



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 6.2

TrebleCLEF Summer School on Multilingual Information Access

Start Date of Project: 01 January 2008

Duration: 24 Months

Organisation Name of Lead Contractor for this Deliverable: ISTI-CNR

Version Final, June 2009

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

Document Information

Deliverable number: 6.2
Deliverable title: TrebleCLEF Summer School on Multilingual Information Access
Actual date of deliverable: 30/06/02009
Author(s): Carol Peters & Francesca Borri, ISTI-CNR
Participant(s): All Consortium members
Workpackage: 6
Workpackage title: Dissemination
Workpackage leader: CELCT
Dissemination Level: Public
Version: 1.00
Keywords: Multilingual Information Access, Cross Language Information Retrieval,

History of Versions

Version	Date	Status	Author (Partner)	Description/Approval Level
1.00	30/06/2009	Final	ISTI-CNR	OK

Abstract

The organisation of the TrebleCLEF Summer School on Multilingual Information Access is described. Details are given concerning the lecturers, their course, the participants, and the logistics. The opinion of the students is presented via a breakdown of the Evaluation forms.

Table of Contents

Document Information	1
Abstract	1
Executive Summary	3
1 Objectives.....	4
1.1 Location & Dates	4
1.2 Social Events.....	4
1.3 Registration Fee	4
1.4 Programme Committee	4
2 Lecturers.....	5
3 Programme of Lectures	8
4 Participants.....	13
4.1 Grant Assignments.....	13
4.2 List of Participants	13
5 Evaluation.....	18

Executive Summary

The aim of the TrebleCLEF Summer School, held in Pisa 15-19 June 2009, was to give participants a grounding in the core topics that constitute the multidisciplinary area of Multilingual Information Access (MLIA). Both theoretical and practical issues were addressed. The focus of the school was on "How to build effective multilingual information retrieval systems and how to evaluate them". The School was intended for advanced undergraduate and post-graduate students, post-doctorial researchers plus academic and industrial researchers and system developers with backgrounds in Computer Science, Information Science, Language Technologies and related areas.

The lecturers at the School were experts in the different sectors that form part of the MLIA domain. Many of the lecturers were also members of the TrebleCLEF consortium, all have been strongly involved in the organisation of the Cross Language Evaluation Forum. The programme was studied to begin with a basic introductory course on multilingual and cross-language retrieval on the first day,. This was followed by a series of lessons on various aspects of the core problem: how to enable multilingual access functionality for diverse types of requirements, and media. Issues covered included multilingual architectures and interfaces, multimedia retrieval, evaluation and applications in a multilingual context. The final lecture was on commercial aspects of cross-language search. An optional mentoring session was also held.

This document describes the organisation of the School. Details are given concerning the lecturers, their course, the participants, and the logistics. The opinion of the students is presented via a breakdown of the Evaluation forms. An additional Appendix provides copies of all slides presented by the lecturers.

1 Objectives

The aim of the Summer School was to give participants a grounding in the core topics that constitute the multidisciplinary area of Multilingual Information Access (MLIA). Both theoretical and practical issues were addressed. The focus of the school was on "How to build effective multilingual information retrieval systems and how to evaluate them".

1.1 Location & Dates

The Summer School was held 15 - 19 June 2009 in the beautiful ex-convent Santa Croce in Fossabanda, Pisa. Santa Croce provides the perfect setting for study and discussions in a peaceful, relaxed atmosphere and is just a short walk from the town centre and the famous Piazza dei Miracoli with its Leaning Tower.

1.2 Social Events

The following events were organised:

- Sunday, 14 June, 18.00 - Reception in the cloisters at Santa Croce in Fossabanda
- Monday, 15 June - Social Dinner at a typical Tuscan country restaurant in the Pisan hills
- Thursday, 18 June, 19.00 – Boat trip on the River Arno with Guide

In addition, the dates of the Summer School were chosen to coincide with the festivities organised by the town to honour the patron saint of Pisa: San Ranieri. The celebrations were held over 2 days, the 16th and 17th of June. The 16th was dedicated to La Luminara di San Ranieri, where thousands of candles are lit up in the buildings along the river and a fireworks display was held. On the following day, there was a Regatta of the Historical Quarters of Pisa and an all day street market all through the city. The social events were much appreciated by the students. The social dinner was organised early – on the evening of the first day of the School – in order to encourage the students to get to know each other and the lecturers (approx. half were present).

1.3 Registration Fee

The School registration fee was 200 euros. Registration covered lectures, coffee breaks, lunches, course material (provided on USB keys and, on request, as printed handouts), plus all social events.

1.4 Programme Committee

The programme committee consisted of one member from each partner in the TrebleCLEF consortium plus members of the TrebleCLEF Advisory Board. The organisation and programme of the School was discussed at the Management Meetings in Aarhus, September 2008 and Paris, January 2009.

Carol Peters (Chair), National Research Council, Italy
Maristella Agosti, University of Padua, Italy
Martin Braschler, Zurich University of Applied Sciences, Switzerland
Amedeo Cappelli, CELCT, Italy
Khalid Choukri, ELRA/ELDA, France
Christian Fluhr, NewPhenix, France
Stefan Gradmann, University of Humboldt, Germany
Donna Harman, National Institute of Standards and Technology, USA
Noriko Kando, National Institute of Informatics, Japan
Mark Sanderson, University of Sheffield, UK
Costantino Thanos, National Research Council, Italy
Felisa Verdejo, Universidad Nacional de Educación a Distancia, Spain
Claire Warwick, University College, London, UK

2 Lecturers

The lecturers at the School are experts in the different sectors that form part of the MLIA domain. Many of the lecturers were also members of the TrebleCLEF consortium, all have been strongly involved in the organisation of the Cross Language Evaluation Forum. The School website contains photos and details on all the lecturers. Here below we list the lecturers with brief CVs. The lecturers stayed a varying amount of time at the School, depending on their other commitments: Giorgio Di Nunzio, Anselmo Peñas and Franca Debole were present throughout the week.

Martin Braschler holds the title of Professor ZFH and a position as lecturer at the Zurich University of Applied Sciences in Winterthur. He studied computer science at the Swiss Federal Institute of Technology ETH in Zurich, leading to a degree of MSc ETH, and received the degree of Dr. sc. from University of Neuchatel in 2004. He has been active in IR research since 1997, with an emphasis on cross-language IR. A stay as visiting researcher at the National Institute of Standards and Technology NIST was the beginning of his involvement with CLIR evaluation, and he was responsible for the technical coordination of the core activities of the Cross Language Evaluation Forum CLEF in its early years. Aside from his academic record, he formerly was head of research and innovation at Eurospider Information Technology AG, a Swiss IR solution provider that develops cutting-edge IR technology for commercial applications.

