

CLEF 2010

Conference on Multilingual and Multimodal Information Access Evaluation

Maristella Agosti
University of Padua, Italy
agosti@dei.unipd.it

Martin Braschler
Zürich University of Applied Sciences, Switzerland
bram@zhaw.ch

Khalid Choukri
ELDA, France
choukri@elda.org

Nicola Ferro
University of Padua, Italy
ferro@dei.unipd.it

Donna Harman
NIST, USA
donna.harman@nist.gov

Carol Peters
ISTI-CNR, Italy
carol.peters@isti.cnr.it

Emanuele Pianta
CELCT, Italy
pianta@fbk.eu

Maarten de Rijke
University of Amsterdam, The Netherlands
derijke@uva.nl

Alan Smeaton
Dublin City University, Ireland
alan.smeaton@dcu.ie

1 Introduction

The CLEF 2010 Conference on Multilingual and Multimodal Information Access Evaluation¹ was held at the University of Padua, Italy, September 20–23, 2010. 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.

Since 2000, *Cross-Language Evaluation Forum (CLEF)* has played a leading role in stimulating research into a wide range of key areas in the information retrieval domain, including cross-language question answering, image and video search, and interactive retrieval. It has also promoted the study and implementation of evaluation methodologies for diverse IR tasks and media. As a result, CLEF has been extremely successful in building a strong, multidisciplinary

¹<http://www.clef2010.org/>

research community, covering a wide range of expertise. This constantly growing and almost completely voluntary community has dedicated an enormous amount of effort to making CLEF happen and is at the core of the CLEF achievements.

CLEF 2010 represented a radical innovation of the “classic CLEF” format and an experiment aimed at understanding how “next generation” evaluation campaigns could be structured. The challenge was to innovate CLEF while still preserving its traditional core business, namely the benchmarking activities carried out in the various tracks and tasks.

After lively and community-wide discussions, the consensus was to make CLEF an independent four days event, no longer organised as a two and half day workshop in conjunction with the European Conference on Research and Advanced Technology for Digital Libraries (ECDL). CLEF 2010 thus consisted of two main parts: a peer-reviewed conference – the first two days – and a series of so-called lab sessions – the last two days. Two forms of labs were envisaged: benchmarking activities and a more “workshop”-style format aimed at exploring issues of information access evaluation and related fields.

In order to experiment with this new format, CLEF renewed its organizational structure. General Chairs were responsible for the overall event, Program Chairs took care of the conference program, and Lab Chairs managed the selection of labs. The Resource Chair addressed issues concerning the licensing and intellectual property rights of the experimental collections and resources made available during the evaluation activities, and the Organization Chair looked after the actual running of the labs, the preparation of the publications related to the event, and the event Web site. A new statute governing the future organization of CLEF, the election of committee members and the selection of Conference and Lab Chairs is now being prepared, and will be released shortly for approval by the CLEF community.

Approximately 140 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.

2 CLEF 2010: The Conference

The conference aimed at advancing research into the evaluation of complex multimodal and multilingual information systems.

Papers were solicited that explored needs and practices for information access, studied new evaluation metrics and methodologies, and discussed and proposed new directions for future activities in multilingual and multimodal *Information Retrieval (IR)* system evaluation. In addition, we encouraged the submission of papers that analyse some of the achievements of 10 years of CLEF through in-depth experiments using existing CLEF collections in interesting ways. A large international program committee, representing not only the multidisciplinary areas which have traditionally been part of CLEF but also covering neighboring areas such as information visualization, was established in order to stimulate the submission of innovative papers.

We received 21 submissions (17 full papers and 4 short papers) and had, on average, 4 reviews for each paper. 12 papers were accepted (8 full papers and 4 short papers) with an overall acceptance rate of 57%. The topics covered included research on resources, tools, and methods; experimental collections and datasets, and evaluation methodologies. Two invited keynote talks highlighted future directions for experimental evaluation 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”, while Ricardo Baeza-Yates (Yahoo! Research and Universitat Pompeu Fabra, Spain) talked about “Retrieval Evaluation in Practice”.

The conference program was structured to ensure that CLEF 2010 was a single integrated event and not simply the juxtaposition of a conference plus workshops. The first two days included overviews of the lab activities and reports from other major worldwide evaluation initiatives: Ellen Voorhees (National Institute of Standards and Technology, USA) presented *Text REtrieval Conference (TREC)* in the US; Noriko Kando (National Institute of Informatics, Japan) introduced *NII-NACSIS Test Collection for IR Systems (NTCIR)* in Japan; Jaap Kamps (University of Amsterdam, The Netherlands) described *INitiative for the Evaluation of XML Retrieval (INEX)*, held this year in Australia; Pavel Braslavski (Yandex, Russia) talked about *Russian Information Retrieval Evaluation Seminar (ROMIP)* in Russia; and, finally, Prasenjit Majumder (DAICT, India) reported on the activities of *Forum for Information Retrieval Evaluation (FIRE)* in India.

