community working on Multilingual Information Access (MLIA) topics, and the maturity of many MLIA technologies, there is still a general lack of commercial MLIA systems available, and a lack of adoption of MLIA technologies in user communities with multilingual information access needs. One of the identified problems, at least with respect to TrebleCLEF activities, is the mutual lack of awareness between CLEF researchers and current – or potential – user communities.

The workshop was aimed at taking a first step towards solving this problem, and had two main goals: (i) provide input for the Treble-CLEF (user-oriented) best-practices report and (ii) bring CLEF closer to current/potential user communities for MLIA services by identifying user communities and application scenarios, increasing the awareness of CLEF results in user communities, and helping in shaping future CLEF campaigns.

The workshop duration was one and a half days, of which half a day was reserved for an intensive working session, where a consensus reaching strategy (“grid of groups”) was implemented to reach a common vision on two specific issues: (i) features that MLIA systems should have from the users’ perspective and (ii) strategy to provide MLIA technology with these features and transfer these technologies to society, considering the use of evaluation forums such as CLEF.

In addition to technical presentations, the workshop was tightly focused on a debate from which consensus emerged regarding (i) the features that MLIA systems should have from the users’ perspective; and (ii) strategies to provide MLIA technology with these features and transfer these technologies to society.

The input from the user communities was very valuable: in particular, a list of desirable features for MLIA systems emerged as a consensus of the workshop participants, and will be reflected in the best practices deliverable. In addition, the workshop produced a few immediate consequences on the CLEF campaign: an initiative to involve user communities representatives in the CLEF steering committee, an initiative for a patent retrieval track in CLEF which is tightly focused on the needs emerging from the patent offices community, and the possibility to use a large MLIA search log from the European Joint Research Centre in the iCLEF track.

This document summarises the audience, goals, agenda and results from the workshop. Together with the actual workshop, it forms Deliverable 3.2 of the TrebleCLEF project. The output from the User Communities Workshop is one of the key sources of information which will be used to compile best practice recommendations for user-orientated evaluation (to be presented as part of Deliverable 3.3).

1 Introduction

This document summarises the organization and results of the TrebleCLEF User Communities Workshop entitled “Workshop on Best Practices for the Development of Multilingual Information Access Systems: the User Perspective” held 24-25 June 2008 in Segovia, Spain.

Despite the large and active research community working on Multilingual Information Access (MLIA) topics, and the maturity of many MLIA technologies, there is still a general lack of commercial MLIA systems available, and a lack of adoption of MLIA technologies in user communities with multilingual information access needs. One of the identified problems, at least with respect to TrebleCLEF activities, is the mutual lack of awareness between CLEF researchers and current – or potential – user communities.

The workshop was aimed at taking a first step towards solving this problem, and had two main goals: (i) provide input for the Treble-CLEF (user-oriented) best-practices report and (ii) bring CLEF closer to current/potential user communities for MLIA services by identifying user communities and application scenarios, increasing the awareness of CLEF results in user communities, and help shaping future CLEF campaigns.

The workshop brought together MLIA researchers and representatives from relevant user communities, namely: Cultural Heritage, European government agencies, news agencies, patent and trademark professionals, enterprise and web search companies, and EU projects. Its duration was one and a half days, of which half a day was reserved for an intensive working session, where a consensus reaching strategy (“grid of groups”) was implemented to reach a common vision on two specific issues: (i) features that MLIA systems should have from the users’ perspective and (ii) strategy to provide MLIA technology with these features and transfer these technologies to society, considering the use of evaluation forums such as CLEF.

The input from the user communities was very valuable: in particular, a list of desirable features for MLIA systems emerged as a consensus of the workshop participants, and will be reflected in the best practices deliverable. In addition, the workshop produced a few immediate consequences on the CLEF campaign: an initiative to involve user communities representatives in the CLEF steering committee, an initiative for a patent retrieval track in CLEF which is tightly focused on the needs emerging from patent and trademark agencies, and the possibility to use a large MLIA search log from the European Joint Research Centre in the interactive CLEF track (iCLEF).