Paul Clough is a Lecturer in Information Systems at the University of Sheffield. He is a member of the Information Retrieval Group and the Database Group. His main interests are Multimedia Information Retrieval (IR), Geographic IR, Multilingual IR and using Natural Language Processing (NLP) technologies within IR applications. Paul is currently PI in three EU-funded projects: MultiMatch, Memoir and TrebleCLEF. The MultiMatch project aims to develop a multilingual search engine for accessing multimedia cultural heritage material. The Memoir project is investigating the use of technologies to improve information access of our personal memories. The TrebleCLEF project aims to promote evaluation, best practice and collaboration for multilingual information access. Recently, Paul has been an overall coordinator of an international evaluation campaign for image retrieval called ImageCLEF. Paul has published over 60 papers in refereed international journals and conferences on a variety of topics including information access, text reuse and evaluation. [<http://ir.shef.ac.uk/cloughie/>]

Franca Debole is a researcher at the Institute of Information Science and Technologies one of the institute of the Italian National Research Council in Pisa, Italy. Her main research interests include text categorization and information retrieval; XML data structure; search engine technologies and multilingual information retrieval. In the last three years she was involved in a EU-funded projects MultiMatch project aims to develop a multilingual search engine for accessing CH material. At the moment she is working on a eContent plus project the European Film Gateway.

Maarten de Rijke is full professor of Information Processing and Internet in the Informatics Institute at the University of Amsterdam. He holds MSc degrees in Philosophy and Mathematics (both cum laude), and a PhD in Theoretical Computer Science. He worked as a postdoc at CWI, before becoming a Warwick Research Fellow at the University of Warwick, UK. He joined the University of Amsterdam in 1998, and was appointed full professor in 2004. He leads the Information and Language Processing Systems group. While relatively young, this group has rapidly established itself as one of the leading academic research groups in information retrieval in Europe. His current focus is on intelligent web information access, with projects on search and discovery for informal text, vertical search engines, question answering, weakly or semi-structured documents, and multilingual information. He currently holds one of the prestigious Pionier grants, has published over 400 papers, has published or edited over a dozen books, is editor for various journals and book series, and until recently coordinated the evaluation efforts of the Crosslingual Web Track at CLEF.

Thomas Deselaers is a postdoc at the Computer Vision Laboratory of the ETH Zürich. He received his diploma and his PhD degree from RWTH Aachen University in Aachen, Germany in 2004 and 2008, respectively. From March 2004 to December 2008 he was a full time researcher at the Human Language Processing and Pattern Recognition Group of the Computer Science Department, where he has been the head of the image processing and understanding group since 2005. In 2002 he was a visiting student researcher at the Instituto Tecnológico de Informática at the Universidad Politécnica de Valencia, Spain, in 2006 he was a research intern at Microsoft Research Cambridge, UK, and in 2008 he was visiting researcher at the Departamento de Sistemas Informáticos y Computación at the Universidad Politécnica de Valencia, Spain. Thomas is vice chair of the IAPR technical committee 5 and Co-organizer of the ImageCLEF image retrieval evaluation. His research interests are object classification and detection in complex scenes, content-based image retrieval, and pattern recognition.

Giorgio Maria Di Nunzio is assistant professor at the Department of Information Engineering of the University of Padua, Italy. His main research interests are probabilistic models and visualization approaches for automated text categorization and information retrieval; evaluation of information access systems; log analysis of digital library systems for personalization of contents. He is also interested in the area of Multilingual Information Retrieval and in particular in the study of new strategies to compensate for limited linguistic resources for multi-language retrieval tasks.

Marcello Federico received the Laurea degree in Computer Science from University of Milan in 1987. After he joined ITC-irst in Trento (Italy) where he has been a permanent researcher since 1991. He is currently co-head of the Human Language Technology research unit at Fondazione B. Kessler, formerly ITC-irst, and consulting professor at the University of Trento. His research interests include statistical machine translation, spoken language translation, statistical language modelling, information retrieval, and speech recognition. His group has participated in several evaluation campaigns on cross-language IR (CLEF) and MT (NIST, IWSLT) and has contributed to the developed of popular open source software, for language modelling (IRSTLM) and statistical machine translation (Moses).

Nicola Ferro is assistant professor in computer science at the Department of Information Engineering of the University of Padua, Italy. His main research interests are digital libraries and archives, their architectures and evaluation, and multilingual information access and its evaluation. He participates in national and international projects: SAPIR which deals with search in audio visual content using peer-to-peer information retrieval; TELplus which concerns the enhancement of The European Library portal towards Europeana; TrebleCLEF which aims at providing best practices, collaboration, and evaluation for multilingual information access systems. He is involved in the coordination of the CLEF evaluation campaigns, and is co-organizing the Ad Hoc and Grid@CLEF tracks. He has published more than 50 papers on digital library architectures, interoperability, and services; multilingual information access and its experimental evaluation; the management of the scientific data produced during evaluation campaigns.

Gareth J. F. Jones is a faculty member in the School of Computing, Dublin City University, Ireland, where he is an Investigator in the Centre for Digital Video Processing and the SFI CSET Centre for Next Generation Localisation. He has previously held positions at the University of Cambridge and University of Exeter, U.K., as a Toshiba Fellow at the Toshiba Corporation Research and Development Center in Kawasaki, Japan, and as a visiting researcher with the Informedia project at Carnegie Mellon University, U.S.A and a JSPS Visiting Fellow at the National Institute of Informatics, Toyko, Japan. He was awarded BEng and PhD degrees from the University of Bristol, U.K. His research is focused on topics in information retrieval. These include multilingual and cross-lingual IR, multimedia IR, speech retrieval, context-aware IR, and personal information search. His research is currently supported by Science Foundation Ireland (SFI) and the European Commission.

Henning Müller studied medical informatics at the University of Heidelberg from 1992-1997 with a specialisation in signal and image processing. After a diploma thesis in the telemedicine project Chile he worked for six months at Daimler-Benz research and technology North America in Portland, OR, USA with a scholarship from the Carl Duisberg Society. From 1998-2002 he was a Ph.D student at the

University of Geneva with a research stay at Monash University in Melbourne, Australia in 2001. He received his Ph.D in content-based image retrieval in 2002. Since 2002 he has been working at the Medical Informatics Service at the University Hospitals and the University of Geneva, where he started the medical image retrieval project medGIFT and initiated the medical image retrieval benchmark ImageCLEFmed. Since 2007 he is a professor at the University of Applied Sciences Western Switzerland in Sierre, while keeping a part-time research position in Medical Informatics in Geneva.

