



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 6.4
MLIA Technology Transfer Day

Start Date of Project: 01 January 2008

Duration: 24 Months

Organisation Name of Lead Contractor for this Deliverable: CELCT

Final Version December 2009

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

Document Information

Deliverable number: [6.4]
Deliverable title: [MLIA Technology Transfer Day]
Due date of deliverable: [10|2009]
Actual date of deliverable: [12|2009]
Author(s): [Pamela Forner, Khalid Choukri]
Participant(s): [CELCT, ELDA]
Workpackage: [6]
Workpackage title: [Dissemination]
Workpackage leader: [CELCT]
Dissemination Level: PU (Public)
Version: [final]
Keywords: [Multilingual Information Access, Technology Transfer]

History of Versions

| Version | Date | Status | Author (Partner) | Description/Approval Level |
|---------|------------|----------|------------------|----------------------------|
| 1.00 | 21/12/2009 | Nr-final | CELCT, ELDA | Circulated for approval |
| 2.00 | 23/12/2009 | Final | CELCT, ELDA | Approved |

Abstract

This document reports the Multilingual Information Access Technology Transfer Day” held on December 8, at the World Business Alliance Center, Berlin, Germany. The event involved 20 invited speakers from various academic and commercial institutions and discussed various aspects and perspectives of Multilingual Information Access. A number of other people attended the event including local businesses and academic institutions. The workshop provided a forum in which to discuss and share experiences in the field of Multilingual Information Access. This report describes the event, the presentations given by the invited speakers and summarises discussions held.

Table of Contents

| | |
|---|----|
| Document Information | 1 |
| Abstract | 1 |
| Executive Summary | 3 |
| 1 Introduction | 4 |
| 2 Workshop Details | 4 |
| 3 Presentations..... | 5 |
| 4 Demo Session..... | 6 |
| 5 Discussion Panel..... | 7 |
| 6 Final summary – Wrap up and Take away | 9 |
| Appendix A – Attendees (random order) | 10 |
| Appendix B – Programme | 13 |

Executive Summary

This document reports the Multilingual Information Access Technology Transfer Day” held on December 8, at the World Business Alliance Center, Berlin, Germany. The event involved 20 invited speakers from various academic and commercial institutions and discussed various aspects and perspectives of Multilingual Information Access. A number of other people attended the event including local businesses and academic institutions. The workshop provided a forum in which to discuss and share experiences in the field of Multilingual Information Access. This report describes the event, the presentations given by the invited speakers and summarises discussions held.

This report describes the organisation of the event which ran in Berlin on December, 8, the invited speakers and their presentations, a summary of discussions and reflections on the event.

1 Introduction

The primary goal of TrebleCLEF is the dissemination of knowledge and spreading of excellence in the domain of Multilingual Information Access (MLIA), by creating a community of researchers and developers and facilitating future collaborative initiatives between them. Its main objective is to support the development and consolidation of expertise in the multidisciplinary research area of Multilingual Information Access and to promote a dissemination action in the relevant application communities.

To this end a MLIA Technology Transfer Day was organized which was aimed at putting together researchers developers and users in order to facilitate a two-way communication of research advances and results, as well as the reverse flow of requirements and ideas from users and developers. Another goal was also to increase the awareness of the importance of MLIA in the information society, in both academic and industrial subjects at an international level.

The final goal was to stimulate the discussion on MLIA technologies and provide a forum in which invited speakers could share their experiences, identify common strategies and needs, and detect future challenges in this field.

The remainder of this deliverable is structured as follows: the setup of the event (audience, goals and audience; Section 2) is presented first, followed by a summary of presentations given by attendees at the workshop (Section 3), an outline of the Demo Session (Section 4), a summary of the discussion panel (Section 5), and finally a short conclusion (Section 6).

2 Workshop Details

The aim of the MLIA Technology Transfer Day is to put researchers, software developers and users together in order to facilitate a two-way communication of research advances and results by researchers, as well as the re-verse flow of requirements and ideas from users and developers.

Another goal was also to increase the awareness of the importance of MLIA in the information society, in both academic and industrial subjects at an international level.

The final goal was to stimulate the discussion on MLIA technologies and provide a forum in which invited speakers could share their experiences, identify common strategies and needs, and detect future challenges in this field.

The Multilingual Information Access Technology Transfer Day was held at the World Business Alliance Center in Berlin, on December 8. The event was jointly organized by CELCT (Center for Evaluation of Language and Communication Technologies, (<http://www.celct.it/>)) and ELDA (Evaluations and Language Resources Distribution Agency, (<http://www.elda.org/>)), while DFKI offered its support for the local organization.