The remainder of this deliverable is structured as follows: the setup of the workshop (audience and agenda; Section 2) is presented first, together with a summary of presentations given by attendees at the workshop. Section 3 discusses the outcome of the workshop debates, and Section 4 gives some concluding remarks.

2 Audience and Agenda

In order to warrant a balanced and representative participation, attendance at the workshop was by invitation only, and included representatives from (actual or potential) MLIA user communities together with MLIA researchers from academia and industry. All workshop attendees were invited to participate in the half-a-day debate, and most of them were invited to make short position statements on the workshop topics in the first day of the workshop.

2.1 First Day Presentations

The first day of the workshop consisted of a series of focused presentations – position statements – by a majority of the workshop participants, organized in five short sessions:

Session 1 The Research Perspective

The goal of this opening session was to inform representatives from user communities about CLEF and TrebleCLEF activities, about the state-of-the-art in MLIA, and about the issues that require their input. These were the contributions:

From Research to Practice: bringing Multilingual Information Access to the Real World

Carol Peters, ISTI-CNR Pisa.

Carol is the general coordinator of CLEF campaigns and principal investigator of the TrebleCLEF project. She gave a brief summary of CLEF goals and activities since its creation, focussing on CLEF achievements and current shortcomings. CLEF has indeed stimulated research activities in previously unexplored areas, has provided de facto standards for MLIA benchmarking, quantitative and qualitative evidence with respect to best practice in cross-language system development, and has built a strong, multidisciplinary research community. However, the CLEF focus has been mainly on system performance, with little attention paid to user needs, and has largely ignored the impact of user interaction. In parallel, there has been a notable lack of take-up by application communities.

Multilingual Information Access from a User-Oriented Perspective

Julio Gonzalo, UNED Madrid.

Julio is co-organizer of user-oriented evaluation studies in the interactive CLEF track and responsible for TrebleCLEF Work Package 3. After Carol's general presentation of CLEF activities, Julio focused on evaluation studies from a user-oriented perspective. He showed that only a tiny fraction of MLIA evaluation includes some user perspective, and this fraction has been mostly done by a few researchers in the context of iCLEF. A summary of iCLEF finding reveals that, while significant research contributions has been made with respect to interactive MLIA, iCLEF has not been sufficiently tied to realistic MLIA scenarios and user communities.

What Systems for what Scenarios? From CLEF Research Challenges to Real-world Opportunities

Maarten de Rijke, U. Amsterdam.

Maarten is one of the most active European researchers in MLIA and one of the key people bringing innovation and self-criticism into CLEF activities. He examined various research cases which are inspired by real-world needs and discussed their distance from current CLEF evaluation activities. Some key issues were: How informed is CLEF about users of multilingual search facilities? To match real world needs, there is a need to start from the users (performing user studies, analysing search logs, contacting potential customers) and simultaneously showcase the technologies produced at a research level. This implies constant innovation in tasks, document types, evaluation metrics, etc. Two concrete suggestions were adding a user group to CLEF, and paying more attention to Machine Translation in the context of complete MLIA applications (notably, CLEF has had little intersection with the Machine Translation research community).

Session 2: User Needs Studies and Surveys

The goal of this session was to provide input for discussion from known user studies related to Multilingual Information Access. We tried to focus on Cultural Heritage as one of the priorities for the European Commission and a natural scenario for CLEF activities.

User Needs Studies and Surveys: MLIA Needs in the Cultural Heritage Domain

Jennifer Marlow, University of Sheffield.

