



PROMISE

Participative Research labOratory for
Multimedia and Multilingual Information Systems

FP7 ICT 2009.4.3, Intelligent Information Management
www.promise-noe.eu

Deliverable 7.2

First PROMISE Annual Conference and Proceedings

The CLEF 2010 Conference on Multilingual and Multimodal Information Access Evaluation
Held in Padua, Italy, 20-23 September 2010

Version 1.0, 30th September 2010



Document Information

Deliverable number:	7.2
Deliverable title:	First PROMISE Annual Conference and Proceedings. The CLEF 2010 Conference on Multilingual and Multimodal Information Access Evaluation, Held in Padua, Italy, 20-23 September 2010
Delivery date:	30/09/2010
Lead contractor for this deliverable	CELCT
Author(s):	Pamela Forner (CELCT), Nicola Ferro (UNIPD)
Participant(s):	CELCT, UNIPD
Workpackage:	7
Workpackage title:	Dissemination, IPR, and Resources
Workpackage leader:	ELDA
Dissemination Level:	PU – Public
Version:	1.0
Keywords:	CLEF 2010 conference, labs and workshops, liaison meeting

History of Versions

Version	Date	Status	Author (Partner)	Description/Approval Level
0.1	27/09/2010	Draft	CELCT	Circulated to all partners
0.2	29/09/2010	Draft	UNIPD	Revised version
1.0	30/09/2010	Final	CELCT	Final version

Abstract

Deliverable 7.2 consists of two main parts: the organization of the CLEF 2010 conference and labs, and the publication of the proceedings. The conference and the labs were held at University of Padua for 20th to 23rd September 2010 – the first two days (20-21 September) were the conference part and were hosted at Palazzo del Bo, the main historical building of the University of Padua; the second two days (22-23 September) were the labs part and were hosted at the Faculty of Statistical Sciences of the University of Padua. CLEF 2010 represents the first event organized by PROMISE and was also the opportunity to publicly launch the project in the research community.

This documentation provides an overview of the CLEF 2010 conference together with a description of the labs that have been offered, details on this year's participation, and a short report on the workshops. Moreover, we describe the website developed for the event and the contents available on it. Finally, we briefly report on the "First Evaluation Initiatives Liaison Meeting" that has been held during CLEF 2010.

In appendix, we provide:

- the list of the people involved in the organization of CLEF 2010
- the list of the institutions involved in the coordination of CLEF 2010 Labs and Workshops
- the Conference Programme
- the Labs and Workshops Programme
- the list of the conference/workshops participants
- the list of the institutions which submitted results in the CLEF 2010 evaluation campaigns (Labs), showing also in which Lab they took part
- CLEF 2010 Sponsors
- The ERCIM News about CLEF 2010.

Table of Contents

Document Information	3
Abstract.....	3
Table of Contents.....	5
Executive Summary	6
1 Introduction	8
2 CLEF 2010 Conference.....	11
3 CLEF 2010 Labs and Workshops	13
4 Participation in CLEF 2010.....	18
5 CLEF 2010 Web Site and Advertising.....	20
5.1 CLEF 2010 Web Site	20
5.2 Advertising.....	21
6 First Evaluation Initiatives Liaison Meeting	22
7 CLEF 2011 and Beyond	23
8 Acknowledgements.....	23
References	24
Appendix A: CLEF 2010 Organization	25
Appendix B: CLEF 2010 Labs and Workshops Coordination.....	26
Appendix C: CLEF 2010 Conference Programme	28
Appendix D: CLEF 2010 Labs and Workshops Programme.....	33
Appendix E: List of participants at the Event.....	40
Appendix F: List of Participating Institutions to the Labs	47
Appendix G: CLEF 2010 Sponsors	52
Appendix F: ERCIM News About CLEF 2010	53

Executive Summary

Deliverable 7.2 comprises two main parts: the organization of the CLEF 2010 conference and labs, and the publication of the proceedings.

CLEF 2010 consisted of two main parts: (i) a peer-reviewed conference on experimental evaluation, and (ii) a series of Labs designed to test different aspects of mono- and cross-language information retrieval (IR). The intention is to encourage experimentation with all kinds of multilingual information access – from the development of systems for monolingual retrieval operating on many languages to the implementation of complete multilingual multimedia search services. Seven Labs were offered:

- **CLEF-IP**: a benchmarking activity on intellectual property
- **ImageCLEF**: a benchmarking activity on image retrieval
- **PAN**: a benchmarking activity on plagiarism and Wikipedia vandalism detection
- **RespubliQA**: a benchmarking activity on question answering using multilingual political data
- **WePS**: a benchmarking activity on Web people search
- **CriES**: a workshop aimed at exploring the evaluation of searching for expertise in social media
- **LogCLEF**: a workshop aimed at exploring methodologies for studying search engine log files

The conference and the labs were held at University of Padua for 20th to 23rd September 2010 – the first two days (20-21 September) were the conference part and were hosted at Palazzo del Bo, the main historical building of the University of Padua; the second two days (22-23 September) were the labs part and were hosted at the Faculty of Statistical Sciences of the University of Padua.

CLEF 2010 represents the first event organized by PROMISE and was also the opportunity to publicly launch the project in the research community.

The outcomes of the Conference and Labs are collected in two different and separated publication forms: conference proceedings are published by Springer in their Lecture Notes for Computer Science (LNCS 6360) series with the title *Multilingual and Multimodal Information Access Evaluation*, while all the experiments carried out by systems during the CLEF 2010 evaluation campaigns are described in the Notebook papers, which were published in digital form only and are available at the CLEF 2010 website (<http://clef2010.org/>). Printed abstracts are collected in the Book of Abstracts. Both Notebook Papers and Book of Abstracts are divided into seven main sections, corresponding to the seven Labs activated this year, they are further subdivided in many subsections corresponding to the different tasks planned.

Both online Notebook papers and Book of Abstracts have been assigned an ISBN and ISSN number, and have also been indexed by DBLP.

Post-conference publications in suitable venues, such as journals, are welcome and under the responsibility of lab organizers.

The CLEF 2010 Conference Proceedings, together with the Labs Book of Abstracts have been distributed to attendees at the conference and workshops, and are thus also to be considered as part of this deliverable as additional appendices.

This documentation provides an overview of the CLEF 2010 conference together with a description of the labs that have been offered, details on this year's participation, and a short report on the workshops. Moreover, we describe the website developed for the event and the contents available on it. Finally, we briefly report on the "First Evaluation Initiatives Liaison Meeting" that has been held during CLEF 2010.

In appendix, we provide:

- the list of the people involved in the organization of CLEF 2010
- the list of the institutions involved in the coordination of CLEF 2010 Labs and Workshops
- the Conference Programme
- the Labs and Workshops Programme
- the list of the conference/workshops participants
- the list of the institutions which submitted results in the CLEF 2010 evaluation campaigns (Labs), showing also in which Lab they took part
- CLEF 2010 Sponsors
- ERCIM News about CLEF 2010.

1 Introduction

The overall goal of PROMISE is to advance the experimental evaluation of complex multimedia and multilingual information systems in order to support individuals, commercial entities, and communities who design, develop, employ, and improve such complex systems.

In order to pursue its overall goal, PROMISE will leverage a number of key activities, one of which is the organization of the PROMISE annual conference on experimental evaluation where the results of the various evaluation activities can be presented and discussed as well as proposals for novel evaluation methods, metrics, and tasks addressing core issues in the evaluation field can be presented and discussed in order to advance the evaluation field itself.

Large-scale evaluation campaigns at the international level have provided a fundamental contribution to the advancement of state-of-the-art techniques and development of innovative information systems through common evaluation procedures, regular and systematic evaluation cycles, comparison and benchmarking of the adopted approaches and solutions, and the spreading and exchange of knowledge and know-how.

CLEF 2010 was the first and launching event organized by the PROMISE network. It is starting to shape CLEF in the directions envisioned by PROMISE which sees the relationship between multilingual and multimedia information access systems and experimental evaluation as a kind of “temple” with three levels, as shown in the below in Figure 1:



Figure 1: Vision for the CLEF conference and labs/workshops.

- **Basement:** the conference aimed at advancing the evaluation of multilingual and multimedia information systems, the evaluation methodologies and metrics developed to embody realistic use cases and evaluation task. The techniques aimed at bringing more automation in the evaluation process constitute the basement of CLEF and will provide the foundations for promoting and supporting the expected scientific and technological advancement.
- **Pillars:** the regular and thorough evaluation activities carried out in the labs, the realistic use cases and evaluation tasks designed for compelling user and industrial needs represent the pillars of CLEF. They stimulate the research and development in the multilingual and multimedia information systems field and they contribute to the creation and driving of a multidisciplinary researchers and developers community which brings together the competencies needed to develop such complex systems.
- **Roof:** the “basement” and the “pillars” of CLEF will give the necessary support for designing and developing the next generation multilingual and multimedia information systems needed to address the emerging user needs and to cope with the interaction among content, users, languages and task which we are seeing nowadays.

Since 2000 the Cross-Language Evaluation Forum (CLEF) played a leading role in stimulating investigation and research in a wide range of key areas in the information retrieval domain, such as cross-language question answering, image and geographic information retrieval, interactive retrieval and many more. It also promoted the study and implementation of appropriate evaluation methodologies for these diverse types of tasks and media. As a result, CLEF has been extremely successful in building a wide, strong, and multidisciplinary research community, which covers and spans the different areas of expertise needed to deal with the spread of CLEF tracks and tasks. This constantly growing and almost completely voluntary community has dedicated an incredible amount of effort to making CLEF happen and is at the core of the CLEF achievements.

CLEF 2010 represents a radical innovation of the “classic CLEF” format and an experiment aimed at understanding how “next generation” evaluation campaigns might be structured. We have had to face the problem of how to innovate CLEF while still preserving its traditional core business, namely the benchmarking activities carried out in the various tracks and tasks.

The consensus, after lively and community-wide discussions, has been to make CLEF an independent four days event, no longer organised in conjunction with the European Conference on Research and Advanced Technology for Digital Libraries (ECDL) where CLEF has been running as a two and half day workshop. CLEF 2010 thus consisted of two main parts: (i) a peer-reviewed conference (20–21 September) on experimental evaluation, which innovated the CLEF tradition. It covered a broad range of issues in the fields of multilingual and multimodal information access evaluation; and (ii) a series of Labs (22–23 September), which continued the CLEF tradition of community-based evaluation.

CLEF 2010 **Conference** aimed at advancing the evaluation of complex multilingual and multimodal information systems in order to support individuals, organizations, and communities who design, develop, employ, and improve such systems.

CLEF 2010 **Labs** continued the CLEF tradition of community-based benchmarking and complemented it with workshops on emerging issues in evaluation methodology. In fact, two different forms of labs were offered: labs could either be run as “campaign-style” during the twelve month period preceding the conference, or adopt a more “workshop-style” format that could explore issues of information access evaluation and related fields. A call was made for proposals for both types of labs, with 9 proposals being submitted in late October of 2009. Five of these were accepted as benchmarking activities, and two were accepted as exploration workshops. Each submitted proposal was reviewed by the CLEF 2010 lab organizing committee, with final decisions of acceptance and length of the labs. The labs culminated in sessions of a half-day, one full day or two days at the CLEF 2010 conference.