The event was widely publicized in the major mailing lists.

Participation in the event was free of charge but subject to prior registration

The one day event involved 20 speakers (including 3 invited talks, 4 presentations of the Best Practices Recommendations, 9 demo presenters, and 3 panelists) coming from academic and commercial institutions. In addition, a certain number of participants (more or less 40 in total) attended the event including local based companies (i.e. ORTEC , vionto) and academic institution (i.e. U. Chemnitz, U. Hildesheim).

By involving representatives from different communities we aimed at stimulating the transfer of knowledge and ideas. By inviting academics we aimed at clarifying current research.

The slides of the Presentations will be made available and a video of all the talks of the event is being prepared and will be downloadable at the website created for the event (<http://celct.isti.cnr.it/MLIA-Day/index.html>).

3 Presentations

The event began with an introductory talk by Carol Peters in which the main objectives of the TrebleCLEF project were presented.

This was followed by three keynote speeches from the business and application communities invited to present their experiences, requirements and problems with respect to Multilingual Information Access:

1. Benefits of Multilingual Media Monitoring - Ralf Steinberger, Language Technology Project Manager at the Joint Research Centre (JRC). Steinberger described the media monitoring software developed by JRC which gathers and analyses between 80,000 and 100,000 news articles per day in up to fifty languages. The software clusters related news, classifies the articles into hundreds of categories, extracts information from them, fuses information found in different languages, detects breaking news, and tracks topics over time and across languages.

The users of this software are European Union (EU) institutions, EU member state organisations, international organisations and other selected partners around the world. Additionally, an average of 30,000 anonymous internet users per day make use of the public systems.

Special attention was given also to the multilingual and cross-lingual functionality of the various tools and on the different user requirements.

2. Multilingual Discovery in Europeana - Sjoerd Siebinga, Senior Developer at Europeana Office. His talk may be summarized in the following points:

Multilingual Information Access is not difficult, just very expensive

Not development cost, but maintenance cost are the real bottleneck

Multilingual Information Access is all about resources and little about technology

Be extremely explicit in what you are going to offer and be certain that the user really wants it, i.e. user-needs should drive business requirements

User Interface and Interaction Design is more important than the number of features you offer.

3. Industrial Needs for Multilingual Information Access and Available Resources - Gregory Grefenstette, Chief Science Officer at Exalead

Grefenstette introduced Exalead as a global software provider in the enterprise and Web search markets, which tries to integrate research functions in one piece of software, combining Internet, desktop and enterprise functionality. The firm's technology makes it possible to integrate structured and unstructured content in a unique way to address critical applications in areas such as extended business intelligence, customer support, compliance, and many others.

Exalead attempts to add value to search results by displaying statistics and information about the results along the left-hand side of the page. This includes categorical and geographical information.

Exalead's features include:

- Proximity search supporting the NEAR operator
- Nested search expressions
- Autocompletion searches

- Faceted search
 - o by geographic location
 - o by language (54 languages recognized)
 - o by file type
 - o by DMOZ categories
- Related terms suggestions
- Phonetic search is also available, which can be used when the orthography is not known
- Search within results
- Audio, video and RSS search
- Search in Wikipedia

Grefenstette also addresses some of the reasons why applications do not use research results, i.e. licensing, API, maintenance.

The day provided also a forum in which to discuss and share experiences and best practices in the MLIA field, as the event also included the presentation of a set of Best Practice Recommendations that were produced by TrebleCLEF in the areas of MLIA System Development and Search Assistance, Test Collection Creation, Evaluation Methodologies and Language Processing Technologies.

All the information about TrebleCLEF Best Practices have been made available to all attendees coming from both the scientific and application communities. The TrebleCLEF Best Practices Report can be accessed also via the T-CLEF website.

4 Demo Session

In addition to the presentations, the event included also a demo session during the lunch break where attendees were given the opportunity to view a number of prototype MLIA systems and components. It resulted in a very lively demo session with nine technical demonstrations which were meant to allow researchers and practitioners that have implemented software systems to describe the design, the components, the development and functionalities of their work in an interactive setting.