Jennifer is a university researcher who has been involved in several user studies regarding MLIA access in the Cultural Heritage domain. She presented results from a survey made of Tate Online users, a user study researching the MLIA needs of cultural heritage (CH) professionals in the context of the MultiMatch European project (www.multimatch.eu), and the results of an interactive experiment with the Google cross-language search service. Some relevant results were (i) A majority of Tate Online users were not native language speakers and would strongly benefit from the possibility of browsing, reading and searching in their own preferred languages, (ii) Cultural Heritage professionals routinely need to use other languages when searching information at work, usually aided by dictionaries and other people. (iii) Machine Translation was not good enough in the context of CH searches.

MLIA User Needs Studies in the Cultural Heritage Domain: The European Library/TELplus Experience

Giorgio di Nunzio, University of Padova.

Giorgio is a researcher involved in *The European Library* related project, *TELplus*. He explained the methodology used in TEL to satisfy user needs, via user studies, HTTP log analysis, Action/Search log analysis, and questionnaires. He highlighted facts relevant for MLIA needs, such as the fact that the most frequent action after entering TEL was changing the interface language, the variety of user languages from which each collection was accessed, and the fact that the interface was one of the major points of concern for users. Altogether these are clear signs that MLIA must be a key point for the European Library, and that user-oriented design and evaluation must be part of the picture.

Session 3 MLIA Systems in the Real World: Cases of Study

Multilingualism in the Europe Media Monitor NewsExplorer

Bruno Pouliquen, European Joint Research Centre.

Bruno presented insights from his experience in the Europe Media Monitor (EMM) NewsExplorer, a highly multilingual information access application capable of cross-language linking of around 50.000 newspaper articles per day from 1.500 news portals in more than 40 languages, and with around 50.000 visits per day. This application would be a powerful source of evidence for user studies at CLEF. According to Bruno, the need for MLIA capabilities has to be anticipated by the researcher: their users at the European Commission did not perceive the need for such a tool, but once it was ready it became an essential tool for their customers.

MLIA Needs in Government Agencies: a Case Study

Michael Kluck, German Institute for International and Security Affairs, Berlin

Michael has an exceptional profile for this workshop, as he was involved in the organization of CLEF for many years from his position at Humboldt University, and is now working in a German government agency (Institute for International and Security Affairs) which handles multilingual information access to online databases such as the *Database World Affairs*. He provided insights into

the challenges of providing in-built multilingual search capability and integration with domain specific multilingual thesauri.

Session 4 Market Opportunities

Opportunities for MLIA in the Corporate Search Market

Paraic Sheridan, Dublin City University

Paraic is currently working at the *Centre for Next Generation Localisation* at Dublin City University, and he was invited to the workshop as an experienced researcher in MLIA who has spent 10 years of his career working for *TextWise*, an enterprise search company with a US focus. Paraic gave an overview of the enterprise search market and the role of Cross-Language search in it. He explained that enterprise search is embedded in larger enterprise information software – Enterprise Resource Planning, Customer Relationship Management, and Enterprise Content Management – and customers expect search, including its multilingual aspects, to be integrated there. Remarkably, MLIA does exist as an offering by most major Enterprise Search players – such as SAP (www.sap.com), FAST (www.fastsearch.com) or Autonomy (www.autonomy.com) – which in a way contradicts the general impression in the research community that MLIA research is not transferring to market. Paraic also summarized some TextWise experiences deploying ML search applications, with valuable lessons learnt: for instance, a Japanese Patent Search producing good quality rankings was judged unacceptable by clients because it was not integrated into their search product, and because the machine translations of the patents was unreadable. This is a clear signal for the necessity of integrating a concern for MT in MLIA research, at least from the point of view of end-to-end systems. From this point of view, sometime bulk translation (as opposed to query translation) is the easier or most practical option. Finally, he gave an overview of the localisation challenges, and pointed out Business Intelligence as a right target for multilingual processing beyond strict cross-language document search, including entity & topic recognition, content classification, intelligent processing, etc.

Web Search Engines in Europe: is there a Demand for MLIA Facilities?