Figure 2: Different moments of CLEF 2010.

2 CLEF 2010 Conference

The first two days of the CLEF 2010 conference (20-21 September) were held in the “Aula Magna” (Main Assembly Hall) at the Palazzo del Bo, the main historical building of University of Padua.

Scientific papers on all aspects of multilingual and multimodal information access evaluation were requested for the conference. All submissions were reviewed on the basis of relevance, originality, importance, and clarity.

Relevant topics for the CLEF 2010 conference included:

- Novel methodologies for the design of evaluation tasks, especially user-centric ones;
- Alternative methods for improving and automating ground-truth creation;
- Task-oriented and easy to communicate metrics;
- Innovative techniques for analysing the experimental results, including statistical analyses, data mining, and information visualization;
- Infrastructures for bringing automation and collaboration in the evaluation process;
- Component-based evaluation approaches;
- Analysis of the impact of multilingual/multicultural/multimodal differences on interface and search design;
- Multilinguality and multimodality in relevant application communities – e.g. digital libraries, intellectual property, medical, music, and video domains – and social media.

A very wide and interdisciplinary program committee has been set up in order to review the papers and also to stimulate submission from areas traditionally not covered by CLEF. 74 members participated in the program committee with about 80% of the people coming from academia and 20% from industry; the geographical spread was broad as well: 60% comes from Europe, 20% from North America, 12% from Asia, 4% from Australia and New Zealand, 3% from South America, and 1 member from Africa. The PC members have a wide range of expertise, covering traditional topics of CLEF but also new ones: information retrieval (image IR, music IR, video IR, geographic IR, contextual IR) and evaluation methodologies (laboratory and user-centered); multilinguality; natural language processing; machine translation; automatic text categorization; question answering; XML and semi-structured data; log analysis; databases and IR; digital libraries; and, human-computer interaction and information visualization.

The papers accepted for the conference comprised research on resources, tools, and methods; experimental collections and datasets, and evaluation methodologies and metrics. Moreover, two keynote talks highlighted future directions for the experimental evaluation both from an academic and an industrial point of view:

- Norbert Fuhr, University of Duisburg-Essen, Germany, explored “IR Between Science and Engineering, and the Role of Experimentation”. In this talk, it is claimed that IR is still in its scientific infancy. Despite the extensive efforts in evaluation initiatives, the scientific insights gained are still very limited - partly due to shortcomings in the design of the testbeds. From a general scientific standpoint, using test collections for evaluation only is a waste of resources. Instead, experimentation should be used for hypothesis

generation and testing in general, in order to accumulate a better understanding of the retrieval process and to develop a broader theoretic foundation for the field.

- Ricardo Baeza-Yates, Yahoo! Research and Universitat Pompeu Fabra, Spain, discussed about “Retrieval Evaluation in Practice”. Nowadays, most research on retrieval evaluation is about comparing different systems to determine which is the best one, using a standard document collection and a set of queries with relevance judgements, such as TREC. However, in an industrial setting, reality is much harder. On the other hand, search engines have hundreds of millions of users and hence click-through data can and should be used for evaluation. This invited talk explored important problems that arise in practice. Some of them are: Which queries are already well answered and which are the difficult queries? Which queries and how many answers per query should be judged by editors? How we can use clicks for retrieval evaluation? What retrieval measure we should use? What is the impact of culture, geography or language in these questions? All these questions are not trivial and depend in each other, so we only give partial solutions. Hence, the main message to take away is that more research in retrieval evaluation is certainly needed.

The main conference program was interleaved with highlights and reports from the actual evaluation activities and the labs: overviews and descriptions of the labs have been given as well as reports from the other major evaluation initiatives worldwide, TREC in the US, NTCIR in Japan, INEX in Australia, ROMIP in Russia, and FIRE in India.

In order to further stimulate the discussion and involve the audience two panels were organized:

- “The Four Ladies of Experimental Evaluation”, involved four key researchers, namely: Donna Harman - National Institute of Standards and Technology (NIST), USA - Noriko Kando - National Institute of Informatics (NII), Japan - Mounia Lalmas - University of Glasgow, UK – and Carol Peters - Institute of Information Science and Technologies, Italian National Research Council (ISTI-CNR), Pisa, Italy, who have created and run the main evaluation initiatives of the last two decades. The goal of the panel was to present some of the main lessons that have been learned in over a decade of experimental evaluation and to promote discussion with respect to what the future objectives in this field should be.
- “A PROMISE for Experimental Evaluation”, was given by some of the partners of Participative Research labOratory for Multimedia and Multilingual Information Systems Evaluation (PROMISE), the EU FP7 Network of Excellence starting in conjunction with this first independent CLEF 2010 conference, and designed to support and develop the evaluation of multilingual and multimedia information access systems, largely through the activities taking place in Cross-Language Evaluation Forum (CLEF) today, and taking it forward in important new ways. The single most important step forward for multilingual and multimedia information access that PROMISE will work towards is to provide an open evaluation infrastructure in order to support automation and collaboration in the evaluation process.

All accepted papers, invited talks, and panels are published by Springer in their Lecture Notes for Computer Science (6360) series, which are attached to the paper version of this deliverable.

Note that, as conference proceedings, the LNCS publication was ready by the conference and was distributed to all the attendees. This is a further innovation with respect to the CLEF tradition where CLEF LNCS post-proceedings containing revised version of the papers of the working notes have been published almost one year after the workshop.



Figure 3: Front and back matter of the CLEF 2010 LNCS proceedings.

The conference Programme is given in Appendix C.

3 CLEF 2010 Labs and Workshops

The labs and workshops (22-23 September) were held in a different venue, at the Faculty of Statistical Sciences in Padua.

Five benchmarking evaluations were run as Labs in CLEF 2010:

CLEF-IP: it was a follow-on to a CLEF 2009 track whose aim was to investigate IR techniques for patent retrieval. There were two tasks in 2010: the Prior Art Candidate Search Task to find patent documents that are likely to constitute prior art to a given patent application, and the Classification Task which classified a given patent document according to the International Patent Classification system (IPC). The track utilizes a collection of more than 1.3 million patent documents derived from EPO (European Patent Office) sources. The collection covers English, French and German with at least 150,000 documents in each language. Relevance judgments are produced using the patent citations for the Prior Art Candidates search task and using the recorded classification codes for the Classification task.

The lab was organized by the Information Retrieval Facility (IRF) in Vienna.

ImageCLEF: it was the eighth running of this track in CLEF. There were four tasks in 2010:

- **Medical Retrieval:** as in 2008 and 2009, the database in 2010 used images and captions from the Radiology and Radiographic journals published by the Radiological Society of North America. Three sub-tasks were conducted within the auspices of the medical task: modality detection, image-based retrieval and case-based retrieval. 2010 was the first year that the modality detection task was held. The goal of this task was to detect the acquisition modality of the images in the collection using visual, textual or mixed methods. The goal of the image-based retrieval task was to retrieve an ordered set of images from the collection that best met the information need specified as a textual statement and a set of sample images, while the goal of the case-based retrieval task was to return an ordered set of articles (rather than images) that best meet the information need provided as a description of a case.
- **Photo Annotation:** the task poses the challenge of automated annotation of 93 visual concepts in Flickr photos. The participants were provided with a training set of 8,000 Flickr images including annotations, EXIF data and Flickr user tags and the test was performed on 10,000 Flickr images. It was differentiated between approaches considering solely visual information, approaches relying on textual information and multi-modal approaches.
- **Robot Vision Challenge:** the track addressed the problem of visual place classification, with a special focus on generalization. Participants were asked to classify rooms and areas of an office environment on the basis of image sequences captured by a stereo camera mounted on a mobile robot, under varying illumination conditions.
- **Wikipedia Retrieval:** using 237,000 Wikipedia images that cover diverse topics of interest and are associated with unstructured and noisy textual annotations in English, French, and German. The task provides a testbed for the system-oriented evaluation of multimedia information retrieval from a collection of Wikipedia images. The aim is to investigate retrieval approaches in the context of a large and heterogeneous collection of images (similar to those encountered on the Web) that are searched for by users with diverse information needs.

- The Lab was coordinated by the University of Applied Sciences Western Switzerland, Switzerland with the help of many other institutions.

PAN: a lab on uncovering plagiarism, authorship, and social software misuse, ran at CLEF for the first time, following three previous workshops at other conferences. It was sponsored by Yahoo Research and had two tasks, namely detection of plagiarism detection and Wikipedia vandalism. Plagiarism detection in text documents is a challenging retrieval task: today's detection systems are faced with intricate situations, such as obfuscated plagiarism or plagiarism within and across languages. Moreover, the source of a plagiarism case may be hidden in a large collection of documents, or it may not be available at all. Vandalism has always been one of Wikipedia's biggest problems. However, the detection of vandalism is done mostly manually by volunteers, and research on automatic vandalism detection is still in its infancy. Hence, solutions are to be developed which aid Wikipedians in their efforts. The Lab was jointly coordinated by the Bauhaus-Universität Weimar, the Universidad Politécnica de Valencia, the University of the Aegean, and Bar-Ilan University.

MLQA10: ResPubliQA was the eighth year for multilingual question answering in CLEF. Similar to the ResPubliQA version in CLEF2009, the exercise is aimed at retrieving answers to a set of 200 questions over EUROPARL and ACQUIS collections. Two tasks have been proposed this year: Paragraph Selection (PS) and Answer Selection (AS). The PS task consisted of extracting a relevant paragraph of text that satisfies completely the information need expressed by a natural language question. In the AS task, the exercise was to demarcate the shorter string of text corresponding to the exact answer supported by the entire paragraph.

The Lab was jointly coordinated by CELCT - Italy, UNED - Spain, and the University of Limerick - Ireland.

WePS: focused on person name ambiguity and person attribute extraction on Web pages and on Online Reputation Management (ORM) for organizations, again dealing with the problem of ambiguity for organization names and the relevance of Web data for reputation management purposes. This was the lab's first year at CLEF, following two previous workshops at other conferences. Given the top web search results for a person name, two tasks were addressed: a clustering task, which consists of grouping together web pages referring to the same person, and an extraction task, which consists of extracting salient attributes for each of the persons sharing the same name. Continuing the path of previous campaigns, this third evaluation aimed at merging both problems into one single task, where the system must return both the documents and the attributes for each of the different people sharing a given name.

The Lab was jointly coordinated by UNED - Spain, NYU - USA, Intelius Corp - USA, University of Illinois, Llorente & Cuenca - Spain.

The following two were run as Workshops in CLEF 2010:

CriES: it addressed the problem of multi-lingual expert search in social media environments. The main topics were multi-lingual expert retrieval methods, social media analysis with

respect to expert search, selection of datasets and evaluation of expert search results. Papers reporting on experiments or proposals for possible benchmarking activities were invited.

The Lab was jointly coordinated by Karlsruhe Institute of Technology - Germany, University of Koblenz - Germany, and Bielefeld University - Germany.