- Jens Kürsten from the Chemnitz University of Technology presented the Xtrieval Framework. It consists of three functional component: the indexer, the search engine and the evaluation toolkit.
- Alessio Bosca from CELI showed the CACAO interface which offers an innovative approach for accessing, understanding and navigating multilingual textual content in digital libraries and Online Public Access Catalogues (OPACs).
- Stéphane Clinchant from Xerox Research Center illustrated his system doing translations of Wikipedia pages.
- Franca Debole from ISTI-CNR described the MultiMatch Web search engine which is specialized in the cultural heritage domain. Queries can be expressed in multiple languages and multilingual and multimedia retrieval is possible. It allows access to multiple sources of information (e.g. museums, cultural institutions, CH educational sites, tourist information portals).
- Paul Libbrecht from DFKI gave an idea of the cross-curriculum search on the I2Geo Platform. The main objective is to make digital content for mathematics teaching in Europe more accessible, usable and exploitable.
- Carlos Amaral from Priberam gave some details about his QA cross-lingual system searching news in many languages.
- Irja Wichert from the German Exalead group showed some of the features that have been deployed by some of their customers in SBA projects. They have tried to split that over several areas, like queries, structuring, reconciliation, analytics with some examples for each parts.

- Mikołaj Wypych from POLENG explained the functionalities of the Globalizer system which gathers together Machine Translation, Computer Aided Translation and Enterprise Resource Planning. The system almost fully organizes the translation bureau work-flow. What is quite unique is that the Globalizer integrates with many popular Content Management Systems, allowing for automatic placing of orders.
- Leonoor van der Beek from Q-go described their web self-service for Multinationals. The core of Q-go's technology is an engine capable of recognizing the endless variations of natural language and matching individual user questions to predefined model or concept questions that form an editorial index of content. This allows the search engine to return a limited number of highly relevant results, obtaining higher click through rates.

5 Discussion Panel

Khalid Choukri introduced the panel that was entitled: "Cross-Lingual Information access, retrieval, extraction, needs, requirements and current state of the art"; and then introduced the topics and experts with complementary experience and interests were invited to ensure a stimulating exchange of ideas a lively discussion and audience participation. The panellists were:

- Allan Hanbury, Senior Scientist, Information Retrieval Facility (IRF), Vienna
- Leonoor van der Beek; Research Manager at Q-go, The Netherlands
- Jochen L. Leidner, Research Scientist, Thomson Reuters Corporation, Corporate Technology, Research & Development, USA and also CEO of Linguist, Dublin, Ireland

The main topics that were addressed were the following:

- a) The needs of large users of multilingual information archives and their requirements, expectations
- b) The current offers for enterprise search and web
- c) What is preventing the take off of cross-lingual search applications

The theme of the panel discussion was why, despite 10 years of MLIA evaluations (within TREC, CLEF, NTCIR programmes), the adoption of such technologies is still a questionable issue.

This session was organized around a series of questions which were circulated to the panelist beforehand. The list of questions was as follows:

- During the last ten years or so, the number of languages and domains, for which research and development on MLIA have been conducted, has increased massively; is there any killer application that emerged from such clusters?
- Have we managed to assess performance and are these in line with application deployments; are these two concepts strongly related (performance, deployments)?
- Did the evaluations consider the right needs and requirements, and by the way do we have a clear picture of these?
- If no success story can be mentioned today, does that mean key players were left out of these evaluations? Why?
- If players were among the participants, does that mean we do have applications/core-technology with performances that can not be offered to real users?
- And what offers do MLIA players have that one can highlight today, in particular for use by professionals?
- Can we imagine fulfilling the needs of all languages and all domains at a reasonable cost (money, time, ...)?
- How can the panelists (and the participants) address these issues and establish pointers to high levels of cooperation between users, service providers and core-technology providers (including computational linguistics)?

Each of the three panelist had about 10 minutes for a focussed presentation but due to the time constraints not all issues and questions were tackled.