Amar Djalil-Mezaour, Exalead

Amar was invited as a representative of Web Search Engines in Europe: Exalead currently has the largest web index developed and maintained by a European company, and is involved in Quaero, a large research project aimed at the development of a multimedia and multilingual search engine. He explained that there is, beyond doubt, a demand for MLIA facilities in this competitive emerging market, and in particular in semantic research projects, where multilingualism is a main parameter to capture the meaning of information.

Session 5 User Communities and MLIA: Position Statements

Cultural Heritage & Digital Libraries

Joern Sieglerschmidt, Bibliotheksservice-Zentrum Baden-Wuerttemberg

Joern has been involved in several European projects involving Cultural Heritage and Multilingualism, and gave an overview of the role of multilingual thesauri in search. It was clear that, from the point of view of humanities scholars, multilingual information access is a problem of crafting high-quality multilingual thesauri, a perspective that is alien to a large percentage of MLIA researchers, and that has little representation in CLEF campaigns (the domain specific track being the strongest exception to date).

Dov Winer, Israel National Library

Dov, from the Minerva affiliated Israel National Library, presented Israel as a remarkable case of study for the need of MLIA within a country: Israel has two official languages (Hebrew and Arabic), and the fact that as a migratory country the mother tongue of many citizens is none of the official languages. He also reviewed opportunities for MLIA applications, such as the eTwinning EC programme (with more than 40,000 European schools participating), the Learning Resources Exchange (LRE), the Minerva cluster of projects, and MICHAEL and the EDL/Europeana.

Steven Krauwer, Utrecht University

Steven spoke as coordinator of the CLARIN network (Common Language Resources and Technology Infrastructure for the Humanities and Social Sciences). He explained CLARIN's basic goal is to achieve a European federation of digital archives with language data, with access to language and speech technology tools to manipulate, enhance, explore and exploit the data. In his opinion, there is no current demand for MLIA in the Social Sciences and Humanities (SSH) area, but it has to be created. He also explained that, from his point of view, technology evaluation is not directly interesting for the SSH user, and best evaluation in this context is through success stories and cost/benefit perception by the user.

News agencies

Denis Teyssou, Agence France-Presse, Paris

Denis represented AFP, a news agency which handles 6 main languages and is present in 165 countries. AFP was presented as a clear case for MLIA needs, and Denis explained how MLIA technologies are used with customers, in the journalists' work flow, and in particular regarding alert and filtering on keywords in one language with the capability of retrieving all pertinent documents on the same subject/context in different languages. Using IPTC (International Press Telecommunications Council) standards, queries in one language routinely retrieve documents in Spanish, French and English. He concluded by summarizing basic issues when designing MLIA systems: understanding what users are searching for, what are their goals and strategies, putting users in the loop (feedback), specific handling of proper names (widely searched in the news domain), transliteration, cross-language as a transparent feature, and making foreign docs useful even if the user cannot understand them. He also signalled a relevant application domain for the short-term future: the phenomenon of "twitterization" of information where real-time breaking news is provided not by professionals but by witnesses. There is an increase of speed and more reactivity that need MLIA applications to be successful.

Patent and Trademark Search

John Tait, Information Retrieval Facility, Vienna

John Tait is currently the Chief Scientific Officer of the Information Retrieval Facility, whose mission is to bring the latest information retrieval technology to the community of patent professionals and other professional searchers in intellectual property. He claimed that Patent Retrieval is one of the most clear cases for the need of high-quality MLIA, and with special characteristics: high recall is needed (unlike typical web search) because a single missed document can invalidate a patent; it is session based: single searches may involve days of cycles of results review and query reformulation, and therefore interactive retrieval aspects are of great importance; and it must be defensible, i.e. the search process and their results may need to be defended in court. Multilingualism is essential because, for instance, a Russian patent can invalidate a British patent. Indeed Machine Translation is already

widely used in the area since 2006, and there are powerful visualization tools coming on the market. A special challenge for search is patent obfuscation (somewhat the reverse of web spamdexing): patents are written in such a way that makes them difficult to search. John’s presentation has eventually led to a pilot patent retrieval track in CLEF 2009.