Anselmo Peñas is Associate Professor in the School of Computing, Distance Learning University of Spain, where he is researcher in the UNED NLP & IR Group since 1998. He obtained the PhD degree with special distinction and award in 2002 and also holds the Award of the Spanish Society for the Natural Language Processing. He has participated in several European projects such as EuroWordNet, CLEF and TrebleCLEF projects, News Engine Web Services project (NEWS), or the European Schools Treasury Browser project (ETB). A. Peñas has a book on Linguistic Techniques applied to Information Retrieval and more than 60 research papers on this topic, with special interest in multilingualism, interactivity and evaluation. Currently, he coordinates the Question Answering Track at CLEF.

Mark Sanderson is a Reader in Information Retrieval (IR) at the University of Sheffield. His research interests are in the usability of searching systems, image retrieval, searching across different languages and evaluation of searching systems. Mark has raised over £4 million in research grant income to study these research areas and published his work extensively. A number of these publications have been cited over 100 times. He sits on the editorial board of four the major IR journals and is currently the programme co-chair of ACM SIGIR, the premier conference in the field. Along with Paul Clough, Mark was the co-founder of the highly successful imageCLEF track of CLEF. He taught tutorials on IR at ACM SIGIR in 2000 and 2001, for many year's on a Master course at "Faculdade de Engenharia da Universidade do Porto" and at IR summer schools in Glasgow and Sydney.

Jacques Savoy is full Professor in Computer Science at University of Neuchatel (Switzerland). J. Savoy received a Ph.D. in quantitative economics from the University of Fribourg (Switzerland) in 1987. From 1987-92 he was on the faculty of Computer Science at the University of Montreal (Canada). His research interests cover mainly information retrieval for other languages than English (European, Asian, and Indian) as well as multilingual and cross-lingual information retrieval. He has participated for many years in various evaluations campaigns (CLEF, NTCIR and TREC) dealing with these questions.

Páraic Sheridan is the Scientific & Operations Manager at the Centre for Next Generation Localisation (CNGL) at Dublin City University. He holds an M.Sc. degree in Computer Applications from Dublin City University and an M.S. degree in Computational Linguistics from Carnegie Mellon. He completed his doctoral work in 1998 at the Swiss Federal Institute of Technology (ETH) Zurich with a dissertation on the topic of Cross-Language Information Retrieval. Dr. Sheridan then joined TextWise LLC, a start-up company in Syracuse, NY which was a spin-out from Syracuse University. At TextWise he worked on the CINDOR cross-language search system, which was commercialised and marketed by TextWise in the enterprise search space. Dr. Sheridan is now at CNGL, a research centre spanning four Irish universities and nine industrial partners and funded by Science Foundation Ireland (SFI) and industrial contributions. CNGL is pursuing an ambitious five year research programme to provide the foundation for the next generation of localisation and currently consists of over 100 researchers in areas of Language Technology, Digital Content Management, Localisation, and Systems Frameworks.

3 Programme of Lectures

The programme was studied to begin with a basic introductory course on multilingual and cross-language retrieval on the first day,. This was followed by a series of lectures on various aspects of the core problem: how to enable multilingual access functionality for diverse types of requirements and media. Issues covered included multilingual architectures and interfaces, multimedia retrieval, evaluation and applications in a multilingual context. The final lecture was on commercial aspects of cross-language search. An optional mentoring session was also held in the middle of the week. Four students presented their research ideas, which were commented by five of the lecturers: Páraic Sheridan (coordinator of the session), Giorgio Di Nunzio, Gareth Jones, Anselmo Peñas and Jacques Savoy. About one third of the students attended this session.

Monday, 15 June

Multilingual Information Retrieval and Cross-Language Retrieval

Martin Braschler(1) and Jacques Savoy(2)

(1) *Institute of Applied Information Technology, Zurich University of Applied Sciences, Switzerland*

(2) *Computer Science Dept., University of Neuchâtel, Switzerland*

Abstract

A lot of early work in Information Retrieval was exclusively focused on retrieval of English text documents. This limitation started to become addressed in the early 80s in earnest, with the advent of evaluation campaigns such as TREC in the 90s being a major force behind this development. In this lecture, we will show how to systematically extend basic monolingual indexing and matching, and adapt them for working with other languages. We will cover issues pertaining to the indexing process such as language identification, tokenization/segmentation (including Asian languages), word normalization, stemming/decompounding. We will discuss the effect of these measures on retrieval effectiveness. Understanding how to adapt IR systems successfully for many languages is a necessary pre-requisite to tackle the problem of Cross-Language Information retrieval (CLIR), i.e. the retrieval of documents written in a language different to the language of the user's request. The lecture will cover the different translation strategies to address the CLIR problem. Issues of cultural differences, translation ambiguity, pivot languages and extension to a large number of languages are covered.

Tuesday, 16 June: Morning

Image Retrieval

Henning Mueller(1) and Thomas Deselaers(2)

(1) *University Hospitals of Geneva, Switzerland*

(2) *Computer Vision Laboratory, Swiss Federal Institute of Technology - ETH Zurich, Switzerland*

Abstract

Image retrieval has been receiving increasing interest due to the vast amount of images publicly available on the Internet. Most image sharing sites, such as Flickr, allow for text/tag-based image searching. In the research community, content-based image retrieval has been under investigation since the early 1990s. The tutorial will show approaches to fusing the efforts from content-based image retrieval with the available text-based image searching solutions, including multilingual approaches. Therefore, a short introduction into content-based image retrieval and text-retrieval will be given. Different approaches to fusing these will be discussed and further topics such as user-interaction, retrieval system architecture, and benchmarking issues will be presented.

The tutorial will focus on image retrieval from different perspectives:

- Content-based image retrieval, i.e. finding images by their visual

- Content text-based image retrieval, i.e. finding images using textual information;
- Combination of the above
- Image retrieval in a multilingual context

Tuesday, 16 June: Afternoon

Information Access and Information Extraction

Anselmo Peñas

Department of Computing Languages and Systems, National Distance Learning University, Spain

Abstract

From the user perspective, Information Access is a cognitive process that involves more than the retrieval of a set of documents. Depending on the kind of information need, the sources of information, their size, media, etc. users must develop different strategies to find the information, explore and finally process it in order to group related pieces, synthesize pertinent information, discover trends, etc. All these tasks share something in common: the need to give structure to information. This is the general goal of Information Extraction. At the same time, all these tasks have to deal with language barriers like domain specific vocabularies or languages not known by the user. Terminology Extraction also provides some basic techniques useful for these challenges in Multilingual Information Access. The lecture will discuss these issues and study some research and evaluation frameworks related to Information Access tasks that need some kind of Information Extraction.