- Allan Hanbury focussed on concrete cases that involved actors of very highly specialised domains: Patent Searchers, Biomedical Researchers, and Medical Doctors. It is clear from these cases that “business” imposes different weights on different languages (e.g. French and German translation not so important – Chinese, Japanese and Korean are), but globally Multi-linguality is crucial. Translation is a crucial component (need to translate up to nuances) and not simply “have an idea of the content”. Scientific publications and patents are critical and hard to access to when drafted in Chinese, Russian, etc while most researchers are proficient at reading English. For the other categories (e.g. General medical practitioners (GPs)) as well as the large public, they are potential user of multi-lingual search systems though today they do prefer monolingual search and searching sources available in their native language and for each category with its appropriate jargon.
- Leonoor (Leonoor van der Beek, Manager Research , Q-go, Netherlands), introduced the topic through Q-Go activities and highlighted the need for a user to get a small number of results, with the most relevant results that are (in the case of Q-Go) presented as questions, which are linked to answers. She emphasised that the need for multi-lingual search is mostly based on the use of parallel content management with Parallel (domain) dictionaries that have to be developed (and the importance of their costs). According to Leonoor, such approaches, that require cooperation of IR and NLP communities, show the mismatch between their focuses. She also made a strong statement about the Cross-lingual website search: **there is no demand for it and no one ready to pay for the deployment of an application**. She insisted that data on usability, use cases, success stories, are urgently needed though it is hard to envisage sharing this within the community. MT for Cross-lingual website search has to be perfect as one deals with marketing issues that require translation, localization, etc.

Leonoor also elaborated on the Website search vs. enterprise search; she highlighted that for enterprise search, there is more data, more multilingual users and less marketing considerations and therefore some players do offer cross lingual search engines.

She concluded that in many case the critical issue is the lack of Language Resources.

- Jochen Leidner introduced his speech through controversial statements such as : “academics make problems instead of asking question” but then focussed on major issues such as how to boost technology transfer through a share of the risks and how to design evaluations that considers real requirements of users. Another important issue pointed out by Joschen is the “cost of Language Resources “ (what is the right cost), and in particular how much are companies (large and small) willing and able to pay for language resources? He gave the example of LDC commercial license (\$21,000.-/year and insisted that “it’s cheaper for a company to re-tag part of the Wall Street Journal internally”. Should language resources (like other data resources) perhaps even be free, at least if curated from taxpayers’ money? Availability of resources in sets of language is important, e.g. for all EU languagesbut then Licenses need to be industry-viable and easy to find on resource catalog Websites.

During the debate, some of the discussants elaborated on various aspects of multi-linguality and cross-linguality. A brief summary is given herein:

- Language Resources constituted an important discussion topic. The issue of sharing data was discussed with the views of the R&D community compared to the views of the industry. The first group focuses on quality and availability as the most important features while the second one insists of IPR clearing and easy licensing. Some participants felt that the evaluation initiatives work on unrealistic scenarios and that for many applications, a performance of 80% means simply that no operational application can be derive from such technology.
- Language Resources costs when integrated within a given technology were also discussed. Khalid Choukri indicated that, from past ELRA surveys, applications such as speech

transcription, machine translation, spoken dialogues, Language Resources contribute with 10%-20% of the total costs of an application. Some indicated that in MLIA this should be rather close to 80% (probably for porting existing technology to another domain /language but not for the development of the core technology). Particular resources such as Named Entities, that felt as very crucial by industry, seem not to attract enough R&D attention.

- The need for CLIR application was also extensively debated by the participants with a clear controversial discussion on real needs vs artificial research scenarios. Connections between research and industry should be strengthened to ensure smooth transition from R&D to applications;
- The involvement of open source community was well advocated for and hopefully this is already happening through the number of systems, components, etc. that are offered in particular by the CLEF participants.

As a conclusion, Khalid Choukri, insisted on the need for visions for the next 2010-2015 period but also for the strong need for media coverage to sell the MLIA to the decision makers (role of public agencies procurements). He also rephrased many participants statements and/or question that “**even if all is not about Language Resources, these play more than a crucial role within MLIA**”.

The slides of the Presentations will be made available and a video of all the talks of the event is being prepared and will be downloadable at the website created for the event (<http://celct.isti.cnr.it/MLIA-Day/index.html>)

6 Final summary – Wrap up and Take away

- The major obstacle to technology transfer is very often ignorance of state of the art and the weak link between R&D and Industry. Some of this is due to organizational problems, many are related to legal issues.
- “All is about Language Resources”: many experts stated that a major factor that hinders business efficiency is the availability of adequate Language Resources. Unfortunately, very often, such resources exist but are hard to identify and harder to license (strong statement: MLIA is about LRs, little about technology”).
- MLIA (research, evaluation, applications) should be extended to other modalities than text (speech/audio, Images, Video, combined modalities, sign languages, etc.) but also to other languages than the usual CLEF/TREC/NTCIR/FIRE ones.
- The CLEF Best practices have to be interpreted as a multiple dimension ones: (a) resources for baseline(s) where a lot is available for some major languages but not a lot for application deployment; (b) the problems of IPR that require clean agreements with resource providers and MLIA developers (unless one crawls data from the Internet without caring about infringement of IPRs); (c) need for a centralized center to access licensable/high-quality LRs is still important; (d) the importance of turning some of the “news” into Language Resources;
- One should be conscious that most of the users are MONOLINGUAL and many consider MT as a must and not as an option; many rely on the Google-translate as the basic feature for MLIA and Translation.
- “Solved problems” is often an assumption of the research community that is not always valid when it comes to deployment (scalability, sustainability, multiple languages, language evolutions, and many user oriented constraints).