In order to further stimulate discussion and involve the audience, two panels were also organised. The first panel, “The Four Ladies of Experimental Evaluation,” saw four key researchers – Donna Harman, Noriko Kando, 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 some of the main retrieval evaluation initiatives of the last two decades, discussing what has been achieved so far and what still needs to be done by evaluation campaigns. The second panel, “A PROMISE for Experimental Evaluation,” was presented by the partners of the *Participative Research labOratory for Multimedia and Multilingual Information Systems Evaluation (PROMISE)* project, an EU FP7 Network of Excellence. PROMISE aims at advancing the experimental evaluation of complex multimedia and multilingual information systems in order to support the decision making process of individuals, commercial entities, and communities who develop, employ, and improve such complex systems.

3 CLEF 2010: The Lab Sessions

The laboratories continued and expanded on the traditional CLEF track philosophy. Two forms of labs were offered: benchmarking activities proposing evaluation tasks and comparative performance analyses, and workshop-style labs that explored issues of information access evaluation and related fields. A lab selection committee reviewed the lab proposals and decided on those to accept for CLEF 2010. Criteria for selection included soundness of methodology, feasibility of task, use case, existence of business case/industrial stakeholders, number of potential participants, clear movement along a growth path, scale of experiments, interdisciplinarity, originality, and reusability of results. The results of the activities of the labs are reported in the working notes, available on-line [2]. It is hoped that they will also give rise to post-conference publications in journals or separate volumes.

Five benchmarking activities were organized.

- **CLEF-IP '10**, sponsored by the Information Retrieval Facility (IRF) in Vienna was a follow-on of a CLEF 2009 track. There were two tasks in 2010: the Prior Art Candidate Search Task, which aimed at retrieving 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 (IPC) codes.

- **ImageCLEF 2010** was the eighth edition of this track, which began in CLEF 2003. There were four tasks in 2010: Medical Retrieval of images from articles published in Radiology and Radiographics, Photo Annotation of a MIR Flickr database of consumer photos with multiple annotations, Robot Vision Challenge, and Wikipedia Retrieval using images that cover diverse topics of interest and are associated with unstructured and noisy textual annotations in English, French, and German.
- **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 the detection of plagiarism and the detection of Wikipedia vandalism.
- **QA@CLEF 2010 - ResPubliQA** was the eighth year for multilingual question answering in CLEF. Similar to the ResPublicQA version in CLEF 2009, the lab used the Europarl Corpus, and had seven monolingual tasks for English, French, German, Italian, Portuguese, Spanish and Romanian.
- **WePS (Web People Search)** 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.

Two exploratory pilot workshops were also organized.

- **CriES** addressed the problem of multilingual expert search in social media environments. The main topics were multilingual 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.
- **LogCLEF** investigated the analysis and classification of queries in order to understand search behavior in multilingual contexts and ultimately to improve search systems. Two different datasets were available: logs from The European Library (TEL), and logs from the Deutscher Bildungsserver (DBS), 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 future labs.

4 CLEF 2011 and Beyond

CLEF 2011 will follow a similar format to CLEF 2010, namely a conference plus labs and workshops, and will be organized by the University of Amsterdam in Amsterdam, The Netherlands, in September 2011. A Call for Lab Proposals will be issued in October 2010 and an initial Call for Papers will be released by the end of the year.

For CLEF 2012 and onwards, the plan is to organise an open bidding process so that interested institutions and organizations will be able to apply to organize and host the event.

In the next few years, we hope to see mutual feed-back and cross-fertilization between conference and labs. For example, after additional analyses and studies, methodologies developed or results achieved in the benchmarking activities one year may give rise to papers in the conference the next year. Alternatively, topics discussed at the conference one year may lead to the organisation of a lab in the following year.

Our main goal is to encourage the interaction between science and engineering in our domain, around a shared experimental methodology, thus helping the CLEF community to grow and progress so that it is ready to face the challenges of an increasingly complex and challenging field.

Acknowledgments

CLEF 2010 was partially supported by the EU FP7 PROMISE Network of Excellence, grant agreement No. 258191.

The event also received the sponsorship of several organizations and institutions: Google, Switzerland; Xerox Research Centre Europe (XRCE), France; Information Retrieval Facility (IRF), Austria; CELI Language Information Technology, Italy; Department of Information Engineering, University of Padua, Italy; University of Padua, Italy; and, Consorzio per la Formazione e la Ricerca in Ingegneria dell'Informazione in Padova (COFRIDIP), Italy.

References

- [1] Maristella Agosti, Nicola Ferro, Carol Peters, Maarten de Rijke, and Alan F. Smeaton, editors. *Multilingual and Multimodal Information Access Evaluation, International Conference of the Cross-Language Evaluation Forum, CLEF 2010, Padua, Italy, September 20-23, 2010. Proceedings*, volume 6360 of *Lecture Notes in Computer Science*. Springer, 2010.
- [2] Martin Braschler, Donna Harman, and Emanuele Pianta, editors. *CLEF 2010 LABs and Workshops, Notebook Papers, 22-23 September 2010, Padua, Italy, 2010*. <http://www.clef2010.org/index.php?page=papers/proceedings.php>