Information Extraction in a Multilingual Context

Maarten de Rijke

Informatics Institute, University of Amsterdam, The Netherlands

Abstract

With the increasing focus on retrieval systems that return information rather than just documents, robust and scalable information extraction (IE) technologies quickly gain in relevance. Opinions, themes, stakeholders, entities as well as their relations and attributes need to be identified reliably. In a European setting, and with tasks such as news monitoring, open source intelligence, and social media analysis, information extraction gets a distinctly multilingual flavour. In the presentation, some of the tasks giving rise to the challenges in multilingual IE will be discussed, together with proposed solutions. The presentation will be organized around examples and solutions that complement the ones offered in Anselmo Penas's presentation.

Wednesday, 17 June: Morning

System Architecture and Multilinguality

Nicola Ferro and Giorgio Maria Di Nunzio

Dept. of Information Engineering, University of Padua, Italy

Abstract

Multilingual Information Access (MLIA) systems are designed for both unstructured and structured data and deal with information needs for which relevant information can exist in languages other than the one the query is formulated in. In order to address different kinds of information needs in a flexible way and to cope with different types of information resources, architectural choices play an important role in the design and development of MLIA systems. Moreover, similar architectural components can play different roles according to the information needs and tasks that are being addressed. This session thus concerns the impact of the different architectural alternatives when designing MLIA systems for specific applicative scenarios.

Lesson I - MLIA Architectures: Standalone Applications and Services to Complex Systems (*Nicola Ferro*)

This lesson will present a conceptual architecture for a MLIA systems and analyse its main building blocks. The proposed architecture is suitable for developing standalone MLIA application focused on searching and retrieving multilingual documents. Moreover, we will discuss how the proposed architecture can be integrated and extended to be part of more complex systems, such as the digital library systems, and we will examine how it impacts on the design of such complex systems.

Lesson II MLIA Architectures: Advanced Services Beyond Searching (*Giorgio Maria Di Nunzio*)

This lesson will present how the building block of the proposed conceptual architecture can be looked from a perspective different from the search and retrieval one. In particular, we will discuss how both end-user services and system services can rely on and exploit the building blocks of the proposed architecture. Examples of the former are multilingual categorisation and filtering; examples of the latter are log keeping and analysis aimed at user personalization and profiling.

Wednesday, 17 June: Afternoon

Mentoring Session *Coordinated by Páraic Sheridan*

Opinion Mining

Alexandra Balahur-Dobrescu, University of Alicante, Spain

Multi-lingual Sentiment Analysis

Taras Zagibalov, University of Sussex, UK

Answers from the Quran: Online Information Seeking Needs for Searches related to Faith

Rita Zaharah Wan-Chik, University of Sheffield, UK

Arabic Multi-Document Summarization

Mahmoud Osman El-Haj, University of Essex, UK

Thursday, 18 June: Morning

Multilingual Search Assistance: Interactive Aspects of Cross Language Information Access

Paul Clough

Dept. of Information Studies, University of Sheffield, United Kingdom

Abstract

The interface acts as the intermediary between users of information retrieval (IR) systems and the search system itself. A well-designed interface should assist users in clarifying their information needs, and subsequently help them formulate suitable queries and understand the results. However, interactive multilingual information access systems provide an additional challenge to designers – users may not have the necessary language skills to find and interact with documents written in multiple languages. To provide effective access to multilingual document collections, users require search assistance. This session will explore approaches to support interaction at various stages within the search process: query formulation (e.g. offering the user additional query terms to refine their search such as synonyms), query translation (e.g. enabling the user to select from multiple query translations such as different word senses), document selection from search results (e.g. providing useable summaries for users to make informed decisions) and document examination (e.g. providing translated versions of documents for use by the end users). In addition to specific cross-language functionality, we will also discuss other aspects of interactive user interfaces such as design principles, evaluation criteria, use cases and general information seeking behaviours. Students will also gain hands-on experience using example interactive cross-language systems.

MultiMatch: A Multilingual/Multimedia Search Engine for Cultural Heritage: Interface Design and Demo

Paul Clough (1) and Franca Debole (2)

(1) University of Sheffield, United Kingdom & (2) ISTI-CNR, Pisa, Italy

Abstract

MultiMatch is a multilingual and multimedia search engine designed to enable users to explore and interact with online accessible cultural heritage content. The functionality of the system will be demoed and the various issues that had to be taken into consideration when designing the system interface to support multilingual search and browsing and multimedia retrieval will be discussed

Thursday, 18 June: Afternoon

Machine Translation for Multilingual Information Processing

Marcello Federico

Human Language Technology Research, Fondazione Bruno Kessler, Trento, Italy

Abstract

My lecture will survey the state-of-the-art in machine translation by mainly focusing on the so called statistical approach, that has been dominating the scene during the last decade. I will review approaches and methods developed to translate written and spoken documents, as well as available language resources and software tools that can be employed for multilingual information processing.

Multilingual Speech and Video Search

Gareth Jones

Centre for Next Generation Localisation, Dublin City University, Ireland

Abstract

Providing access to spoken and video content presents challenges in terms of indexing this content. This most obviously involves knowing what has been said, but can also involve use of supporting metadata. Manual transcription of spoken content is generally uneconomic, meaning that speech recognition techniques must be used. Searching spoken content in a multilingual environment introduces further issues of translation between the query and document languages involved. The visual data stream of video material is often independent of the spoken soundtrack meaning that it can be useful for language independent indexing and also as a retrieval target in multilingual search.

This tutorial will introduce techniques for speech and video retrieval, including speech recognition and visual indexing. These will then be explored in the context of multilingual search examining its opportunities and challenges.

Friday, 19 June: Morning

Evaluation for (Cross-Language) Information Retrieval

Mark Sanderson

Dept. of Information Studies, University of Sheffield, United Kingdom

Abstract

Evaluation of search output is a critical component of any form of Information Access, however in MLIA, evaluation is additionally challenging. In this session, I will provide a broad introduction to evaluation of information access systems, where the main focus of will be on evaluation using test collections. Here the design and use of test collections will be described along with the measures employed to determine the effectiveness of a searching system. Alternative evaluation methods such as query log analysis and user studies will also be described. Next, the means by which evaluation

methods have been adapted to work with multi-lingual systems are described. Finally, the additional assessments that need to be conducted to fully understand the success of MLIA are detailed.

Friday, 19 June: Afternoon

Commercial Cross-Language Search

Páraic Sheridan

Centre for Next Generation Localisation, Dublin City University, Ireland

Abstract

This session will begin by examining the economic imperatives of Multilingual Information Access. We will examine trends in participation and make-up of those acting as publishers and consumers of information on the World Wide Web, and the languages they are using. We will see examples of innovative use of technologies (e.g. Machine Translation) and user participation to break down language barriers and provide access to information across languages and markets. I will also present some analysis of the Enterprise Search market from the viewpoint of Multi-Lingual Information Access. Where are MLIA features being used? What are customers' expectations? Who are the players in this space and what do they offer? I will conclude by asking, "For researchers and developers, where are the opportunities?" and provide some ideas based on my own experience and observations in this field.