LogCLEF: it investigated the analysis and classification of queries in order to understand search behavior in multilingual contexts and ultimately to improve search systems. Another important long-term aim is to stimulate research on user behavior in multilingual environments and promote standard evaluation collections of log data. Two different log sets were used, the The European Library (TEL) logs, and the Deutscher Bildungsserver (DBS) logs, a quality controlled internet directory for educational resources. Participants were invited to investigate a variety of questions with the end goal of defining a benchmarking task for follow-on labs. LogCLEF differs from other evaluation tracks since its goal is not the production of a gold standard for a specific task, but to create a forum for the creative exploration of user behaviour based on logs.

The Lab was jointly coordinated by the University of Padua - Italy and the University of Hildesheim - Germany.

The results of the experiments conducted within CLEF 2010 evaluation campaigns were presented and discussed at the workshop, 22-23 September, Padua, Italy. The workshop plays an important role by providing the opportunity for all the groups that have participated in the evaluation campaign to get together comparing approaches and exchanging ideas. The Labs workshop were run in parallel sessions covering two days, while a general poster session was arranged at the end of the second day, where all posters from all the different Labs had the opportunity to present their works.

Each Lab had its own schedule and programme, but they were almost structured in two main parts: presentation of the main results and reports from the participants. Some Labs had also some important invited talks:

- Ellen Voorhees, NIST, USA. In her talk “Community Evaluations for Machine Reading” she provided a history of the Question Answering track, which was started at TREC. The TREC QA track reinvigorated the question answering research community, fostering extensive research on different question types and finding answers in different kinds of corpora. In recent years, the TAC QA track evolved into the Knowledge Base Population (KBP) track, where the task is (essentially) factoid-question answering combined with entity resolution: answer strings must be resolved into the appropriate nodes of a pre-existing ontology. KBP is itself viewed as an initial test of capability for Machine Reading systems. In this talk she reflected on what the lessons learned from the QA track suggest for creating a Machine Reading evaluation with the goal of creating a vibrant Machine Reading research community.
- Maarten de Rijke, University of Amsterdam, The Netherlands. His talk “Sentimental About People” is a mixture of descriptive and algorithmic aspects of searching for people. It started surveying search scenarios in which people are the target. In several use cases, something more is desired: not just people but also expressions of opinions

about people. The performance of current sentiment retrieval methods applied to people search was described and more details were given about ongoing work in which syntactic information in a bootstrapping fashion is used to improve the effectiveness of finding sentiments about people in social media.

- Heng Ji, NLP@The City University of New York, USA. She gave a talk on “Overview of the TAC2010 Knowledge Base Population Track”. The main goal of KBP is to promote research in discovering facts about entities and expanding a knowledge source. KBP2010 includes the following four tasks: (1). Regular Entity Linking, where names must be aligned to entities in the KB; (2). Optional Entity linking, without using Wikipedia texts; (3). Regular Slot Filling, which requires a system to automatically discover answers from the source document collection and return answers for a query entity with specified slots, and use them to expand KB; (4). Surprise Slot Filling, which requires a system to return answers regarding new slot types. KBP2010 has attracted many participants from Information Extraction (IE) and Question Answering (QA) communities (54 teams registered, 24 teams submitted results). Then a state-of-the-art slot filling system is presented which combines IE and QA approach, and exploits statistical answer Re-ranking and Markov Logic Networks for Cross-slot reasoning.

Each Lab also planned a final discussion session in order to give space to more in-depth exam of the results and where intentions for the future were expressed.

Overall, the workshops provided an ample panorama of the current state-of-the-art and the latest research directions in the multilingual information retrieval area.

More information on the CLEF 2010 Labs can be found in the electronic Notebook papers which are available online at the CLEF website (<http://clef2010.org/>).

Moreover, a book of abstract with ISBN has been prepared and distributed to all the attendees at the event.



Figure 4: Front and back matter of the CLEF 2010 "Book of Abstracts"

The Labs programmes are given in Appendix D.

4 Participation in CLEF 2010

In general, the event was attended by approximately 140 people from different academic and industrial institutions.

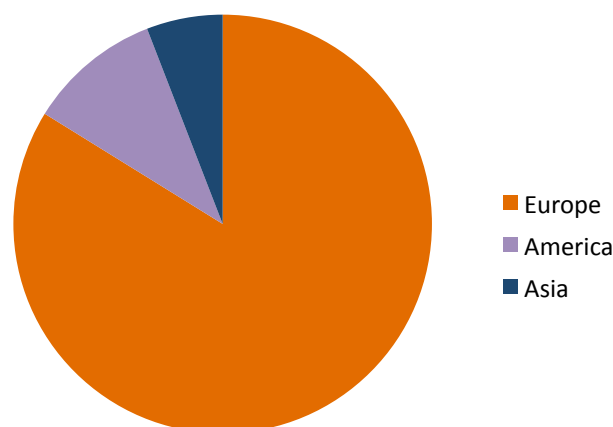


Figure 5: Participation according to the continents.

As figure 1 shows, a massive increase in the participation from Europe is registered this year, while the other continents remained more or less stable with respect to last years. The list of participants at the conference is given in Appendix E.

As for the Labs, about 200 groups initially registered showing interest in the CLEF evaluation campaigns. This is a great outcome which proves that CLEF has achieved a high visibility. About 110 people both researchers and system developers submitted runs and papers to the CLEF 2010 evaluation Labs and Workshops. Figure 2 shows the breakdown of participation per Lab: CLEF-IP 19; ImageCLEF 49; PAN 31; MLQA 24; WePS 21; CriES 7; LogCLEF 19. The most popular Lab at CLEF is ImageCLEF. This gives some cause for reflection as ImageCLEF is the Lab least concerned with multilinguality. However, recognising that multilingual information access is a truly multidisciplinary domain; one of the main objectives was to create a forum where groups working in many different areas (e.g. Information Retrieval, Natural Language Processing, Image Processing, Patents, and Log Analysis) have a chance to gather together, see what others are doing, and discuss and compare ideas.

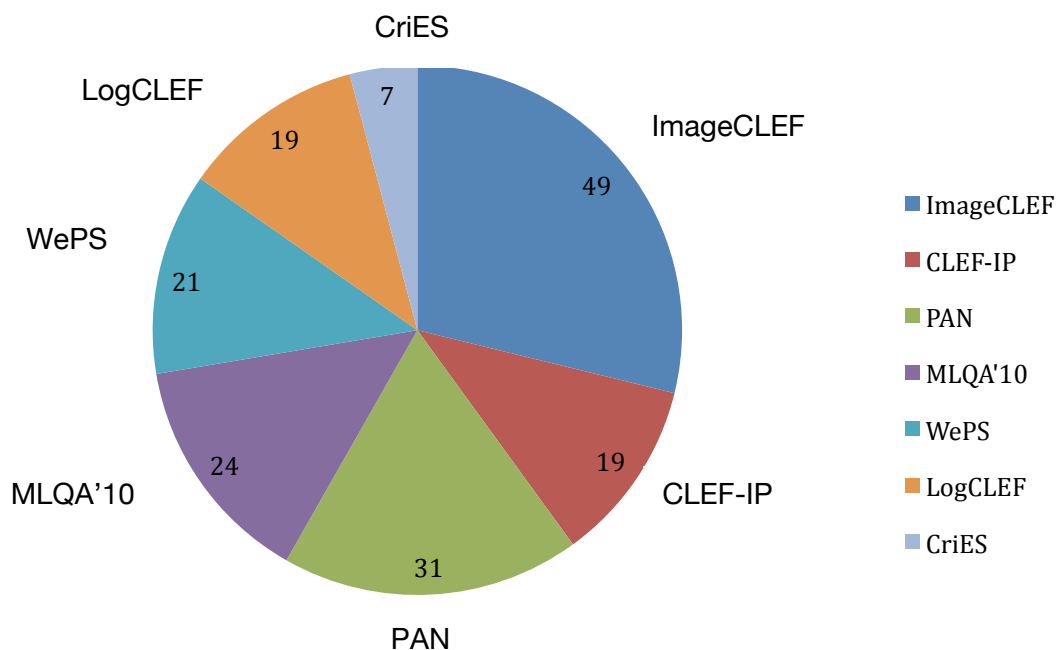


Figure 6: Breakdown of participation to the Labs.

Participation in the two new Labs at CLEF (PAN and WePS) is encouraging, while the CriES workshop attracted few groups.

A list of groups and indications of the tracks in which they participated is given in the Appendix F to this document.

5 CLEF 2010 Web Site and Advertising

5.1 CLEF 2010 Web Site

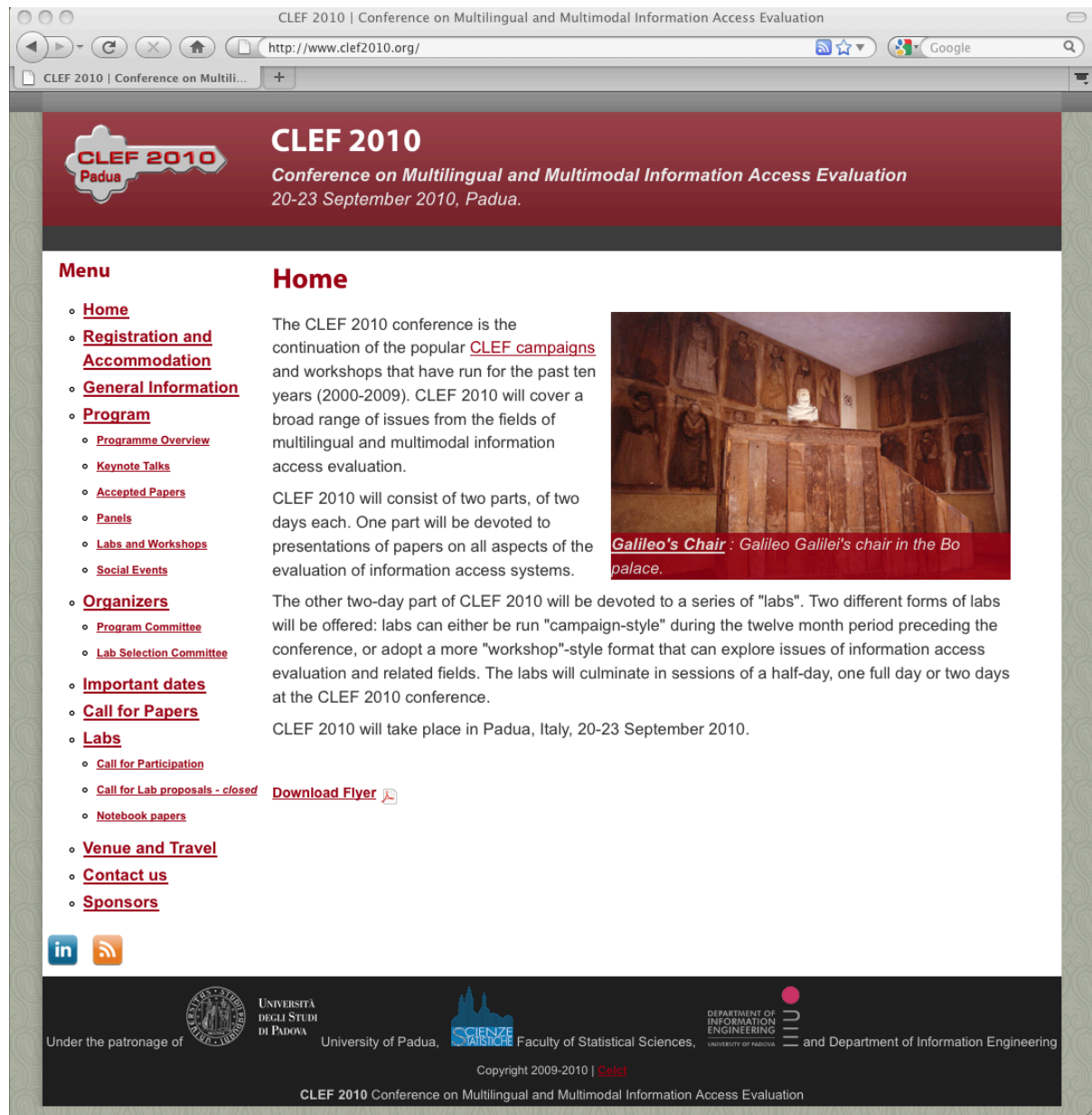


Figure 7: Home page of the CLEF 2010 Web site.