2.2 Second Day Working Session

The second day of the workshop (half a day), chaired by Felisa Verdejo, was organized as a working session with the participation of all first-day speaker plus Felisa and two additional TrebleCLEF representatives: Amedeo Cappelli (CELCT) and Ana García-Serrano (UNED).

The goal was to reach consensus on two main issues:

- **Which are the essential features that MLIA systems should have from a user’s perspective?**
- **Which strategies should be developed to provide MLIA technologies with such features and transfer these technologies to society?**

Instead of a moderated discussion, in order to encourage participation and to obtain a common view of these issues, we used two different mechanisms:

- A questionnaire (see annex A) which was distributed to all participants prior to the workshop, thought to focus and prepare the discussion.

For the debate, we used a known consensus-reaching strategy called “grid of groups”. With this technique, 16 participants receive a task, and they must find a single position which represents everyone at the end of the session. Each participant receives a position in a grid (a letter {A,B,C,D} and a number {1,2,3,4}). The grid is a 4x4 matrix.

In a first step, each column forms a discussion group, where consensus must be reached. Once the allotted time expires, there is a break where secretaries (one per group) exchange their minutes. Then, each row (with one participant from each of the former groups) forms a new group, where again consensus must be reached. Then a representative of each new group prepares a presentation with the conclusions of the group. Finally, each representative makes a presentation to all the community, and a final debate unifies the four positions. The grid used was the following:

	A	B	C	D
1	Julio Gonzalo (Secretary)	Amedeo Cappelli	Michael Kluck	Bruno Pouliquen
2	Paraic Sheridan	Carol Peters (Secretary)	Denis Teyssou	Khalid Choukri
3	Dov Winer	John Tait	Felisa Verdejo (Secretary)	Giorgio Maria Di Nunzio
4	Steven Krauwer	Joern Sieglerschmidt	Jennifer Marlow	Ana García Serrano (Secretary)

3 Outcome of the Second Day Discussion

The 16 representatives from user communities and MLIA researchers achieved a consensus with respect to the two issues presented. This is a summary of the results, which constitutes a valuable input for Deliverable 3.3 (Best Practice Recommendations for the Design of MLIA Systems).

3.1 Required features for MLIA Systems from a User's Perspective

There was general agreement on the fact that a use case must be specified before making an exhaustive list of features: the starting point is a model of a user and a model of the task. For instance, there is little in common between a monolingual web surfer and a patent retrieval specialist with passive knowledge of six languages. However, it was still possible to make some practical observations at a very general level:

Integration

- Systems must be transparent regarding cross-language search. There is no such thing as a Cross-Language Information need: there are simply information needs that cannot be fully satisfied without finding information in other languages. So, by default, there is no need to know what the system is doing and how it is working (but still full control should be available if required, see *search interface* below)
- Multilingualism is just a feature of Information Access systems, and it must be seamlessly integrated.

Search interface

- There should be an advanced search mode that gives user full control over multilingual features (target languages, query translations) for the small percentage of advanced users that want control when things go wrong. There were, however, some disagreements between participants as to whether this still holds when query expansions made by the system are very complex, as happens in the case of the JRC system.
- If possible, link structured sources that help mapping the meaning of the query (e.g. with named entities). Consider Wikipedia.

Results presentation

- Interfaces should be flexible about how to organize results. There must be at least two choices: separated by target language, or merged. The default view depends on the application (and the user profile).
- There should be a choice of seeing the original document or a translation. If translation for a certain language pair is not available, one option is to show metadata: named entities, categories, etc. Another option is to translate into the language which is more familiar to the user (according to his/her profile), or perform some approximate translation (summarization, key concept translation, word-by-word in the worst case) if possible.
- When few monolingual results are available, the system should alert the user whenever there is more information available in other target languages.