Some recommendations from the Treble-CLEF best practices and the MLIA debates:

- Language Resource packages have been designed to be used as a Basic Language Resource Kit (BLARK) for MLIA and are available to all.
- Major outcomes of the Treble-CLEF work are recorded in the various documents (blueprints), covering system designs, Language Resources, evaluation methodologies, results presentations, etc.
- “All is about user satisfaction”, both interfaces and content summaries, good quality of MT (and metadata), etc.
- Need to go beyond the current resources and incorporate more “semantics” and other hot features (ontologies, sentiment analysis, web 2.0 resources, etc.), linked to real world resources (wikis, linkedData, world Gumbo
- Technology transfer suffers more from the complexity of licensing from R&D labs than from the “non-mature” feeling about technology (complex, slow, legally binding, royalties very high, etc.)
- Very often APIs are missing or not engineered by the right skilled people (API is an engineering problem , not a research topic);
- Technology transfer requires maintenance and long term partnership; not a core concern for most of the R&D labs;
- Encourage the R&D and HLT community to focus on the Basic Language Resources such as word lists as a community cooperative work and provide it as a gift to all interested parties, for as many languages as possible;
-

Appendix A – Attendees (random order)

Complete list of participants (including speakers, demo presenters and panelist)

| | Name | Affiliation |
|----|----------------------|----------------------|
| 1 | Emanuele Pianta | CELECT - Italy |
| 2 | Carol Peters | ISTI-CNR -Italy |
| 3 | Julio Gonzalo | UNED - Spain |
| 4 | Martin Braschler | ZHAW - Switzerland |
| 5 | Khalid Choukri | ELDA - France |
| 6 | Nicolas Moreau | ELDA - France |
| 7 | Paul Clough | U. Sheffield - UK |
| 8 | Gregory Grefenstette | Exalead - France |
| 9 | Ralf Steinbergher | JRC - Italy |
| 10 | Sjoerd Siebienga | Europeana |
| 11 | Alessio Bosca | CELI - Italy |
| 12 | Jens Kuersten | U. Chemnitz, Germany |
| 13 | Stephane Clinchant | Xerox - France |
| 14 | Paul Libbrecht | DFKI - Germany |
| 15 | Franca Debole | ISTI-CNR - Italy |
| 16 | Carlos Amaral | Priberam - Portugal |
| 17 | Irja Wichert | Exalead - Germany |
| 18 | Mikołaj Wypych | Globalizer - Poland |

| | | |
|----|-------------------------|--|
| 19 | Jochen Leidner | Thomson-Reuters |
| 20 | Leonoor van der Beek | Q-GO- the Netherlands |
| 21 | Allan Hanbury | IRF – Austria |
| 22 | Pamela Forner | CELCT- Italy |
| 23 | Giovanni Moretti | CELCT - Italy |
| 24 | Thomas Wilhelm | U. Chemnitz - Germany |
| 25 | Maximilian Eibl | U. Chemnitz - Germany |
| 26 | Avarè Stewart | L3S Research Center - Germany |
| 27 | Thomas Mandl | U. Hildesheim - Germany |
| 28 | Vivien Petras | U. Berlin – Germany |
| 29 | Francesca Borri | ISTI-CNR - Italy |
| 30 | Christa Womser-Hacke | U. Hildesheim - Germany |
| 31 | Georg Rehm | vionto GmbH, - Germany |
| 32 | Maria Gaede | U. Berlin - Germany |
| 33 | Juliane Stiller | U. Berlin - Germany |
| 34 | Giorgio Maria Di Nunzio | U. Padua, Italy |
| 35 | Marco Dussin | U. Padua, Italy |
| 36 | Alexandre Khan | WallMuse |
| 37 | Adedotun Kola Sijuade | VITAL CLICKS SOLUTION |
| 38 | Michael Kluck | SWP - Germany |
| 39 | Jan Odijk | U. Utrecht - the Netherlands |
| 40 | Dirk Beerbohm | Exalead - Germany |
| 41 | Xiwen Cheng | YocoY technologies - Germany |
| 42 | Marie-Véronique LEROI | French Ministry of Culture and Communication |
| 43 | Matthias Ort | ORTEC GmbH |
| 44 | Joacim Zschimmer | ORTEC GmbH |