A dedicated Web site has been prepared by CELCT in order to provide all the necessary information about CLEF 2010 and promote the visibility of the event. The Web site is available at: <http://www.clef2010.org/>.

As explained in Section 3, the Web site contains the electronic version of all the papers submitted to the labs accessible under the “Labs → Notebook papers” menu.

Moreover, the slides of the presentations given during the conference and the labs are accessible and available for download on the CLEF 2010 Web site. Publications, both of the conference and of the Labs are linked to the bibliographic social networks BibSonomy or Citulike. The photos taken during the event have been posted to Flickr and tagged, at least, with the “CLEF 2010” and “PROMISE” tags. This photo album has been linked also from the CLEF 2010 Web site. Finally, the main sections (opening addresses, invited talks, other campaigns overviews, panels) have been shot and the videos have been posted to YouTube and tagged, at least, with the “CLEF 2010” and “PROMISE” tags. The videos have been linked also from the CLEF 2010 Web site¹.

5.2 Advertising

The CLEF 2010 Conference and Labs (call for labs, call for papers, call for participation) have been announced and advertised on several mailing lists:

- Evaluation Inivitatives
 - CLEF (Cross-Language Evaluation Forum community, Europe)
 - TREC (Text Retrieval Conference, USA)
 - TRECvid (TREC Video Retrieval Evaluation, USA)
 - NTCIR (NII Test Collection for IR Systems, Japan)
 - FIRE (Forum for Information Retrieval Evaluation, India)
- Research field news letters
 - SIG-IRList (ACM Special Interest Group on Information Retrieval (SIGIR) moderate newsletter)
 - DBWorld (ACM Special Interest Group on Management of Data (SIGMOD) moderated newsletter)
 - BCS-IRSG list (Information Retrieval Special Interest Group of the BCS, The Chartered Institute for IT, UK)
 - ASIS-L (American Society for Information Science, General List)
 - SIGDL-L (American Society for Information Science, Digital Libraries Interest Group)
 - DBItaly (Italian information management systems – databases, information retrieval systems, digital libraries, ... - moderated newsletter)
- European Projects
 - Europeana (<http://www.europeana.eu/>)
 - DL.org (<http://www.dlorg.eu/>)
 - TrebleCLEF (<http://www.trebleclef.eu/>)
 - DELOS (<http://www.delos.info/>)

A brief report on CLEF 2010 has been published in ERCIM News 83, issue of October 2010 (<http://ercim-news.ercim.eu/>) and it is reported in Appendix F.

¹ Some of these activities are still ongoing at the time of the writing of the present deliverable.

A more detailed report has been submitted to SIGIR Forum (<http://www.sigir.org/forum/>) and it will appear in the issue of December 2010.

A press release about CLEF 2010 has been published in the “Corriere del Veneto” (the attachment of the Italian Veneto Region – where Padua is located – to the “Corriere della Sera” the main Italian newspaper) on Tuesday, September 21st, at page 22.



Figure 8: Excerpt from "Corriere del Veneto" about CLEF 2010.

6 First Evaluation Initiatives Liaison Meeting

On Tuesday, September 24th, a liaison meeting with the organizers and representatives of the other main evaluation initiatives in the field has been organized. The objectives of the meeting were to explore whether there are shared themes and issues evaluation initiatives might be interested to discuss all together and for which they might envision some kind of medium/long term strategies where coordinated actions can produce higher benefits.

The attendees to the meeting were:

- Pavel Braslavski, Yandex, Russia, for ROMIP (Russian Information Retrieval Evaluation Seminar, Russia)
- Nicola Ferro, University of Padua, Italy, for CLEF (Cross-Language Evaluation Forum, Europe)
- Jaap Kamps, University of Amsterdam, The Netherlands, for INEX (Initiative for the Evaluation of XML Retrieval, Australia)
- Noriko Kando, NII, Japan, for NTCIR (NII Test Collection for IR Systems, Japan)
- Prasenjit Majumder, DAICT, India, for FIRE (Forum for Information Retrieval Evaluation, India)
- Ellen Voorhees, NIST, USA, for TREC (Text Retrieval Conference, USA)

The meeting has been an innovation and a step forward with respect to the past since, so far, evaluation initiatives have mostly been “good neighbours” due to the personal relationships among organizers and the reciprocal exchange of information but it has never

been attempted to explore the possibility of setting a common framework/strategy on some shared topics. Moreover, it was the first that CLEF has contacts with ROMIP and that all these initiatives met together at the same time (usually a subset of the organizers meets at the respective workshops and events).

The major outcome of the meeting is a common agreement on the need of further liaisons and cooperation on shared topics among evaluation initiatives, even if this is not trivial due to different organizational and funding constraints, as well as lack of specific funding for this kind of activity.

A first topic of cooperation has been identified: the need for a better description of the scientific data produced by the evaluation initiatives in terms of proper metadata formats.

This topic fits very well in the PROMISE agenda which is going to propose metadata formats for the evaluation data in the development of its open evaluation infrastructure. This will facilitate both the design of the infrastructure which will take into account a set of wider requirements and the adoption of the proposed formats which will be agreed also with the other evaluation initiatives.

The follow-ups of the meeting will be the set up of a Google group for carrying out the discussion in a coordinated way and the preparation of a position paper precisely defining the aims and goals of this activity.

7 CLEF 2011 and Beyond

CLEF 2011 will follow a similar format as CLEF 2010, namely conference plus labs and workshops, and will be organized by University of Amsterdam (UvA) in Amsterdam, The Netherlands, in September 2011. A call for labs proposals will be issued beginning of October 2010 and an initial call for papers will be issued by the end of the year.

For CLEF 2012 and onwards, the plan is to open a bid so that interested institutions and organizations will be able to apply and compete to organize and host the event.

8 Acknowledgements

We would like to warmly thank all the people who worked hard to make CLEF 2010 possible. We should also like to thank the members of the CLEF Steering Committee who have assisted us with their advice and suggestions.

References

- [Agosti et al. 2010] M. Agosti, N. Ferro, C. Peters, M. de Rijke, and A. Smeaton (Eds.). *Multilingual and Multimodal Information Access Evaluation*. Proceedings of the International Conference of the Cross-Language Evaluation Forum, CLEF 2010, Padua, Italy, September 2010, Lecture Notes in Computer Science (LNCS 6360) Springer 2010
- [Braschler et al. 2010] M. Braschler, D. Harman, and E. Pianta (Eds.) *CLEF 2010 LABs and Workshops, Notebook Papers*, 22-23 September, Padua, Italy. Web. September 2010
(<http://clef2010.org/index.php?page=papers/proceedings.php>)
ISBN 978-88-904810-0-0.

Appendix A: CLEF 2010 Organization

CLEF 2010 was organized by the Information Management Systems (IMS) research group of the Department of Information Engineering (DEI) of the University of Padua, Italy.

Honorary Chair

Carol Peters, Institute of Information Science and Technologies (ISTI), Italian National Research Council (CNR), Italy

General Chairs

Maristella Agosti, University of Padua, Italy

Maarten de Rijke, University of Amsterdam, The Netherlands

Program Chairs

Nicola Ferro, University of Padua, Italy

Alan Smeaton, Dublin City University, Ireland

Labs Chairs

Martin Braschler, Zurich University of Applied Sciences, Switzerland

Donna Harman, National Institute of Standards and Technology (NIST), USA

Resource Chair

Khalid Choukri, Evaluations and Language resources Distribution Agency (ELDA), France

Organization Chair

Emanuele Pianta, Center for the Evaluation of Language and Communication Technologies (CELCT), Italy

Appendix B: CLEF 2010 Labs and Workshops Coordination

It would be impossible to run the CLEF evaluation initiative and organize the annual workshops without considerable assistance from many groups. CLEF is organized on a distributed basis, with different research groups being responsible for the running of the various Labs. A list of the main institutions responsible for the coordination of the different CLEF 2010 Labs is given here below:

- Bar-Ilan University, Israel
- Bauhaus-Universität Weimar, Germany
- Bielefeld University, Germany
- Center for the Evaluation of Language and Communication Technologies (CELCT), Italy
- Computer Vision and Multimedia Laboratory, University of Geneva, Switzerland
- Department of Computer Science University of Limerick, Ireland
- Fraunhofer Institute for Digital Media Technology, Ilmenau, Germany
- Georgia Institute of Technology, Atlanta, GA, USA
- IDIAP Research Institute, Martigny, Switzerland
- Information Retrieval Facility, (IRF), Austria
- Intelius, Inc., Bellevue, Washington, USA
- Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
- Karlsruhe Institute of Technology, Germany
- Leiden Institute of Advanced Computer Science, Leiden University, The Netherlands
- Llorente & Cuenca, Madrid, Spain
- New York University, New York, USA
- NLE Lab, Universidad Politécnica de Valencia, Spain
- Oregon Health & Science University, Oregon, USA
- Royal Institute of Technology, Stockholm, Sweden
- TELECOM Bretagne, France
- NLP & IR Group, Universidad Nacional de Educación a Distancia, Spain
- University of Applied Sciences Western Switzerland (HES-SO), Switzerland

- University of Hildesheim, Germany
- University of Illinois at Chicago, USA
- University of Koblenz, Germany
- University of Padua, Italy
- University of the Aegean, Greece