4 Participants

The School was intended for advanced undergraduate and post-graduate students, post-doctorial researchers plus academic and industrial researchers and system developers with backgrounds in Computer Science, Information Science, Language Technologies and related areas. We accepted 40 applications for the School. 3 cancelled at the last moment. This was unfortunate as we had had to reject several late applications due to restrictions imposed by the size of the lecture room.

4.1 Grant Assignments

Two types of grants were awarded:

- Grant A: covering accommodation for 6 nights and breakfast; Grant A does not include the 200 euro registration fee
- Grant Bs; Registration fee only

30 applications for grants were received. Applications consisted of a letter from the applicant motivating why participation in the School was important for them accompanied by a letter of recommendation by their supervisor or equivalent. The applications were reviewed by 5 members of the PC: Maristella Agosti, Martin Braschler, Donna Harman, Costantino Thanos and Felisa Verdejo. 13 Grant As and 7 Grant Bs were awarded. In addition 3 students from TrebleCLEF institutions were given Grant Bs on the condition that they assisted in the local organisation.

When assigning the grants, in addition to consideration of the application and the recommendation, the following criteria were adopted as criteria of preference:

- students should be on a PhD course (and preferably towards the beginning rather than the end of their studies)
- we aimed at having a mixed scientific background, i.e. not all computer science or NLP people but a mix.
- we wanted to have a good geographic balance - this meant that if they met our other criteria we automatically accepted students that were the only applicant from a given nation.
- we didn't think that student's coming from CLEF groups should have preferential treatment - we thought that the Summer School was a way of widening our community
- we wanted to have a reasonable M/F balance, e.g. not below 1/3 of either sex.

4.2 List of Participants

In order to enable the students to acquainted before the beginning of the School, and also to give the lecturers the opportunity to know something about the students background, we asked all students to send a brief CV and photo to be posted on the Summer School website. This was a very popular decision.

Daniela Alderuccio (daniela.alderuccio at enea.it) actually works in the ICT domain at ENEA (Italian National Agency for New Technologies, Energy and the Environment) in the Research Centre of Casaccia (Rome). She graduated cum laude in Modern Foreign Languages and Literatures at the University of Rome "La Sapienza", with an interdisciplinary ENEA thesis on semantic information retrieval in multilingual digital libraries (harmonizing Humanities Computing, AI and Computational Linguistics). Her research activities focus on Multilingual Text Processing for Intelligent Access to Multilingual e-Content, by using linguistic software packages for content analysis. Her research interests include: Multilinguism and Knowledge Representation (Ontologies) - Cross Language Information Retrieval - Web Archiving - e-Humanities. Application Fields: Digital Libraries - Semantic Web – GRID.

Aliksandr Autayeu (Aliksandr.Autayeu at disi.unitn.it) is a PhD student of the Doctorate School in Information and Communication Technologies of the University of Trento. His current research

interests are connected with application of natural language processing techniques to metadata, such as web directory classification labels, business directory classes, picture titles, and other. Aliaksandr Autayeu graduated from Mechanics and Mathematics Faculty of Belarusian State University with specialization in Computer Mathematics. His past research interests included applications of wavelets to content-based image retrieval.

Stefano Baccianella (stefano.baccianella at isti.cnr.it). I actually works as a researcher at ISTI-CNR, Pisa and PhD Student in Computer Engineering at Università di Pisa. I have a master's degree in Computer Science and my research topics concerns all the aspects of the classification of multimedia documents. In particular my focus is on the classification of textual documents and the analysis of the sentiment carried by them. Main research areas include the study of: Text Mining, Text Classification, Sentiment Analysis and Opinion Mining.

Alexandra Balahur-Dobrescu (abalahur at dlsi.ua.es). I am a graduate in Computer Science from the "Al. I. Cuza" University in Iasi, Romania. My graduate thesis was in Textual Entailment, in which I was working in 2007 and 2008. In parallel, I have started working in multilingual Question Answering and Answer Validation. I am currently a second year Ph.D Student at the University of Alicante, Spain, where my main research area is sentiment analysis. At the moment (April 2009 - September 2009) I am doing a traineeship with the European Commission's Joint Research Centre in Ispra, Italy, within the Institute for the Protection and Security of the Citizens, where I am applying my doctoral research in sentiment analysis over newswire texts. I speak Romanian, English, German, Spanish, French, some Portuguese and I am currently learning Italian.

Guillaume Bernard (gbernard at limsi.fr). I am from France and I am currently a Phd Student, in my second year. I am working in the LIMSI-CNRS laboratory, in Orsay, on computational linguistic, on the field of questions-answering systems. Even if I am studying all kind of linguistic phenomena, I am more specialized in textual implication : in the context of a questions-answering system (Ritel), I am trying to identify cases where the answer to a question asked by a user is not stated explicitly in our documents.

Annalina Caputo (acaputo at di.uniba.it). am a second year Ph.D student in Computer Science at the University of Bari (Italy). My research activity is principally focused on studying new hybrid Information Retrieval models and applying new methods to index documents content. The main Research Area is Information Retrieval (IR). In particular, the ultimate goal of my Ph.D is to investigate performance improvements in Information Retrieval systems due to the use of Natural Language Processing techniques which are able of extracting semantic information from documents. Currently, my research is concerning to the integration of Named Entities in IR systems.

Balint Daroczy (daroczyb at yahoo.co.uk). As a PhD student at the Eötvös Lóránd University Budapest currently I am working on my dissertation on field of image and text search on the web. In the last two years I have been working on an image segmentation method based on three types of image growing method for an image ranking system combined with wide range of classifiers and a cross-media clustering method (bi-clustering). I am interested in the fields of image processing, classification techniques and large scale image retrieval systems.

Le Thanh Dinh (thanh190882 at yahoo.com). I am doing the second year master of the European Language and Communication Technology program (LCT). My first year was in Free University of Bozen-Bolzano and this second year I am studying in the Charles University of Prague. My current interests are mainly about Machine Learning and Statistical methods, Information Retrieval and Multilingual. I am doing my master thesis about Question Classification, Topic shift Classification using machine learning methods.

