

Introduction to the CLEF 2012 Labs

Jussi Karlgren¹ and Christa Womser-Hacker²

¹Gavagai, Stockholm, Sweden

²University of Hildesheim, Dept. of Information Science & Natural Language Processing, Germany
jussi@sics.se; womser@uni-hildesheim.de

The Third International Conference of the Cross-Language Evaluation Forum, CLEF, has as one of its most central features a broad palette of empirical evaluation activities for information access systems of various types. These are proposed and operated by groups of organisers volunteering their time and effort to define, promote, administer and run an evaluation activity. In 2012, as in 2011 and 2010, two participation types were offered:

1. **CLEF Evaluation Labs** that follow the tradition of “campaign-style” known from former CLEF events. They evaluate practice for specific information access problems (during the twelve month period preceding the conference).
2. **CLEF Lab Workshops** organised as speaking and discussion sessions to explore issues of evaluation methodology, metrics, and processes in information access.

A call for participation was distributed after CLEF 2011 to solicit proposals for CLEF activities. Twelve creative proposals of very varying scope and domain were submitted by the end of November 2011 and eight of these were accepted as activities into the CLEF 2012 program. In some cases, proposals were judged to be similar to each other, in which case only one of them was accepted - we see this as a good indication of the timeliness of the task in question. Seven of the accepted proposals are evaluation labs:

1. CHiC Cultural Heritage in CLEF a benchmarking activity to investigate systematic and large-scale evaluation of cultural heritage digital libraries and information access systems.
2. CLEF-IP a benchmarking activity to investigate IR techniques in the patent domain
3. ImageCLEF a benchmarking activity on the experimental evaluation of image classification and retrieval, focusing on the combination of textual and visual evidence
4. INEX a benchmarking activity on the evaluation of XML retrieval
5. PAN a benchmarking activity on uncovering plagiarism, authorship and social software misuse
6. QA4MRE a benchmarking activity on the evaluation of Machine Reading systems through Question Answering and Reading Comprehension Tests
7. RepLab a benchmarking activity on reputation management technologies

One activity is organised as a lab workshop:

CLEFeHealth 2012 workshop on Cross-Language Evaluation of Methods, Applications, and Resources for eHealth Document Analysis

Besides well-established criteria from previous years' editions of CLEF such as topical relevance, novelty, potential impact on future world affairs, likely number of participants, and the quality of the organising consortium, this year we stressed movement beyond previous years' efforts and connection to real-life usage scenarios. This is in keeping with work performed in the PROMISE project, which stresses the connection and necessary linkage between the two sides of evaluating information systems: quantitative benchmarking information access systems on the hand, and validating hypotheses of usage and application on the other. Benchmarking has traditionally been the main focus of CLEF and other related evaluation campaigns, but we want to move towards evaluation activities which are comparable across tasks, and to enable validation of the evaluation efforts.

Therefore, this year, we required the lab proposals to address the issue of validation through explicitly stated hypotheses of usage. An evaluation lab should be concrete with respect to situation, context, platform and user preferences for which the suggested evaluation benchmark is valid; a lab workshop should discuss how participants with domain and usage experience and expertise can be recruited to the workshop to provide a grounding of evaluation methodology in application to real-world task. We hope this principle will lead to wider contact surfaces between evaluation campaigns in CLEF and industrial stakeholders.

In previous years, lab workshops have resulted in a proposal for an evaluation lab for the following year. This was again the case this year: the CHiC lab workshop from 2011 submitted a proposal for an evaluation laboratory in 2012. This progression from a lab workshop to an evaluation lab is a development track the CLEF Lab Organisation Committee wishes to encourage to introduce the organisers of an evaluation campaign to discuss practicalities and the make-up of an evaluation task. However, we do not expect that lab workshops are limited to planning future campaigns: their scope may well be more abstract, more far-reaching or more specific.

Acknowledgements

We would like to thank the members of CLEF-LOC (the CLEF Lab Organisation Committee) for their thoughtful and elaborate contributions to assessing the proposals during the selection process:

Paul Clough, The University of Sheffield
Hideo Joho, University of Tsukuba
Jaana Kekäläinen, University of Tampere
Vanessa Murdock, Yahoo! Research
Doug Oard, University of Maryland

Last but not least without the important and tireless effort of the enthusiastic and creative proposal authors, the organisers of the selected labs, the colleagues and friends involved in running them, and the participants who contribute their time to making the labs and workshops a success, the CLEF labs would not be possible.

Thank you all very much!