CLEF 2011
Conference on Multilingual and Multimodal Information Access Evaluation

Paul Clough
University of Sheffield, UK
p.d.clough@sheffield.ac.uk

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

Pamela Forner
CELCT, Italy
forner@celct.it

Julio Gonzalo
UNED, Spain
julio@lsi.uned.es

Bouke Huurnink
University of Amsterdam, The Netherlands
bhuurnink@uva.nl

Jaana Kekäläinen
University of Tampere, Finland
Jaana.Kekalainen@uta.fi

Mounia Lalmas
Yahoo! Research, Spain
mounia@acm.org

Vivien Petras
Humboldt University Berlin, Germany
vivien.petras@ibi.hu-berlin.de

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

1 Introduction

The CLEF 2011 Conference on Multilingual and Multimodal Information Access Evaluation was held at the University of Amsterdam, the Netherlands, September 19–22, 2011. CLEF 2011 was organized by the Intelligent Systems Laboratory Amsterdam (ISLA) of the Informatics Institute at the University of Amsterdam.

Since 2000 the Cross-Language Evaluation Forum (CLEF) has played a leading role in stimulating research and innovation in a wide range of key areas in the domain of information retrieval and search. It has become a landmark in the annual research calendar of the international Information Retrieval and Search community. Through the years, CLEF has promoted the study and implementation of evaluation methodologies for diverse types of retrieval task and search scenario. As a result, a broad, strong, multidisciplinary and international research community

has been created, which covers and spans the different areas of expertise needed to deal with the evaluation of solutions to challenging information tasks.

Until 2010, the outcomes of experiments carried out under the CLEF umbrella were presented and discussed at annual workshops in conjunction with the European Conference on Research and Advanced Technology for Digital Libraries (renamed in 2011 as the International Conference on Theory and Practice of Digital Libraries). CLEF 2010 [1] represented a radical departure from this "classic" CLEF format. While preserving CLEF's traditional core business and goals, namely benchmarking activities carried out in various tracks, we complemented these activities with a peer-reviewed conference component aimed at advancing research in the evaluation of complex information systems for cross-language tasks and scenarios. CLEF 2010 was thus organized as an independent four-day event consisting of two main parts, a peer-reviewed conference followed by a series of laboratories and workshops.

CLEF 2011 continued to implement this new format, with keynotes, contributed papers and lab sessions, but with some changes. First, in this year's event we interleaved the conference presentations and the laboratories over a three and a half day period. Second, we added "community sessions" aimed at creating awareness of funding opportunities, offering networking opportunities, and infrastructure developed to help support large-scale retrieval experiments.

Overall, CLEF 2011 was attended by more than 170 people from different academic and industrial institutions; an increase of over 20% from last year. The majority of participants came from Europe, but attendance from all around the world increased this year, in particular from the United States with 15 participants. During the closing session participants were asked to fill in a short online SurveyMonkey questionnaire indicating their thoughts about the format of the event, with 3 multiple choice questions and room for text comments [1]. A total of 47 participants (27%) responded. The results of the survey indicated that the CLEF community largely supported the new, hybrid format for the conference which interleaved the usual lab (formerly track) sessions with peer-reviewed paper presentations, community sessions and keynote speeches. We found doing an online survey during the closing session to be a rapid and effective method of obtaining participant feedback, and would recommend this to the organizers of other conferences.

2 CLEF 2011: The Conference

An initial call for papers was issued before the end of 2010. Overall, 23 papers were submitted (19 full and 4 short), a slight increase with respect to last year. In total, 14 papers (10 full and 4 short) were finally accepted with an overall acceptance rate of 61%. The accepted papers and the abstracts of the invited talks were published by Springer in their Lectures Notes for Computer Science (LNCS) series [2].

This year, the papers accepted for the conference included research on evaluation methods and settings, natural language processing within different domains and languages, multimedia, and reflections on CLEF. Most of the papers (10 out of 14) were experimental studies exploiting standard IR collections (e.g. ImageCLEF 2010, FIRE 2010, TREC 7&8, WEPS-3), but also non-IR and own collections, as well as web, were utilized as a test environment. Two papers discussed evaluation metrics with no empirical setting, one paper proposed a new framework for experiments

---

[1] http://tinyurl.com/clef2012

and one was a bibliometric analysis.