Marco Dussin (dussinma at dei.unipd.it). Since February 2006 I am member of the research team of the Department of Information Engineering of the University of Padua, where I study. I deal with the study and the design of Graphical User Interfaces (GUIs) – in particular Web UIs – for prototypes made by the Information Management Systems research group (IMS). My latest research activity is focused on the interface of Distributed Information Retrieval Evaluation Campaign Tool (DIRECT), a prototype of a Digital Library System (DLS) able to support the course of an evaluation initiative, and

to manage, curate and enrich the scientific data produced during it. DIRECT has been adopted and tested in the Cross-Language Evaluation Forum (CLEF) since 2005. I'm interested in human-computer interaction, web design and information architecture.

Mahmoud El-Haj (melhaj at essex.ac.uk) Position: Computer Science PhD student. Organization: University of Essex, UK. Previous Studies: MSc, Information Systems, University of Jordan BSc, Computer Information Systems: University of Jordan Fields of Interest: • Arabic Natural Language Processing • Automatic Text Summarization • Cross-Lingual Information Retrieval Research activities: Member at the Language and Computation (LAC) group at Essex University

Ingrid Falk (ingrid.falk at loria.fr). Current position: Research assistant at INRIA Nancy-Grand Est, Nancy, France and Ph. D. student at Université Nancy 2. Description: After a degree in Mathematics I worked as a programmer with several academic teams in Nancy involved in language and document engineering. I was principally concerned with modelling, creating and maintaining morpho-syntactic, syntactic and semantic lexical resources for French. In 2008 I received a master's degree in Computational Linguistics and work since fall 2008 as a research assistant and Ph.D. student at the INRIA research center in Nancy. My current research interests lie in exploring the interface between (multilingual) lexical information and ontology elements.

Pamela Forner (forner at celct.it). I graduated in Foreign Languages and Literatures at the University of Trento in 2000. From 2001 to 2006 I worked at ITC-irst doing research in the field of contrastive linguistics, computational lexicography (Italian MultiWordNet), and multilingual corpora. From early 2006 to the present, I have been working at CELCT on the Organization of evaluation campaigns, and annotations at different levels.

M. Rami Ghorab (ghorabm at cs.tcd.ie). Current: PhD Candidate at School of Computer Science & Statistics, Trinity College Dublin, Ireland. Received his MSc in IT from the School of CSIT, University of Nottingham, UK (2003). Received a diploma in IT from the Java Department, Information Technology Institute, Egypt (2002). Received his BSc in Computer Science from the CS Department, Modern Academy in Maadi, Egypt (2001). Held the position of Head of Java Department at the Information Technology Institute, Egypt (2007-2008) Teaching Assistant at Information Technology Institute, Egypt (2003-2008). Teaching Assistant at the Modern Academy in Maadi, Egypt (2001-2002). Research interests: Multilingual Information Retrieval, Adaptive Hypermedia, and User Modeling. Other research interests include: Peer-to-Peer Networks, Artificial Intelligence, and Data Mining.

Eniko Héja (eheja at nyud.hu). I am a junior researcher at the Dept. of Language Technology, Research Institute for Linguistics (HAS) and also a PhD student at the Dept. of Theoretical Linguistics, University of Budapest. I hold MA degrees in philosophy and theoretical linguistics. I am currently working on the automatic generation of bilingual dictionaries. The topic of my thesis is the semantic representation of (Hungarian) verbs from a word sense disambiguation point of view. I am particularly interested in methods that combine symbolic and statistical approaches.

Charlotte Lecluze (charlotte.lecluze at pertimm.com). After a Master Degree in Linguistics at the University of Caen (UCBN, France), a year ago, I started a professional PhD in Computer Sciences at Pertimm, Information Management Experts. I'm also a member of the Interaction Semiotics: Language, Diagrams Group (ISLanD) at UCBN. My current research projects are centred on the implementation of a manual methodology for semantic alignment of sub sentential segments, in order to improve cross-lingual information retrieval.

Vanessa Lopez Garcia (v.lopez at open.ac.uk). Vanessa Lopez is a research fellow at the Open University's Knowledge Media Institute, where she is also a part-time PhD student. Her research interests are in natural-language front ends to query the Semantic Web. Lopez received her MSc in computer engineering from the Technical University of Madrid.

Yashar Mehdad (mehdad at fbk.eu). PhD student in ICT (Information and Communication Technologies) International Doctorate School at the University of Trento. He is working on Textual Entailment, as his main research activity, with Bernardo Magnini in HLT (Human Language

Technology) group in Bruno Kessler Foundation (FBK). After his degree in Engineering (Iran, 2000), he studied Information Technology at University of Malaya in Malaysia, which led to achieve his master degree with distinction in 2006. Moving towards his research enthusiasm, he obtained his second level master in Human Language Technology at University of Trento, in 2008. His main interests, besides Textual Entailment, are Statistical Natural Language Processing, Computational Semantics, Machine Learning, Named Entity Recognition, Wikipedia-based NLP, Multilingual Information Access and Computational Linguistics for Persian (Farsi) Language.

Jinming Min (jmin at computing.dcu.ie) is a PHD student of the Computing School at Dublin City University. His current research interests are multilingual query translation with the IBM translation model and short-length document expansion from the external resource such as Wikipedia. Jinming Min graduated from Chinese Academic of Sciences majored in Computer Science. His past research interests included cross language information retrieval especially the out-of-vocabulary words translation and translation disambiguation.

Soto Montalvo (soto.montalvo at urjc.es). Assistant Professor at Computer Science Department at Universidad Rey Juan Carlos (Madrid, Spain) and PhD student at Universidad Rey Juan Carlos. Previous studies: M. S. in Computer Science and Information Engineering (Universidad Rey Juan Carlos) B. S. in Computer Science and Information Engineering (Universidad Politécnica, Madrid, Spain). Primary Research interests: Automatic Multilingual News Clustering, Cognate Identification and Information Extraction

Sergio Navarro (snavarro at dlsi.ua.es) After 6 years of experience working in a software engineering private company, two years ago I started my PhD in the Natural Language Processing and Information Systems Group in the University of Alicante, Spain. There my research work has been focused on multimedia information retrieval area, concretely on looking for the most suitable ways of fusing multimodal systems and sources in order to achieve systems with better precision and better diversity handling. This summer school is a great opportunity for me in order to complete my formation expanding my view of the related state of the art technologies.

Alexandre Patry (patryale at iro.umontreal.ca). I am completing my PhD on the topic of statistical machine translation at the Université de Montréal in Canada. I currently work on a better integration of context in phrase-based machine translation and on parallel documents retrieval from bilingual corpora.

Mari-Sanna Paukkeri (mari-sanna.paukkeri at tkk.fi) I am a post-graduate researcher at the Adaptive Informatics Research Centre at the Helsinki University of Technology. At the moment I am visiting the School of Informatics at the University of Edinburgh. I work on unsupervised language-independent methods, especially text mining and clustering. I am also interested in social networks and subjective language use. I have conducted studies for eleven European languages.