Appendix C: CLEF 2010 Conference Programme

Monday, 20 September 2010, Aula Magna, Palazzo del Bo	
09:30–10:00	Welcome and Opening
10:00–11:00	Keynote Talk <i>IR Between Science and Engineering, and the Role of Experimentation</i> Chair: Maarten de Rijke, University of Amsterdam, The Netherlands Norbert Fuhr, University of Duisburg-Essen, Germany
11:00–11:30	Coffee Break
11:30–13:00	Resources, Tools, and Methods Chair: Johannes Leveling, Dublin City University, Ireland <i>A Dictionary- and Corpus-Independent Statistical Lemmatizer for Information Retrieval in Low Resource Languages</i> Aki Loponen, University of Tampere, Finland Kalervo Järvelin, University of Tampere, Finland <i>A New Approach for Cross-Language Plagiarism Analysis</i> Rafael Corezola Pereira, Universidade Federal do Rio Grande do Sul (UFRGS), Brazil Viviane P. Moreira, Universidade Federal do Rio Grande do Sul (UFRGS), Brazil Renata Galante, Universidade Federal do Rio Grande do Sul (UFRGS), Brazil <i>Creating a Persian-English Comparable Corpus</i> Homa Baradaran Hashemi, University of Tehran, Iran Azadeh Shakery, University of Tehran, Iran Heshaam Faili, University of Tehran, Iran
13:00–14:30	Lunch
14:30–15:30	Panel: A PROMISE for Experimental Evaluation Chair: Jussi Karlgren, Swedish Institute of Computer Science, Sweden Panelists: Nicola Ferro, University of Padua, Italy Allan Hanbury, Information Retrieval Facility (IRF), Austria Maarten de Rijke, University of Amsterdam, The Netherlands Giuseppe Santucci, Sapienza University of Rome, Italy
15:30–16:30	Experimental Collections and Datasets (1) Chair: Jaques Savoy, University of Neuchâtel, Switzerland

	Validating Query Simulators: An Experiment Using Commercial Searches and Purchases Bouke Huurnink, University of Amsterdam, The Netherlands Katja Hofmann, University of Amsterdam, The Netherlands Maarten de Rijke, University of Amsterdam, The Netherlands Marc Bron, University of Amsterdam, The Netherlands
	<i>Using Parallel corpora for Multilingual(Multi-Document) Summarisation Evaluation</i> Marco Turchi, European Commission - Joint Research Centre (JRC), Italy Josef Steinberger, European Commission - Joint Research Centre (JRC), Italy Ralf Steinberger, European Commission - Joint Research Centre (JRC), Italy Mijail Kabadjov, European Commission - Joint Research Centre (JRC), Italy
16:30–17:00	Coffee Break
17:00–18:30	Labs Overviews Chair: Henning Müller, University of Applied Sciences Western Switzerland, Switzerland
	ImageCLEF RespubliQA WePS CLEF-IP PAN
19:00–21:00	Reception

Tuesday, 21 September 2010, Aula Magna, Palazzo del Bo

09:30–10:00	Experimental Collections and Datasets (2) Chair: Theodore Kalamboukis, Athen University of Economics and Business, Greece
	<i>MapReduce for information retrieval evaluation: “Let's quickly test this on 12 TB of data”</i> Djoerd Hiemstra, University of Twente, The Netherlands Claudia Hauff, University of Twente, The Netherlands
	<i>Which Log for which Information? Gathering Multilinguality Data from Different Log File Types</i> Maria Gäde, Humboldt-Universität zu Berlin, Germany Vivien Petras, Humboldt-Universität zu Berlin, Germany Juliane Stiller, Humboldt-Universität zu Berlin, Germany
10:00–11:00	Other Evaluation Initiatives Chair: Nicola Ferro, University of Padua, Italy
	TREC, Ellen Voorhees, NIST, USA NTCIR, Noriko Kando, NII, Japan

	INEX, Jaap Kamps, University of Amsterdam, The Netherlands ROMIP, Pavel Braslavski, Yandex, Russia FIRE, Prasenjit Majumder, DAICT, India
11:00–11:30	Coffee Break
11:30–12:30	Invited Talk Chair: Maristella Agosti, University of Padua, Italy Ricardo Baeza-Yates, Yahoo! Research, Spain
12:30–13:00	Evaluation Methodologies and Metrics (1) Chair: Christa Womser-Hacker, University of Hildesheim, Germany
	<i>Examining the Robustness of Evaluation Metrics for Patent Retrieval with Incomplete Relevance Judgements</i> Walid Magdy, Dublin City University, Ireland Gareth Jones, Dublin City University, Ireland
	<i>On the Evaluation of Entity Profiles</i> Maarten de Rijke, University of Amsterdam, The Netherlands Krisztian Balog, University of Amsterdam, The Netherlands Toine Bogers, Royal School of Library & Information Science, Denmark Antal van den Bosch, Tilburg University, The Netherlands
13:00–14:30	Lunch
14:30–15:30	Panel: The Four Ladies of Experimental Evaluation Chair: Donna Harman, NIST, USA
	Panelists: Noriko Kando, NII, Japan Mounia Lalmas, University of Glasgow, UK Carol Peters, ISTI CNR, Italy
15:30–17:00	Evaluation Methodologies and Metrics (2) Chair: Ray Larson, University of California, Berkeley, USA
	<i>Evaluating Information Extraction</i> Andrea Esuli, ISTI-CNR, Italy Fabrizio Sebastiani, ISTI-CNR, Italy
	<i>Tie-breaking Bias: Effect of an Uncontrolled Parameter on Information Retrieval Evaluation</i> Guillaume Cabanac, Université de Toulouse, France Gilles Hubert, Université de Toulouse, France Mohand Boughanem, Université de Toulouse, France Claude Chrisment, Université de Toulouse, France
	<i>Automated Component-Level Evaluation: Present and Future</i> Allan Hanbury, Information Retrieval Facility, Austria

	Henning Müller, University of Applied Sciences Western Switzerland, Switzerland
17:00–17:30	<i>Logistics of Social Events</i>
18:00–23:00	<i>Walk around and Social Dinner</i>

Wednesday, 22 September 2010, Faculty of Statistical Sciences	
09:00–09:30	<i>Labs Kick-off</i>
09:30–11:00	<i>Parallel Lab Sessions</i>
	ImageCLEF CLEF-IP PAN
11:00–11:30	<i>Coffee Break</i>
11:30–13:00	<i>Parallel Lab Sessions</i>
	ImageCLEF CLEF-IP PAN
13:00–14:30	<i>Lunch</i>
14:30–16:00	<i>Parallel Lab Sessions</i>
	ImageCLEF CLEF-IP RespubliQA
16:00–16:30	<i>Coffee Break</i>
16:30–18:00	<i>Parallel Lab Sessions</i>
	ImageCLEF CLEF-IP RespubliQA
18:00–20:00	<i>Poster Session (with reception)</i>

Thursday, 23 September 2010, Faculty of Statistical Sciences	
09:00–09:30	<i>Reports out from the Labs</i>
	Chair: Martin Braschler, Zurich University of Applied Sciences, Switzerland
	CLEF-IP PAN

	RespubliQA
09:30–11:00	<i>Parallel Lab Sessions</i>
	ImageCLEF WePS LogCLEF
11:00–11:30	<i>Coffee Break</i>
11:30:–13:00	<i>Parallel Lab Sessions</i>
	ImageCLEF WePS LogCLEF
13:00–14:30	<i>Lunch</i>
14:30:–16:00	<i>Parallel Lab Sessions</i>
	ImageCLEF WePS CriES
16:00–16:30	<i>Coffee Break</i>
16:30:–18:00	<i>Parallel Lab Sessions</i>
	ImageCLEF WePS CriES
18:00–18:40	<i>Reports out from the Labs</i> Chair: Emanuele Pianta, CELCT, Italy
	ImageCLEF WePS LogCLEF CriES
18:40–19:00	<i>Closing</i>

Appendix D: CLEF 2010 Labs and Workshops Programme

ImageCLEF Lab, Workshop Program, 22-23 September (two-day workshop)

9:30-11:00 **Task overviews**

- General Overview, Henning Müller
- Wikipedia Retrieval, Theodora Tsikrika
- Medical Retrieval, Steven Bedrick

11.00- 11:30 Coffee Break

11:30-13:00 **Task overviews and feedback**

- Photo annotation, Stefanie Nowak
- Robot Vision, Barbara Caputo
- feedback from participants, ideas for 2011

13:00—14:30 Lunch

14:30-16:00 **Wikipedia retrieval** - Chair: Jana Kludas

- XRCE's Participation in Wikipedia Retrieval of ImageCLEF 2010, Stephane Clinchant, Gabriela Csurka, Julien Ah-Pine, Guillaume Jacquet, Florent Perronnin, Jorge Sanchez and Keyvan Minoukadeh
- UNT at ImageCLEF 2010: CLIR for Wikipedia Images, Miguel Ruiz, Jiangping Chen, Karthikeyan Pusapathy, Pok Chin and Ryan Knudson
- Document Expansion for Text-based Image Retrieval at WikipediaMM 2010, Jinming Min, Johannes Leveling and Gareth Jones
- Experiences at ImageCLEF 2010 using CBIR and TBIR mixing information approaches, Joan Benavent, Xaro Benavent Garcia, Esther DeVes, Ruben Granados and Ana Garcia-Serrano
- Discussion

16:00—16:30 Coffee Break

16:30-18:00 **The Medical Retrieval task** - Chair: Henning Müller

- Medical Case-based Retrieval by Leveraging Medical Ontology and Physician Feedback: UIUC-IBM at ImageCLEF 2010, Parikshit Sondhi
- IRT at ImageCLEF 2010: medical retrieval task, Duy Dinh
- Semantic Image Retrieval, Theodore Kalamboukis
- OHSU at ImageCLEF 2010, Steven Bedick

- MedGIFT at ImageCLEF 2010, Ivan Eggel

18:00-20:00 **Poster Session**

9:30-11:00 **The photo annotation task** - Chair: Stefanie Nowak

- Detection of Visual Concepts and Annotation of Images using Predictive Clustering Trees, Ivica Dimitrovski, Dragi Kocev, Suzana Loskovska, and Saso Dzeroski
- LEAR and XRCE's participation at the Visual Concept Detection Task - ImageCLEF 2010, Thomas Mensink, Gabriela Csurka, Florent Perronnin, Jorge Sánchez, Jakob Verbeek
- GMM based image classification on GPU, Balint Daroczy
- A Text-Based Approach to the ImageCLEF 2010 Photo Annotation Task, Li, W., Min, J., Jones, G.J.
- Discussion

11:00—11:30 Coffee Break

11:30-13:00 **The robot vision task** - Chair: Jesus Martinez

- Combination of Classifiers for Indoor Room Recognition, Walter Lucetti
- IDIAP at the robot vision task 2010, Jesus Martinez

13:00—14:30 Lunch

14:30-16:00 **ImageCLEF 2011: breakout and feedback**

- first analysis of the survey and global feedback (all)
- future of the wikipedia task (Jana Kludas)
- future of the medical retrieval task (Henning Müller)

16:00—16:30 Coffee Break

16:30-18:00 **ImageCLEF 2011: breakout and feedback**

- photo annotation task (Stefanie Nowak)
- robot vision task (Jesus Martinez)
- free discussion

=====

CLEF-IP Lab, Workshop Program, 22 September (one day workshop)

9:30—11:00 **Prior Art Search and Classification Session**

- W. Magdy, G. Jones, Applying the KISS principle for the CLEF-IP 2010 Prior Art Candidate Patent Search Task
- Julien GOBEILL, Douglas TEODORO, Emilie Pasche, Dina Vishnyakova, Patrick Ruch and Christian Lovis, Automatic prior art searching and patent encoding at CLEF-IP '10