Two keynote speakers highlighted important developments in the field of evaluation. Elaine Toms, from the University of Sheffield (UK), focused on the role of users in evaluation. In her talk titled "Would you trust your IR system to choose your date? Re-thinking IR Evaluation in the 21st Century", she argued that at present, evaluation has moved from an emphasis on topical relevance, to an emphasis on measuring almost anything that can be quantified. The talk examined some of the pitfalls in existing approaches, and discussed the issues involved in designing more effective evaluation approaches for assessing interactive IR systems. Omar Alonso, from the Microsoft Corporation, California (USA) presented a framework for the use of crowdsourcing experiments in the setting of retrieval evaluation. In his talk titled "Crowdsourcing for Information Retrieval Experimentation and Evaluation", he explored some directions that may influence the outcome of a task and presented a framework for conducting crowdsourcing experiments making some emphasis on a number of aspects that should be of importance for all sorts of IR-like tasks.

The community session at CLEF 2011 was organized around a strategic EU meeting, a networking session organized by the Chorus Network of Excellence, an *Evaluation Initiatives* session with overviews from other benchmarking fora, and an infrastructure session dedicated to the DIRECT system for handling and visualizing data resulting from retrieval experiments.

3   CLEF 2011: The Lab Sessions

The CLEF Labs are a continuation of tracks from previous CLEF workshops. In 2010, CLEF went from being a workshop collocated with an existing conference to a conference in its own right. The CLEF Labs are an integral part of the conference and two different types of lab are offered: (1) benchmarking or "campaign-style" and (2) workshop-style. The benchmarking labs on the whole follow the traditional ("campaign-style") cycle of activities in a large-scale information retrieval evaluation experiment set-up. Workshop-style labs are offered as a way of exploring possible benchmarking activities and to provide a means to discuss information retrieval evaluation issues from different perspectives.

A call for lab proposals was distributed in October 2010 and a lab selection committee was formed from leading IR researchers (see Acknowledgements). Lab proposals were requested to include a detailed description of the topics and goals of the lab, the target audience, potential opportunities for future versions of the lab, as well as details about the planning and organisation of the lab (e.g., suggested tasks, the data collections used in the lab, the background of lab organizers and steering committee members, etc.).

Seven lab proposals were accepted for CLEF 2011: 6 benchmarking labs and 1 workshop. By August 2011, 96 research groups had submitted experimental results in a benchmarking activity and 107 participants registered to attend one of the lab sessions at CLEF. The results of the activities of the labs are reported in the working notes, available on-line [3]. It is hoped that they will also give rise to post-conference publications in journals or separate volumes.

The following benchmarking labs ran in CLEF 2011:

CLEF-IP: IR in the IP domain: CLEF-IP used a collection of ca. 2 million patent documents in English, German, and French, together with patent images. Four tasks were organized: the

Prior Art Candidate Search task for finding potential prior art for a document, the Image-based Document Retrieval for finding patent documents or images for a patent document containing images, the Classification task for classifying documents according to the International Patent Classification scheme and the Image-based Classification task for categorising patent images into pre-defined image categories.

Cross-Language Image Retrieval (ImageCLEF): This was the 9th year of running this track in CLEF. Four tasks were offered this year: image retrieval from Wikipedia, medical image retrieval with a data collection from the scientific literature, visual plant species classification of leaf images and a photo annotation task based on user-generated annotations/tags from Flickr.

LogCLEF: In its 3rd CLEF year, the aim of LogCLEF was to analyse transaction logs and classify queries in order to help with understanding user's search behavior in multilingual contexts. Three tasks were organized: a query language identification task, a query classification task against a set of predetermined categories and the analysis of logs with respect to the "success" of a query. Three different log files were made available: logs from The European Library; logs from the Sogou Chinese search engine; and logs from Deutsche Bildungsserver.

MusiCLEF: This was a newcomer to CLEF and the first year of running the lab. The goal of this lab was to aid the development of novel methodologies for both content-based and contextual-based (e.g. tags, comments, reviews, etc.) access and retrieval of music. Two tasks were offered in 2011: content- and context-based music retrieval for music categorisation and music identification for the clustering of the same (or similar) works. The test collection consisted of 10,000 songs in MP3 format.

QA4MRE - Question Answering for Machine Reading Evaluation: This was a new task proposed by organizers of the previous Question Answering (QA) track of CLEF. The aim of this lab was to evaluate machine reading ability through question answering and reading comprehension tests. The task involved answering multiple choice questions based on reading a single document and with some inference from previously-acquired background knowledge. Additionally, a pilot task was offered that involved detection of negation and modality in texts.