Alberto Pérez García-Plaza (alpgarcia at lsi.uned.es). I am a Ph.D. candidate and teaching assistant at the Department of Computer Systems and Languages and member of the UNED group in Natural Language Processing and Information Retrieval. My main research interests are Web Page Clustering, HTML Document Representation, Fuzzy Logic and Self-organizing maps. Currently I am interested in testing my methodology with more than one language. For more detailed information you can take a glance to my home page <http://nlp.uned.es/~alpgarcia>

Ginevra Peruginelli (peruginelli at ittig.cnr.it) is a researcher of the Italian National Research Council. She has a degree in Law (1999) and holds a MA/MSc Diploma in Information Science awarded by the University of Northumbria, Newcastle, UK (2005) and a Ph.D. in Telematics and Information Society from the University of Florence (2009). Since 2000 she has been working at the Institute of Legal Information Theory and Techniques (ITTIG), in Florence. Her main research areas involve: techniques and methods for accessing legal documentation; cross language legal information retrieval, open access to law. From 2004 she is professor under contract of Legal informatics at the Faculty of Law, University of Perugia. In 2003, she was admitted to Bar of the Court of Florence as a lawyer. She has been involved in various European projects on implementing multilingual semantic tools (ontology, thesauri) for accessing European legal information (legislation, case law, legal literature). She has

published several academic articles and delivered national and international conference contributions on law and legal language documentation, legal standards and knowledge extraction. Recently she has published two independent volumes on multilingual access to law.

Fabien Poulard (fabien.poulard at univ-nantes.fr). I'm doing a PhD at Nantes University (France) on content reuse and plagiarism detection. During my master thesis, I've worked on detecting quotations in journalistic articles. I'm interested in this summer school to extend my knowledge on multilingual approaches. The information retrieval techniques can be particularly useful to filter content reuse candidates.

Valeria Quochi (vquochi at gmail.com). I currently work as a junior researcher at the CNR-ILC in Pisa. I hold a PhD in Linguistics and a Masters degree in Foreign Languages and Literature. I worked on the representation and acquisition of lexical semantic information esp. of (semi-) transparent multiword expressions or collocations (complex nominals and light verb constructions). At ILC I have also been involved in projects related to the standardization of lexical resources. Currently, I'm interested in automatic acquisition / extraction / annotation of lexical information and semantic relations, in a cross-lingual perspective, for possible applications in MT.

Alejandro Revuelta Martínez (Alejandro.Revuelta at uclm.es). I studied computer science at Castilla-La Mancha University in Spain. Currently I am a PhD student and part of a research project at the same university. I have studied statistical machine translation and automatic speech recognition systems but I am interested in the whole field of natural language processing

Megan Richardson (megan.richardson at gmail.com). LIMSI - CNRS, France (No CV)

Dávid Siklósi (sdavid at ilab.sztaki.hu). MTA SZTAKI, Budapest, Hungary sdavid@ilab.sztaki.hu (No CV)

Adrian Smales (a.smales at nhm.ac.uk). Head of ICT, Natural History Museum, London. Previous Studies: BEng(Hons) in Electronic and Computer Engineering, Masters in Business Administration, Edinburgh University, Dissertation: Food Information Logistics Current Activities: • Developing IT Technology and Business Strategy for BHL Europe, Chief Technical Architect (BHL WP3), Harmonising data between 28 European institutions. • Architecting and raising sponsorship funding for "World's Greenest Data centre" via Dell, IBM, 3PAR, EMC, Intel, Cisco and others. Approx Value circa £7M - £10M. • Developing sustainability business model for BHL Europe, incorporating Storage, Scanning/Digitisation and Data Centre.

Diana Irina Tanase (tanasedi at wmin.ac.uk). Diana Irina Tanase is a PhD student at University of Westminster, London, UK. Her research is focused on integrating user context to cross-language information retrieval. Her other projects include development work on the Computational Science Education Reference Desk (a NSDL pathway), and a number of collaboration web tools for Design Interaction, Royal College of Art. Her initial training was received at Ovidius University, Romania (2001), followed by a Master of Science at University of Northern Iowa, USA (2003).

Marco Turchi (marco.turchi at gmail.com). I'm Research Assistant at University of Bristol, Department of Engineering Mathematics, Pattern Analysis and Intelligent Systems group. I got my PhD at University of Siena, Italy, on "Computer Engineering, adaptive information processing". During my PhD studies, I was visiting student at University of California Davis, Statistical Department, and intern at Yahoo Research Lab. Before coming in Bristol two years ago, I had a temporary research position at Xerox Research Centre Europe (XRCE). My current research is centered about the European project SMART, applying machine learning techniques to statistical machine translation problems, and I am also involved in a media analysis project aimed at modelling the mediasphere based on text mining and cross-language analysis techniques.

Ville Turunen (ville.t.turunen at tkk.fi). I am a PhD student in Computer Science at Helsinki University of Technology. My main research topic is speech retrieval and I am currently working on an approach for Finnish speech retrieval based on statistically discovered morphemes. I am especially interested in language independent approaches. My previous studies include language technology and machine learning.

Evgenia Vassilakaki (EVGENIA.VASSILAKAKI at student.mmu.ac.uk) is a second year Ph. D Student in the Dept. of Information & Communications at Manchester Metropolitan University studying under a 3 years, full-time studentship offered by The Information Research Institute. Evgenia's research interests lie in users' image seeking behaviour in multilingual environments. In particular, she is focusing on users' trust and confidence when searching for images across languages. Evgenia is also working as a part time junior researcher on some of the projects in the research center, CERLIM.

Rita Zaharah Wan-Chik (rita.zaharah at gmail.com). I am originally from Malaysia and have just started my PhD in Information Studies with the University of Sheffield early this year. I am currently on a study leave from teaching at the Universiti Kuala Lumpur, Malaysia. My research interest is leaning towards multi-language and cross-language web search particularly on the information seeking related to faith in general and to Islam and Al Quran in specific.

Taras Zagibalov (taras8055 at googlemail.com). I am currently a PhD student in University of Sussex, UK. I've been doing some research in unsupervised sentiment and subjectivity extraction, trying to create an approach that could be domain and language independent. I've been experimenting with English, Chinese and Japanese data, also planning to include texts in Russian. I'm quite fluent in English, Chinese (Mandarin) and Russian.

Veronika Zenz (v.zenz at matrixware.com). I'm an Information Retrieval Scientist at Matrixware and member of the Matrixware Research Group. I've studied computer science at the Technical University of Vienna, where I received my master degree in Software Engineering & Internet Computing in 2007. I've written my master thesis on music information retrieval, more precisely on chord detection in polyphonic audio. I'm currently helping in the organization of the CLEF-IP track and am also involved in the creation of a digital library for matrixware and the IRF.