- The system should warn about the quality of Machine Translation and about how much authorized is the translation, to avoid wrong expectations from the user.

Personalization

- There must be some support to specify language skills and translation preferences in the user profile, and the interface should adapt to this profile. For instance, default translation should only be provided for languages unknown to the user. If a user profile is not available, the system should have clever default behaviour, e.g. not translating Portuguese documents by default if the query is in Spanish (because Spanish speakers have passive, reading abilities in Portuguese).
- Ideally there should not be an up front log profile. It is better to start with a sensible default, and allow changes/updates.

3.2 Strategies to Improve User Awareness in MLIA and Transfer Technologies to Society

There were some key issues that were agreed as relevant factors to bring MLIA research closer to the real world. We have grouped them in two classes: general recommendations and suggestions for MLIA evaluation.

General Recommendations

- Research should start from real users and real problems. However, research should not wait for demand; because the utility for MLIA is often perceived by users a posteriori (see JRC EMM application case).
- Do rapid prototyping and showcasing to motivate user communities (see above).
- Consider tasks where named entity recognition is relevant.
- Focus translation on query translation and user interface design for query translation.
- Engage with open source communities.
- Offer shared index/search system.
- Find critical tasks and application niches. Patent retrieval as presented by John Tait is a good example. Others are tourism, education.
- Look for successful reusability experiences between sectors.

Suggestions for MLIA Evaluation

- In controlled evaluations, start with monolingual problem, then go to cross-language and see where problems arise.
- Move evaluation to realistic scenarios, and do more log analysis. Open issue is on which collections. There was a consensus that the iCLEF move into Flickr and search log analysis was a good idea.
- Try to perform live evaluations of systems (such as the QA live evaluation conducted at CLEF 2006 in Alicante).
- Compare end-to-end systems (and here there was agreement that *how* doing that is still an open issue).
- Test the same system in many different locations (again, this has already been tackled by iCLEF in the 2008 task)

Some of these points have already been taken on by the interactive CLEF track. However, the activities of this track are appreciated by the CLEF community, but still marginal in terms of participation.

3.3 Immediate Actions for CLEF

At the end of the workshop, and as a result of all presentations and debate, there emerged a few immediate actions to be taken by CLEF:

- Having representation from user communities in the CLEF steering committee. There was a proposal to invite Denis Teysou (from Agence France Press) and Bruno Pouliquen (from JRC) into CLEF steering committee. Indeed Denis Teysou has already been accepted by the steering committee meeting at their meeting in Aarhus in September.
- Having a Patent Retrieval track at CLEF. John Tait made a proposal at the committee meeting that was accepted for 2009 as a pilot track, and two representatives from patent offices attended the CLEF workshop and had a discussion session with researchers interested in the track.
- Using a 2 million visits search log from JRC as part of iCLEF. Currently this is work in progress and it won't be part of iCLEF 2009.

4 Conclusions

The Treble CLEF user communities workshop was widely regarded as a successful event by all attendees. We had the opportunity of confronting MLIA researchers with representatives from several user communities: Cultural Heritage, European government agencies, news agencies, patent and trademark professionals, enterprise and web search companies, and EU projects.

From the workshop presentations and the structured debate, consensus emerged around (i) features that any MLIA system should have from a user's perspective, and (ii) possible strategies to bring MLIA research closer to real world needs and to effectively transfer MLIA technologies to society, including ideas to improve CLEF evaluation activities. The new evaluation methodology adopted by the interactive CLEF track was strongly supported by the workshop participants, and in particular the choice of search scenario and log analysis approach.

In addition, three immediate actions were taken on board as a direct result of the workshop: inclusion of representatives from user communities in the CLEF steering committee, launch of a patent retrieval track, and contacts with JRC to use logs from a heavily multilingual search application (EMM) in CLEF evaluations.