- P. Lopez, L. Romary – Experiments with Citation Mining and Key-term Extraction for Prior Art Search

11:00—11:30 Coffee Break

11:30—13:00 Classification Session

- S. Verberne, Merijn Vogel and Eva D'hondt, with contributions from Jean Beney - Patent classification experiments with the Linguistic Classification System LCS
- J. Guyot, K. Benzineb, G. Falquet – myClass: A Mature Tool for Patent Classification
- F. Derieux, M. Bobeica, D. Pois, J.-P. Raysz – Combining Semantics and Statistics for Patent Classification

13:00—14:30 Lunch

14:30—16:00 Poster Session

16:00—16:30 Coffee Break

16:30—18:00 Panel Discussion

=====

PAN LAB, Workshop Program, 22 September (half day workshop)

Task 1 - **Plagiarism Detection**, Chair: Martin Potthast

9:30-9:45 Overview of the 2nd International Competition on Plagiarism

Detection, Martin Potthast, Alberto Barrón-Cedeño, Andreas Eiselt, Benno Stein, and Paolo Rosso

9:45-10:10 Improving the Reliability of the Plagiarism Detection System - Lab

Report for PAN at CLEF 2010, Jan Kasprzak and Michal Brandejs

10:10-10:35 External and Intrinsic Plagiarism Detection Using a Cross-Lingual

Retrieval and Segmentation System - Lab Report for PAN at CLEF 2010, Markus Muhr, Roman Kern, Mario Zechner, and Michael Granitzer

10:35-10:50 FastDocode: Finding Approximated Segments of N-Grams for

Document Copy Detection - Lab Report for PAN at CLEF 2010
Gabriel Oberreuter, Gaston L'Huillier, Sebastián A. Ríos, and Juan D. Velásquez

10:50-11:05 UFRGS@PAN2010: Detecting External Plagiarism - Lab Report for

PAN at CLEF 2010, Rafael Corezola Pereira, Viviane P. Moreira, and Renata Galante

11:00-11:30 Coffee Break

Task 2 - **Wikipedia Vandalism Detection**, Chair: Efstathios Stamatatos

11:30-11:45 Overview of the 1st International Competition on Wikipedia

Vandalism Detection, Martin Potthast, Benno Stein, and Teresa Holfeld

11:45-12:10 Wikipedia Vandalism Detection Through Machine Learning:

Feature Review and New Proposals - Lab Report for PAN at CLEF 2010,
Santiago Moisés Mola Velasco

12:10-12:35 Detecting Wikipedia Vandalism using WikiTrust - Lab Report for

PAN at CLEF 2010, B. Thomas Adler, Luca de Alfaro, and Ian Pye

12:35-12:50 Novel Balanced Feature Representation for Wikipedia Vandalism

Detection Task - Lab Report for PAN at CLEF 2010, István Hegedűs, Róbert
Ormándi, Richárd Farkas, and Márk Jelasity

11:50-12:05 Wiki Vandalysis - Wikipedia Vandalism Analysis - Lab Report for

PAN at CLEF 2010, Manoj Harpalani, Thanadit Phumprao, Megha Bassi,
Michael Hart, and Rob Johnson

=====

MLQA'10 Lab, Workshop Program, 22 September (half day workshop)

14.30 - 14.45 **Welcome and Additional Notes on ResPubliQA 2010**, Alvaro
Rodrigo, UNED, Spain

14.45 - 16.00 **Reports from ResPubliQA 2010**

- The LogAnswer Project at ResPubliQA 2010, Ingo Glöckner, FernUniversität in
Hagen, Germany
- Document Expansion for Cross-Lingual Passage Retrieval, Xabier Saralegi, R&D,
Elhuyar Foundation, Basque Country
- Question Answering on Romanian, English and French Languages, Adrian Iftene,
"Al.I.Cuza" University of Iasi, Romania
- Contextual Approach for Paragraph Selection in Question Answering Task, Hapnes
Toba, University of Indonesia, Indonesia

16.00 - 16.30 : Coffee break

16.30 - 17.00 **Invited talk:** Community Evaluations for Machine Reading, Ellen Voorhees,
NIST, USA

17.00 - 17.30 : **Panel - Position Papers**

- Question Answering for Machine Reading Evaluation: Alvaro Rodrigo, UNED, Spain
- Revamping Question Answering with a Semantic Approach over World Knowledge: Johannes Leveling, Dublin City University, Ireland

17.30 – 18.00 : **Final discussion, Wrap up and Conclusion**

=====

WePS-3 Lab, Workshop Program, 23 September (one day workshop)

9.30 – 11.00 **Task-1: Clustering session-** chair: Satoshi Sekine

- WePS overview (Julio Gonzalo)
- Task 1 overview (Javier Artiles)

Presentations by Participants

- Web Person Name Disambiguation by Relevance Weighting of Extended Feature Sets. Chong Long and Lei Shi
- Using Web Graph Structure for Person Name Disambiguation, Elena Smirnova, Konstantin Avrachenkov and Brigitte Trousse,
- Cross-document coreference for WePS. Iustin Dornescu, Constantin Orasan and Tatiana Lesnikova

11.00 – 11.30 Coffee Break

11.30 – 12.00 **Task-1: Clustering session-** chair: Satoshi Sekine

Presentations by Participants

- DAEDALUS at WePS-3 2010: k-Medoids Clustering using a Cost Function Minimization. Sara Lana-Serrano, Julio Villena-Román and José-Carlos González-Cristóbal.
- WePS3 person attribute extraction from the textual parts of Web pages. István Nagy and Richárd Farkas

12.00 – 13.00 **Invited Talk** session chair: Julio Gonzalo

Maarten de Rijke, Sentimental About People

13.00 14.30 Lunch

14.30 – 16.00 **Task-2: Online Reputation Management** session chair: Javier

Artiles

- Task-2 overview (Enrique Amigó)
- Mechanical Turk methodology (Andrew Borthwick)

Presentations by participants

- It was easy, when apples and blackberries were only fruits. Surrender Reddy Yerva, Zoltá, Miklós and Karl Aberer.
- ITC-UT:Tweet Categorization by Query Categorization of On-line Reputation Management. Minoru Yoshida, Shin Matsushima, Shingo Ono, Issei Sato, and Hiroshi Nakagawa.
- The University of Amsterdam in WePS3. Manos Tsagkias and Krisztian Balog
- Bootstrapping Websites for Classification of Organization Names on Twitter. Paul Kalmar.

16.00 – 16.30 Coffee Break

16.30 – 17.30 **Invited talk** session chair: Andrew Borthwick

Heng Ji, Overview of the TAC2010 Knowledge Base Population Track

17.30 – 18.00 **Discussion on the future of WePS** moderator: Andrew Borthwick

=====

CriES Workshop Program, 23 September (half day workshop)

14:30 Introduction and Overview of the CriES Pilot Challenge, Philipp Sorg

15:00 HITS and Misses: Combining BM25 with HITS for Expert Search,
Johannes Leveling and Gareth J. F. Jones

15:30 Identify Experts from a Domain of Interest, Adrian Iftene, Bogdan Luca,
Georgiana Cărauşu, and Madălina Merchez

16:00 – 16.30 Coffee Break

16:30 Multilingual Expert Search using Linked Open Data as Interlingual
Representation, Daniel M. Herzig and Hristina Taneva

17:00 Invited talk, Krisztian Balog

17:50 Wrap up and final discussion.

=====

LogCLEF Workshop Program, 23 September (half day workshop)

9:30 – 9:40 Thomas Mandl, Giorgio di Nunzio: Overview

9:40 – 10:00 Bosca, A.& Dini, L: Language Identification Strategies for Cross Language Information Retrieval

10:00 – 10:20 Perea-Ortega, J.; Montejo Ráez,, A.; Garcia Cumbreras, M. & Ureña-López, L.A.. SINAI at LogCLEF 2010

10:20 – 10:40 Stiller, J.; Gaede, M. & Petras V. Ambiguity of Queries and the Challenges for Query Language Detection.

10:40 – 11:00 Lana-Serrano, S.; Villena-Román, J. & González-Cristóbal, J-C. DAEDALUS at LogCLEF 2010: Analyzing the Success of Search Queries

11:00-11:30 Coffee Break

11:30 – 11:50 Verberne, S; Hinne, M.; van der Heijden, M; Hoenkamp, E.; Kraaij, W. & van der Weide, T. How does the Library Searcher behave?

11:50 – 12:10 Takaku, M.; Egusa, Y.; Saito, H.; Kando, N.; Teraki, H.; Miwa, M.. CRES at LogCLEF 2010: Towards Understanding the User Behaviors through an Analysis of Search Sessions, Search Units and Click Ranks.

12:10 – 12:30 Leveling, J.; Ghorab, M.R.; Magdy, W.; Jones, G. & Wade, V.: DCU-TCD@LogCLEF 2010: Re-ranking Document Collections and Query Performance Estimation.

12:30 – 13:00 Wrap-up: questions, discussions, ideas for future directions

Appendix E: List of participants at the Event

Surname	Name	Affiliation	Country
AGOSTI	MARISTELLA	UNIVERSITY OF PADUA	ITALY
AMIGO' CABRERA	ENRIQUE	UNED, MADRID	SPAIN
ARTILES	JAVIER	CITY UNIVERSITY OF NEW YORK	UNITED STATES OF AMERICA
ATZORI	MANFREDO	UNIVERSITY OF PADUA	ITALY
BAEZA-YATES	RICARDO	YAHOO! RESEARCH	SPAIN
BALOG	KRISZTIAN	UNIVERSITY OF AMSTERDAM	NETHERLANDS
BARADARAN HASHEMI	HOMA	UNIVERSITY OF TEHRAN	IRAN
BARRÓN-CEDEÑO	ALBERTO	POLYTECHNIC UNIVERSITY OF VALENCIA	SPAIN
BECKS	DANIELA	UNIVERSITY OF HILDESHEIM	GERMANY
BEDRICK	STEVEN	UNIVERSITY OF PORTLAND	UNITED STATES OF AMERICA
BENAVENT GARCIA	MARIA ROSER	UNIVERSITY OF VALENCIA	SPAIN
BERENDSEN	RICHARD	UNIVERSITY OF AMSTERDAM	NETHERLANDS
BERNINI	MARIA	UNIVERSITY OF PADUA	ITALY
BOBEICA	MIHAELA	JOUE, LENS	FRANCE
BORTHWICK	ANDREW	INTELIUS, INC., BELLEVUE	UNITED STATES OF AMERICA
BOSCA	ALESSIO		ITALY
BRASCHLER	MARTIN	UNIVERSITY OF ZURICH	SWITZERLAND
BRASLAVSKIY	PAVEL	YANDEX, MOSCOW	RUSSIA
BROTTO	DENIS	UNIVERSITY OF PADUA	ITALY