Arkaitz Zubiaga (azubiaga at lsi.uned.es). I'm a PhD student at the Languages and Computing Systems Department of the National University of Distance Education (UNED). I'm member of the Natural Language Processing and Information Retrieval Group at UNED. My main research interests revolve around social tagging and the knowledge offered by this kind of metadata, specially to exploit them for data mining and information retrieval tasks. In this way, I would like to learn interesting approaches to work against social networks, which commonly have a multilingual nature.

5 Evaluation

Following normal practice for this type of event, students were asked to compile evaluation forms at the end of the School; we received just 29 forms. With the exception of one student who clearly had mistaken their choice of School, the other comments were for the most part very positive. A breakdown of the evaluation results is given in Appendix A. A number of thank you letters were received from students after the event; Appendix B reports one of these as it gives a good idea of the atmosphere of the School. Additional information and copies of the course material distributed can be found on the School website at <http://www.trebleclef.eu/summerschool.php>.

Appendix C which is given separately contains copies off all the slides presented at the School. One of the students recorded all the presentations and intends to prepare a slide show of all the course with the lecturers comments. We have obtained the consent of the lecturers and if the slide show is of sufficiently good quality it will be made publicly available on the TrebleCLEF website.

Appendix A: TrebleCLEF Evaluation results		1= poor - 5 = excellent				
	total	1	2	3	4	5
Organisation and results						
1. Overall Organisation	29				7	22
					24%	76%
2. Location	29			3	8	18
				10%	28%	62%
3. Lodging	27			5	11	11
				19%	38%	38%
4. Social Events	29			1	13	15
				3%	45%	52%
5. Assistance	29			0	5	24
				0%	17%	83%
COMMENTS	I liked that the school was at the same time as Pisa's feast					
	The school is very well organised. Ms. Francesca Borri and the other organizers did an amazing job helping the students for every need					
	I have no issues. Fantastic Event!					
	It was good. Lodging and Location don't have anything to do with the quality of the school					
	The organization was perfect and Francesca gave us all the assistance we needed. The food was quite good, except for the coffee					
	Extremely well organised, thank you!					
	Good organization with a lot of social events that make friendship easier					
	how to go from the main bus station to Fossabanda could have been more precise (name of the station)					
	Just thanks to Francesca for all she did					
	Keeping the timing of the sessions was excellent					
Lectures						
1. Overall how effective were the lectures	29		1	2	20	6
			3%	7%	69%	21%
2. How would you rate the structure of the school?	29		0	2	13	14
				7%	45%	48%
3. How well did the school address the MLIA area?	29		1	2	12	14
			3%	7%	41%	48%
4. Were the handouts	26		1	1	13	11

sufficient?						
			4%	4%	50%	42%
5. What else would you have liked to have seen included?	More hand-on exercises					
	The school really surprised me in terms of quality and pertinence of the courses. The speakers give a clear idea of different aspects of IR					
	More emphasis on linguistic aspects of MLIA from a contrastive perspective (ex. Which IR systems / IR strategies perform better in German, French, Arabic and so on.					
	More about language independent approaches and possibilities. Lectures were good but perhaps too general for my needs (I knew most of the things beforehand). Anyway it is good to have some repetition to know things better.					
	More statistical Systems. Having some experience from the "trenches" was really interesting (what to do with unclear data, tricks and things like that which are learned from experience)					
	Something on ML lexical resources					
	More exercises. A summer school should also incorporate lab sessions to get people to apply some of the methods that are explained. Maybe students can be split into groups based on interests and work together on mini-projects.					
	Some lectures had too many slides and ran out of time. This forced them to avoid more technical parts and reduced some lessons to only an introduction to various problems (i.e. Deselaers and Muller lessons).					
	We didn't have all the handouts and that's because lecturers didn't send them on time. So, maybe in the future there will be a deadline for them to submit the material. It is better to have a sample of them even if it is not complete so as to write notes down rather than nothing! (<i>Note: there was a deadline</i>)					
	Broader coverage on data mining					
	More practical sessions, more interaction					
	If I would provide a bad issue, that would be that in a general way there were a lot of exposed information and access obstacles but not so much solutions and perspectives of evolutions.					
	Labs exercises were really interesting, more of them (in each lecture) would be appreciated!					
	I appreciated the several practical sessions and the interaction with professors during the lectures					
	I would have liked to see more in-depth details about the topics. Excercises were great					
	Very good idea to give papers for reading and commenting during the session					

	Excellent programme and lecturers. Dense (one more day would have helped. Since we finished at 18,00 and then we had social events, we didn't have time to rest or to have a nap which made us lazy somehow)					
	Most of the speakers were very good indeed and presented/spoke clearly					
	total	1	2	3	4	5
Overall Opinion						
1. did the school meet your expectations?	29		1	1	16	11
			3%	3%	55%	38%
2. Will participation in the school be useful for your future studies/career?	29	1	1	4	14	9
		3%	3%	14%	48%	31%
3. Can you suggest ways for improvement in the future?	The school does not need any improvements. Thanks					
	Put the video and lectures on the summer school web site					
	Add more days in order to shorten the daily hours. Up until 16,00 is fine.					
	A little bit more practical approach. More exercises would have been useful for me.					
	More details					
Overall Judgement of the school	29			1	13	15
		0%	0%	3%	45%	52%
	Congratulations, I really enjoyed.					
	Impeccably organized, great talks! Thank you! One observation regarding some of talks. The speakers were heavily biased and presented a lot of their own work, more than I felt was necessary.					
	I learned many new things and met many lovely friends, so you deserve more than 5. Thank you					

Dear All Coordinators, Participants and Lecturers,

Hope you are doing very well and happy. I want to thank all of you for this lovely school; everything was amazing and well organized.

When I registered for this summer school, I was wondering what Multilingual could really mean. Is it just a series of voice waves interpreted in many languages? Or is it just a different way to say whatever you want in a different tune? Not until I was there, since the first minute in the airport I realized the true meaning of multilingual, when you figure out that what you speak is a series of gibberish words! When you feel that your hands and body language could speak better than your tongue. I remember asking that girl about which bus should I take, I was speaking in English, she responded in fluent Italian and still we understand each other, but not a single word! Being from a culture that is only known by worst news headlines I was afraid that I'll not be able to mix with other cultures! Not until I was there, met by everyone's smiles and kindest hospitality. I felt home being surrounded by all those tact and adorable people from everywhere.

I admit that this summer school was one of the best experiences I ever had, many great programs, best lecturers, amazing events and Great friends.

Best regards,
Mahmoud EL-Haj