The output from the User Communities Workshop is one of the key sources of information which will be used – together with a state-of-the-art review and other sources – to compile best practice recommendations for user-orientated evaluation (to be delivered by the TrebleCLEF consortium as part of D3.3).

Annex A: Questionnaire Distributed to Participants

(Think of Multilingual Information Access (MLIA) technologies beyond Searching facilities, i.e. Filtering, Extraction, Classification, Browsing, etc.)

1. User Communities
 - a. What user communities with Information Access needs are you familiarized with?
Chose one of these communities to answer the following questions:
 - b. Do you think that users perceive a need of MLIA technologies?
 - c. What MLIA needs are already covered or can be covered with the current MLIA technologies?
 - d. Imagine an ideal future and detect possible MLIA needs that can't be covered with current technologies.
2. Use Cases
 - a. Chose one of the user communities you know and describe a case of use with need of current MLIA technologies.
 - b. Imagine a case of use with a MLIA need that can't be covered with current technologies.
 - c. What steps or directions should be explored in order to develop such technologies?
3. User Requirements. Chose one of the use cases above and answer the following questions:
 - a. Which are the user functional requirements related to MLIA for that case? (Imposed by the task, by the user interaction, by the temporal availability of resources, by the inter-operation with other systems, etc.)
 - b. Which are the user functional MLIA requirements not covered with current technologies?
 - c. What maturity level has the MLIA technologies in order to cover these requirements?
4. User-oriented Evaluation.
 - a. Think an evaluation task that could help to bridge the MLIA limitations you have detected above.
 - b. What would you add to CLEF in order to bridge the gap between research and applications?
 - c. How would you approach MLIA user-oriented evaluation?
 - d. Enumerate some good characteristics that MLIA interfaces must have.
5. Technology transfer
 - a. Is there a demand for MLIA services? If so, please specify what kind of demand:
 - i. User communities demanding MLIA facilities [0-5]
 - ii. Software companies demanding expertise in the area of MLIA [0-5]
 - iii. Government agencies funding research in the area [0-5]Please, elaborate.
 - b. Are there current applications integrating MLIA services?
 - i. Generic commercial applications [such as Google appliance] [0-5]

- ii. Tailored applications developed specifically for a particular service (such as the US government portal search service) [0-5]
- iii. Research prototypes without a consolidated user base [0-5]

Please, specify.

- c. Is there a potential market for MLIA? Why? If the answer is yes, In which scenarios? What would be the technical requisites to make MLIA viable in that search scenarios?
- d. Which MLIA technologies have more potential to address real world needs, in your opinion? Argue.
- e. What aspects of CLEF research do you find more useful for real-world applications?

Annex B: Alphabetical list of participants

1. Amedeo Capelli (CELCT, Trento)
2. Giorgio Maria Di Nunzio (University of Padua)
3. Danilo Giampiccolo (CELCT, Trento)
4. Ana García-Serrano (UNED, Madrid)
5. Julio Gonzalo (UNED, Madrid)
6. Michael Kluck (SWP, Berlin)
7. Steven Krauwer (University of Utrecht)
8. Jenniffer Marlow (USFD, Sheffield)
9. Amar-Djalil Mezaour (Exalead, Paris)
10. Nicolas Moureau (ELDA, Paris)
11. Anselmo Peñas (UNED, Madrid)
12. Carol Peters (ISTI-CNR, Pisa)
13. Bruno Pouliquen (JRC, Ispra)
14. Maarten de Rijke (University of Amsterdam)
15. Valentín Sama (UNED, Madrid)
16. Paraic Sheridan (DCU, Dublin)
17. Joern Sieglerschmidt (Bibliotheksservice-Zentrum Baden-Wuerttemberg)
18. John Tait (Information Retrieval Facility, Vienna)
19. Denis Teysou (Agence France Press, Paris)
20. Felisa Verdejo (UNED, Madrid)
21. Dov Winer (Israel National Library, Jerusalem)