Surname	Name	Affiliation	Country
CABANAC	GUILLAUME	IRIT - UNIVERSITY OF TOULOUSE	FRANCE
CAMPORESE	ANTONIO	UNIVERSITY OF PADUA	ITALY
COREZOLA PEREIRA	RAFAEL	FEDERAL UNIVERSITY OF RIO GRANDE DO SUL	BRAZIL
CSURKA	GABRIELA	XEROX RESEARCH CENTRE EUROPE, MEYLAN	FRANCE
DAROCZY	BALINT ZOLTAN	MTA SZTAKI, HUNG, BUDAPEST	HUNGARY
DE ALFARO	LUCA	GOOGLE/UC SANTA CRUZ	USA
DE RIJKE	MAARTEN	UNIVERSITY OF AMSTERDAM	NETHERLANDS
DE CUEVA VES	ESTHER	UNIVERSITY OF VALENCIA	SPAIN
DEBOLE	FRANCA	CNR, PISA	ITALY
DERIEUX	FRANCK	JOUE, LENS	FRANCE
DI BUCCIO	EMANUELE	UNIVERSITY OF PADUA	ITALY
DI NUNZIO	GIORGIO MARIA	UNIVERSITY OF PADUA	ITALY
DIMITROVSKI	IVICA	IJS, LJUBLJANA	SLOVENIA
DINH	DUY	IRIT - UNIVERSITY OF TOULOUSE	FRANCE
DUSSIN	MARCO	UNIVERSITY OF PADUA	ITALY
EGGEL	IVAN	HES-SO VALAIS, SIERRE	SWITZERLAND
EISELT	ANDREAS	BAUHAUS UNIV. WEIMAR	GERMANY
FARKAS	RICHARD	UNIVERSITY OF SZEGED	HUNGARY
FERRO	NICOLA	UNIVERSITY OF PADUA	ITALY
FORNER	PAMELA	CELCT, TRENTO	ITALY
FUHR	NORBERT	UNIVERSITY OF DUISBURG-ESSEN	GERMANY
GAEDE	MARIA	UNIVERSITY OF BERLIN	GERMANY
GARCÍA	MIGUEL	UNIVERSITY OF JAÉN	SPAIN

Surname	Name	Affiliation	Country
CUMBRERAS	ÁNGEL		
GKANOIANNIS	ANESTIS	UNIVERSITY OF ATHENS	GREECE
GLÖCKNER	INGO	FERNUNIVERSITÄT IN HAGEN	GERMANY
GOBEILL	JULIEN	UNIVERSITY OF CAROUGE	SWITZERLAND
GONZALO	JULIO	UNED. MADRID	SPAIN
GUYOT	JACQUES	SIMPLE-SHIFT	SWITZERLAND
HANBURY	ALLAN	INFORMATION RETRIEVAL FACILITY, VIENNA	AUSTRIA
HARMAN	DONNA	NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, GAITHERSBURG	UNITED STATES OF AMERICA
HART	MICHAEL A.	STONY BROOK UNIVERSITY	UNITED STATES OF AMERICA
HENTSCHEL	CHRISTIAN	FRAUNHOFER HEINRICH HERTZ INSTITUTE	GERMANY
HERZIG	DANIEL M.	KARLSRUHE INSTITUTE OF TECHNOLOGY	GERMANY
HIEMSTRA	DJOERD	UNIVERSITY OF TWEENTE, ENSCHEDE	NETHERLANDS
HOLFELD	TERESA	BAUHAUS UNIV. WEIMAR	GERMANY
HUBERT	GILLES	IRIT - UNIVERSITY OF TOULOUSE	FRANCE
HUURNINK	BOUKE	UNIVERSITY OF AMSTERDAM	NETHERLANDS
IFTENE	ADRIAN	AL.I.CUZA UNIVERSITY OF IASI	ROMANIA
KALAMBOUKIS	THEODORE	UNIVERSITY OF ATHENS	GREECE
KALMAR	PAUL	KALMAR RESEARCH, LA MESA	UNITED STATES OF AMERICA
KAMPS	JAAP	UNIVERSITY OF AMSTERDAM	NETHERLANDS
KANDO	NORIKO	UNIVERSITY OF	JAPAN

Surname	Name	Affiliation	Country
		TOKYO	
KARLGREN	JUSSI	SICS, KISTA	SWEDEN
KASPRZAK	JAN	UNIVERSITY OF MASARYK, BRNO	CZECH REPUBLIC
KEKÄLÄINEN	JAANA	UNIVERSITY OF TAMPERE	FINLAND
KERN	ROMAN	KNOW-CENTER, GRAZ	AUSTRIA
KLUDAS	JANA	UNIVERSITY OF GENEVA	SWITZERLAND
LALMAS	MOUNIA	UNIVERSITY OF GLASGOW	UNITED KINGDOM
LANA-SERRANO	SARA	UNIVERSITY CARLOS III OF MADRID	SPAIN
LARSON	RAY	UNIVERSITY OF CALIFORNIA, BERKELEY	UNITED STATES OF AMERICA
LEVELING	JOHANNES	DUBLIN CITY UNIVERSITY	IRELAND
LI	WEI	DUBLIN CITY UNIVERSITY	IRELAND
LLOPIS	FERNANDO	UNIVERSITY OF ALICANTE	SPAIN
LOPONEN	AKI	UNIVERSITY OF TAMPERE	FINLAND
LUCETTI	WALTER	SCUOLA SUPERIORE DI STUDI SANT'ANNA, LA SPEZIA	ITALY
MADZIN	HIZMAWATI	UNIVERSITY OF MALAYA, KUALA LUMPUR	MALAYSIA
MAGDY	WALID	DUBLIN CITY UNIVERSITY	IRELAND
MAHDABI	PARVAZ	UNIVERSITY OF LUGANO	SWITZERLAND
MAJUMDER	PRASENJIT	DAIICT, INDIA	INDIA
MANDL	THOMAS	UNIVERSITY OF HILDESHEIM	GERMANY
MARTINEZ	JESUS	IDIAP RESEARCH INSTITUTE, MARTIGNY	SWITZERLAND
MASIERO	IVANO	UNIVERSITY OF PADUA	ITALY

Surname	Name	Affiliation	Country
MATSUSHIMA	SHIN	UNIVERSITY OF TOKYO	JAPAN
MBANYA	EUGENE MUKI	FRAUNHOFER HEINRICH HERTZ INSTITUTE	GERMANY
MELUCCI	MASSIMO	UNIVERSITY OF PADUA	ITALY
MIOTTO	RICCARDO	UNIVERSITY OF PADUA	ITALY
MOLA VELASCO	SANTIAGO MOISÉS	POLYTECHNIC UNIVERSITY OF VALENCIA	SPAIN
MORENO SCHNEIDER	JULIÁN	UNIVERSITY CARLOS III OF MADRID	SPAIN
MOTOHASHI	NAOKI	MEIJI UNIVERSITY, KANAGAWA	JAPAN
MUHR	MARKUS	KNOW-CENTER, GRAZ	AUSTRIA
MÜLLER	HENNING	HES-SO VALAIS, SIERRE	SWITZERLAND
NAGY	ISTVAN	UNIVERSITY OF SZEGED, RGAI	HUNGARY
NEMESKEY	DAVID	MTA SZTAKI, HUNG, BUDAPEST	HUNGARY
NOWAK	STEFANIE	FRAUNHOFER INSTITUTE FOR DIGITAL MEDIA TECHNOLOGY	GERMANY
OBERREUTER	GABRIEL	UNIVERSITY OF CHILE, SANTIAGO	CHILE
ORIO	NICOLA	UNIVERSITY OF PADUA	ITALY
PAL	DIPASREE	INDIAN STATISTICAL INSTITUTE, KOLKATA	INDIA
PALKOVSKII	YURII	ZHYTOMYR STATE UNIVERSITY	UKRAINE
PETERS	CAROL ANN	CNR, PISA	ITALY
PETRAS	VIVIEN	UNIVERSITY OF BERLIN	GERMANY
PIANTA	EMANUELE	CELCT, TRENTO	ITALY
PIROI	FLORINA	INFORMATION RETRIEVAL FACILITY, VIENNA	AUSTRIA

Surname	Name	Affiliation	Country
POTTHAST	MARTIN	UNIVERSITY OF WEIMAR	GERMANY
QUOCHI	VALERIA	CNR, PISA	ITALY
RODRIGO	ALVARO	UNED, MADRID	SPAIN
ROSSO	PAOLO	POLITECNICA VALENCIA	SPAIN
RUIZ	MIGUEL	UNIVERSITY OF NORTH TEXAS	UNITED STATES OF AMERICA
SANTUCCI	GIUSEPPE	UNIV. LA SAPIENZA - ROMA	ITALY
SARALEGI	XABIER	ELHUYAR FUNDAZIOA, USURBIL	SPAIN
SAVOY	JACQUES	UNIVERSITY OF NEUCHÂTEL	SWITZERLAND
SCHULTZ	ANTJE	UNIVERSITY OF KOBLENZ-LANDAU	GERMANY
SEBASTIANI	FABRIZIO	CNR, PISA	ITALY
SEKINE	SATOSHI	NEW YORK UNIVERSITY	USA
SILVELLO	GIANMARIA	UNIVERSITY OF PADUA	ITALY
SMIRNOVA	ELENA	INRIA, SOPHIA ANTIPOLIS	FRANCE
SONDHI	PARIKSHIT	UNIVERSITY OF ILLINOIS AT URBANA, CHAMPAIGN	UNITED STATES OF AMERICA
SORG	PHILIPP	KARLSRUHE INSTITUTE OF TECHNOLOGY	GERMANY
SPINA VALENTI	DAMIANO	UNED, MADRID	SPAIN
STAMATATOS	EFSTATHIOS	UNIVERSITY OF THE AEGEAN, KARLOVASSI	GREECE
STANEK	MICHAL	UNIVERSITY OF WROCLAW	POLAND
STEFANESCU	DAN	ROMANIAN ACADEMY, BUCHAREST	ROMANIA
STILLER	JULIANE	UNIVERSITY OF BERLIN	GERMANY
TARANTO	CLAUDIO	UNIV. DI BARI "ALDO	ITALY

Surname	Name	Affiliation	Country
		MORO"	
TEODORO	DOUGLAS	UNIVERSITE' DE GENEVE	SWITZERLAND
TOBA	HAPNES	UNIVERSITY OF INDONESIA	INDONESIA
TREVISAN	MARCO	CELI SRL	ITALY
TSIKRIKA	THEODORA	UNIVERSITY OF AMSTERDAM	NETHERLANDS
TURCHI	MARCO	EUROPEAN COMMISSION - JOINT RESEARCH CENTRE (JRC)	ITALY
UREÑA LÓPEZ	L. ALFONSO	UNIVERSITY OF JAÉN	SPAIN
VERBERNE	SUZAN	RADBOUD UNIVERSITY OF NIJMEGEN	NETHERLANDS
VILLENA-ROMÁN	JULIO	UNIVERSITY CARLOS III OF MADRID	SPAIN
VOORHEES	ELLEN	NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, GAITHERSBURG	UNITED STATES OF AMERICA
WANG	XIN	INTELIUS, INC., BELLEVUE	UNITED STATES OF AMERICA
WOMSER-HACKER	CHRISTA	UNIVERSITY OF HILDESHEIM	GERMANY
YERVA	SURENDER REDDY	EPFL, LAUSANNE	SWITZERLAND