Uncovering Plagiarism, Authorship, and Wikipedia Vandalism (PAN): This is the second year of running PAN at CLEF and three tasks were offered to participants: plagiarism detection, author identification and Wikipedia vandalism detection. The last task sought to support users of Wikipedia in their search for "vandalized" pages.

The following workshop-style lab was also held at CLEF 2011:

CHiC 2011 - Cultural Heritage in CLEF: This workshop aimed at surveying use cases for information access to cultural heritage materials and review evaluation initiatives and approaches in order to identify future opportunities for novel evaluation experiments and measures. Workshop participants were asked to introduce their ideas for possible evaluation scenarios.

4 CLEF Charter

During CLEF 2011 we prepared the CLEF charter, which explicitly defines the objectives, scope, and organization of the CLEF Initiative. We summarize these here.

The CLEF Initiative is a self-organized body whose main mission is to promote research, innovation, and development of information access systems with an emphasis on multilingual and multimodal information with various levels of structure. CLEF promotes research and development by providing an infrastructure for: multilingual and multimodal system testing, tuning and evaluation; investigation of the use of unstructured, semi-structured, highly-structured, and semantically enriched data in information access; creation of reusable test collections for benchmarking; item exploration of new evaluation methodologies and innovative ways of using experimental data; discussion of results, comparison of approaches, exchange of ideas, and transfer of knowledge.

The CLEF Initiative is organized by an Executive Board and a Steering Committee. The main purpose of the Steering Committee is to ensure the continuity and the quality of the CLEF Initiatives. The Executive Board is responsible for successful running of the CLEF Initiative. It is comprised of the Steering Committee Chair and two Deputy Chairs, responsible for assisting the Steering Committee Chair in overseeing the organization of the Evaluation Labs and the Conference, ensuring continuity of the event organization over the years.

Each yearly CLEF cycle is structured in two main parts. Firstly, a series of Evaluation Labs, i.e., laboratories to conduct evaluation of information access systems and workshops to discuss and pilot innovative evaluation activities. Secondly, a peer-reviewed Conference on a broad range of issues, including: investigation continuing the activities of the Evaluation Labs; experiments using multilingual and multimodal data; and research in evaluation methodologies and challenges. Each yearly cycle is organized by the following Chairs and Committees: the General Chair(s) and the Organizing Committee; the Evaluation Labs Chair(s) and the Evaluation Labs Selection Committee; and the Conference Program Chair(s) and the Program Committee.

5 CLEF 2012 and Beyond

CLEF 2012 will follow a similar format as CLEF 2010 and 2011, namely conference plus labs and workshops, and will be organized by the Sapienza University of Rome, Italy, in September 2012. The general co-chairs will be Tiziana Catarci, from Sapienza University of Rome (Italy), and Djoerd Hiemstra, from University of Twente (The Netherlands), the program co-chairs will be Anselmo Peñas, from National Distance Learning University (Spain), and Giuseppe Santucci, from Sapienza University of Rome (Italy), and the lab co-chairs will be Jussi Karlgren, from Swedish Institute of Computer Science (Sweden) and Christa Womser-Hacker, from University of Hildesheim (Germany). A call for lab proposals will be issued in October 2011 and an initial call for papers will be issued by the end of the year. We are planning the introduction of a best paper award for the conference portion of CLEF 2012.

Looking towards CLEF 2013, the plan is to open a bid around December 2011. The purpose of the bid will be to allow interested institutions and organizations to apply and compete for organization and hosting of the event. Details of the selection criteria will be made available when the bid is opened.

Acknowledgments

References

[1] M. Agosti, N. Ferro, C. Peters, M. de Rijke, and A. 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] P. Forner, J. Gonzalo, J. Kekäläinen, M. Lalmas, and M. de Rijke, editors. *Multilingual and Multimodal Information Access Evaluation, International Conference of the Cross-Language Evaluation Forum, CLEF 2011, Amsterdam, the Netherlands, September 19-22, 2011. Proceedings*, volume 6941 of *Lecture Notes in Computer Science*. Springer, 2011.

[3] V. Petras, P. Forner, and P. D. Clough, editors. *CLEF 2011 Labs and Workshop, Notebook Papers, 19-22 September 2011, Amsterdam, The Netherlands*, 2011. `http://clef2011.org/index.php?page=pages/proceedings.php`.