Registered but unable to come:

| | | |
|---|---------------------------|--|
| 1 | Blanca Rodriguez-Bravo | U. Leon - Spain |
| 2 | Jolanta Mizera-Pietraszko | U. Wroclaw - Poland |
| 3 | Patrice Lopez | European Patent Office |
| 4 | Anastasios Patrikakos | Institute for Language and Speech Processing, Greece |
| 5 | OSAKPAMWAN OSARUMWENSE | SEQUENCE GLOBAL COMMUNICATIONS – South Africa |
| 6 | Ernesto William De Luca | University Berlin -Germany |
| 7 | KRISHNA PARAJULI | TRIBHUVAN UNIVERSITY, Depart. Of |

| | | |
|----|---------------------|----------------------|
| | | Linguistics |
| 8 | Manfred Stede | U. Potsdam - Germany |
| 9 | Valia Kordoni | DFKI - Germany |
| 10 | Peter Adolphs | DFKI, Germany |
| 11 | Tina Klüwer | DFKI - Germany |
| 12 | Nicolaas Bongaerts | DFKI - Germany |
| 13 | ahmed alsow mohamud | kaah enterprises co. |

Appendix B – Programme

| | |  |
|---------------|---|---|
| 9:00 - 9:20 | Registration | |
| 9:20 - 9:30 | Welcome, Emanuele Pianta, CELCT, Italy | |
| 9:30 - 9:50 | Introduction to TrebleCLEF, Carol Peters, ISTI-CNR, Italy | |
| 9:50 - 10:30 | Invited talk: Benefits of Multilingual Media Monitoring, Ralf Steinberger, Joint Research Centre, Italy | |
| 10:30 - 11:10 | Invited talk: Multilingual Discovery in Europeana, Sjoerd Siebinga, Europeana Office | |
| 11:10 - 11:30 | Coffee Break | |
| 11:30 - 12:30 | Treble-CLEF Best Practice Recommendations <ul style="list-style-type: none"> • Best Practices in Language Resources for MLIA, Khalid Choukri and Nicolas Moreau, ELDA, France • Best Practices in System-oriented MLIA, Martin Braschler, Zurich University of Applied Science, Switzerland • Best Practices in User-oriented MLIA, Julio Gonzalo, UNED, Spain • Best Practices for Test Collection Creation, Evaluation Methodologies and Language Processing Technologies, Paul Clough, University of Sheffield, United Kingdom | |
| 12:30 - 12:45 | Discussion | |
| 12:45 - 13:00 | DEMO Boaster Session | |
| 13:00 - 15:30 | Buffet Lunch and DEMO session <ul style="list-style-type: none"> • Extensible Retrieval and Evaluation Framework: Xtrieval, Jens Kürsten, Dept. Computer Science and Media, Chemnitz University of Technology, Germany • CACAO: A Multilingual Interface to Library Catalogs, Alessio Bosca, CELI, Italy • Cross Lingual Search on Wikipedia, Stephane Clinchant, Xerox Research Center, France • MultiMatch: Access to Cultural Heritage in a MultiModal Way, Franca Debole, ISTI-CNR, Italy • Cross-Curriculum Search on the I2Geo Platform, Paul Libbrecht, Competence Center for E-Learning, German Center for Research in Artificial Intelligence, DFKI GmbH, Germany • Multilingual News Retrieval with Priberam Search, Carlos Amaral, Priberam, Portugal • From Disparate Data to Meaningful Building Blocks of Knowledge, Irja Wichert, Exalead, Germany • globalizer.eu: Reducing Costs in CMS Translations, Mikołaj Wypych, POLENG, Ltd., Poland • Q-go: Web Self-service for Multinationals, Leonoor van der Beek, Q-go, the Netherlands | |
| 15:30 - 16:10 | Invited talk: Industrial Needs for Multilingual Information Access and Available Resources, Gregory Grefenstette, Exalead, France | |
| 16:10 - 17:10 | Discussion Panel - <i>Moderator: Khalid Choukri</i> <i>Jochen Leidner, Allan Hanbury, Leonoor van der Beek</i> | |
| 17:10 - 17:30 | Conclusions and Wrap Up, Khalid Choukri, ELDA, France |    |