Appendix F: List of Participating Institutions to the Labs

Affiliation	Country	CLEF-IP	Image	PAN	MLQA	WePS	CriES	Log CLEF
Aichi University of Education	Japan							x
Anna University Chennai	India			x				
Athens University of Economics and Business	Greece		x					
Barcelona Media Innovation Center	Spain			x				
Cairo University	Egypt		x					
CELCT	Italy				x			
Celi	Italy							x
Centrum Wiskunde & Informatica	The Netherlands	x						
CloudFlare, Inc.	USA			x				
CNRS Telecom ParisTech	France		x					
DAEDALUS, S.A.	Spain	x	x	x	x	x		x
DA-IICT	India			x				
Democritus University of Thrace	Greece		x					
Dhirubhai Ambani Institute of Information and Communication Technology	India				x			
EPFL	Switzerland					x		
ETH Zuurich	Switzerland		x					
Faculty of Electrical Engineering and Information Technology	Macedonia		x					
Fraunhofer Institute FIRST	Germany		x	x				
Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute	Germany		x					
Fujitsu Labs of America	USA			x				
Google, Inc.	USA			x				

Affiliation	Country	CLEF-IP	Image	PAN	MLQA	WePS	CriES	Log CLEF
Harbin Institute of Technology	China		x			x		
Humboldt Universität Berlin	Germany	x					x	x
Hungarian Academy of Sciences	Hungary		x	x	x			
IBM Thomas J. Watson Research Center	USA		x					
Idiap Research Institute	Switzerland		x					
I.E.S. "José Caballero"				x				
INRIA Sophia Antipolis	France					x		
Institute for Infocomm Research	Singapore	x	x	x				x
Jadavpur University	India				x	x		
JOUVE R&D	France	x						
Jozef Stefan Institute	Slovenia		x					
Kalmar Research	USA					x		
Karlsruhe Institute of Technology	Germany						x	
Know-Center Graz	Austria			x				
LEAR-INRIA	France		x					
LIG - Grenoble University	France		x					
LIMSI-CNRS	France				x			
Lister Hill National Center for Biomedical Communications	USA		x					
Masaryk University	Czech Republic			x				
Meiji University	Japan		x					
National Institute of Informatics	Japan							x
National Institute for Materials Science	Japan							x
National Institute for Educational Policy Research	Japan							x
National Polytechnic Institute	Mexico				x			
National University of Colombia	Colombia		x					
NIST	USA				x			

Affiliation	Country	CLEF-IP	Image	PAN	MLQA	WePS	CriES	Log CLEF
NLPR, CASIA	China		x					
Oregon Health & Science University	USA		x					
Otto-von-Guericke-University Magdeburg	Germany		x					
Priberam	Portugal				x			
Radboud University Nijmegen	The Netherlands	x						x
Ritsumeikan University	Japan		x					
R&D	France	x						
R&D, Elhuyar Foundation, Usurbil	Spain				x			
Romanian Academy	Romania				x	x		
School of Computing - Dublin City University	Ireland	x	x		x	x	x	x
Scuola Superiore di Studi e Perfezionamento Sant'Anna	Italy		x					
SEMPRIA	Germany				x			
Simple Shift	Switzerland	x						
South China University of Technology	China			x				
Stony Brook University	USA			x				
Taif Univeristy	Saudi Arabia			x				
Technische Universitat Munchen	Germany		x					
TELECOM Bretagne	France		x					
The Open University of Japan	Japan							x
The Robert Gordon University	UK		x					
Trinity College Dublin	Ireland						x	x
UNED	Spain	x	x		x	x	x	x
Universidad Carlos III de Madrid	Spain		x	x	x	x		x
Universidad de Huelva	Spain			x				
Universidad Politécnica de Madrid	Spain		x	x		x		x
Universidad Politécnica de Valencia	Spain	x			x			

Affiliation	Country	CLEF-IP	Image	PAN	MLQA	WePS	CriES	Log CLEF
Universidade Federal do Rio Grande do Sul	Brazil			x				
Universitas Indonesia	Indonesia	x		x	x			
Universität Koblenz Landau	Germany			x	x			
Universitat Politècnica de Catalunya	Spain					x		
Universitatea Politehnica din Bucuresti	Romania		x					
Université de Lyon	France	x						
University of Liège	Belgium		x					
Université Pierre et Marie Curie	France		x					
University of Alicante	Spain			x				
University of Amsterdam	The Netherlands	x	x	x		x	x	x
University of Applied Sciences	Switzerland	x	x					
University of California, Berkeley	USA		x					
University of California, Irvine	USA			x		x		
University of Chile	Chile			x				
University of Electronic Science and Technology of China	China		x					
University of Geneva	Switzerland	x	x					
University of Hagen	Germany				x			
University of Hildesheim	Germany	x						
University of Iasi	Romania	x	x	x	x	x	x	x
University of Illinois	USA		x					
University of Jaén	Spain		x			x		x
University of Lisbon	Portugal				x	x		x
University of North Texas	USA		x					
University of Sheffield	UK			x				
University of Szeged	Hungary			x		x		
University of the Basque Country	Spain				x			
University of Tokyo	Japan					x		

Affiliation	Country	CLEF-IP	Image	PAN	MLQA	WePS	CriES	Log CLEF
University of Toulouse	France		x					
University of Wolverhampton	UK	x		x	x	x		
Univ Paul Cézanne	France		x					
Univ du Sud-Toulon-Var	France		x					
USC-ISI	USA				x			
Wroclaw University of Technology	Poland		x					
Xerox Research Centre Europe	France		x					
Yahoo! Beijing R&D	China					x		
Zhytomyr State University	Ukraine			x				

Appendix G: CLEF 2010 Sponsors

Beyond the support of PROMISE, for the first time, CLEF 2010 has also benefited from the sponsorship of several organizations and institutions:

- Google, Switzerland
- Xerox Research Centre Europe (XRCE), France
- Information Retrieval Facility (IRF), Austria
- CELI Language and Information Technology, Italy
- Associazione Italiana per l'Informatica ed il Calcolo Automatico (AICA), Italy
- Department of Information Engineering, University of Padua, Italy
- University of Padua, Italy
- Consorzio per la Formazione e la Ricerca in Ingegneria dell'Informazione in Padova (COFRIDIP), Italy



CLEF 2010
Conference on Multilingual and Multimodal Information Access Evaluation
20-23 September 2010, Padua.

Menu

- [Home](#)
- [Registration and Accommodation](#)
- [General Information](#)
- [Program](#)
 - [Programme Overview](#)
 - [Remote Tests](#)
 - [Accepted Papers](#)
 - [Panels](#)
 - [Labs and Workshops](#)
 - [Social Events](#)
- [Organizers](#)
 - [Program Committee](#)
 - [Lab Selection Committee](#)
- [Important dates](#)
- [Call for Papers](#)
- [Labs](#)
 - [Call for Participation](#)
 - [Call for Lab group calls - closed](#)
 - [Webbook papers](#)
- [Venue and Travel](#)
- [Contact us](#)
- [Sponsors](#)

Sponsors

DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

xerox

UNIVERSITÀ DEGLI STUDI DI PADOVA

Google

promise

INFORMATION RETRIEVAL FACILITY

COFRIDIP
UNIVERSITÀ DI PADOVA

CELI
LANGUAGE & INFORMATION TECHNOLOGY

AICA

Under the patronage of
UNIVERSITÀ DEGLI STUDI DI PADOVA
University of Padua, **SCIENCE** Faculty of Mathematical Sciences, and Department of Information Engineering

Copyright 2010-2011 | [Legal](#)
CLEF 2010 Conference on Multilingual and Multimodal Information Access Evaluation

Appendix F: ERCIM News About CLEF 2010

CLEF 2010: Innovation, Science, Experimentation

by Nicola Ferro

Since 2000 the Cross-Language Evaluation Forum (CLEF) has played a successful role in stimulating research and promoting evaluation in a wide range of key areas in the information retrieval domain, becoming well-known in the international IR community. The results have been presented and discussed at annual workshops in conjunction with the European Conference for Digital Libraries.



CLEF 2010 represented a radical innovation of this traditional format. The aim is to continue with the benchmarking evaluation experiments, but also to provide more space for analysis, reflection and discussion. It was thus organised as an four-day event: a peer-reviewed conference followed by a series of laboratories and workshops.

CLEF 2010: The Conference

The conference aimed at advancing research into the evaluation of complex multimodal and multilingual information systems. Topics covered included experimental collections and datasets, evaluation methodologies, resources and tools. Two keynote talks by Norbert Fuhr, University of Duisburg-Essen and Ricardo Baeza-Yates, Yahoo! Research and Universitat Pompeu Fabra discussed future directions for experimental evaluation from academic and industrial perspectives. Reports from other major evaluation initiatives were presented: the Text REtrieval Conference, USA, the NII-NACSIS Test Collection for IR Systems, Japan, the INitiative for the Evaluation of XML Retrieval, Australia, the Information Retrieval Evaluation Seminar, Russia, and the Forum for Information Retrieval Evaluation in India. There were also two panels. The first presented the objectives of a new EU FP7 Network of Excellence, PROMISE, which aims at advancing the experimental evaluation of complex multimedia and multilingual information systems. In the second, researchers

responsible for creating and running some of the main evaluation initiatives of the last two decades, discussed what has been achieved so far by these initiatives and, more importantly, what still needs to be done.

CLEF 2010: The Labs and The Workshops

The laboratories continue and expand on the traditional CLEF track philosophy. Two different forms of labs are offered: benchmarking activities proposing evaluation tasks and comparative performance analyses, and workshop-style labs that explore issues of information access evaluation and related fields.

There were five benchmarking activities:

- CLEF-IP '10, sponsored by the Information Retrieval Facility (IRF) in Vienna investigated patent retrieval in a multilingual context.
- ImageCLEF 2010 offered several tasks for both context and content-based retrieval of medical images from the journal of Radiology and Radiographics and photographic images from Flickr and Wikipedia.
- PAN, sponsored by Yahoo Research, investigated the detection of plagiarism and of Wikipedia vandalism.
- QA@CLEF 2010 used the Corpus of the European parliament and offered monolingual question answering tasks for English, French, German, Italian, Portuguese, Spanish and Romanian.
- WePS (Web People Search) focused on person name ambiguity and person

attribute extraction and online reputation management

And two exploratory workshops:

- CriES addressed the problem of multi-lingual expert search in social media environments.
- LogCLEF studied search behaviour in a multilingual context via query analysis and classification.

CLEF 2010 was hosted by the Department of Information Engineering, University of Padua, Italy and was partially supported by the PROMISE project. Approximately 130 researchers participated in the event; with most of them staying for the full four days. All the presentations given at the conference and in the workshops can be found on the CLEF 2010 website

CLEF 2011

CLEF 2011 will be hosted by the University of Amsterdam in September 2011. A Call for Lab proposals will be issued at the beginning of October 2010 and an initial Call for Papers will be released in December 2010.

Links:

CLEF 2010: <http://www.clef2010.org/>
PROMISE: <http://www.promise-noe.eu/>

Please contact:

Nicola Ferro
University of Padua, Italy
E-mail: ferro@dei.unipd.it