

Introduction to the CLEF 2010 labs

Martin Braschler¹ and Donna Harman²

¹ Zurich University for Applied Sciences, Switzerland
bram@zhaw.ch

² National Institute of Standards and Technology (NIST), USA
donna.harman@nist.gov

CLEF 2010 had two different types of labs: benchmarking evaluations for specific information access problem(s), encompassing activities similar to previous CLEF campaign tracks, and workshops for exploration of specific issues in information access. 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 CLEF2010 lab organizing committee, with final decisions of acceptance and length of labs determined by the following criteria.

- Benchmarking labs: soundness of methodology, feasibility of task, importance of problem, number of potential participants, clear movement along a growth path (for labs in previous CLEFs), and appropriate number of subtasks.
- Workshop-style labs: appropriateness of lab to the overall information access agenda pursued by CLEF, number of potential participants, likelihood that the outcome of the workshop would constitute a significant contribution to the field, and adequate focus such that useful conclusions are likely to result.
- Additional other factors for both types of labs included innovation, minimal overlap with other evaluation initiatives and events, vision for a potential continuation, and possible interdisciplinary character.

These five benchmarking evaluations ran as labs in CLEF 2010.

- CLEF-IP '10, sponsored by the Information Retrieval Facility (IRF) in Vienna was a follow-on to a CLEF 2009 track. 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 IPC codes.
- ImageCLEF 2010 was the eighth running of this track in CLEF. There were four tasks in 2010: Medical Retrieval from 77,000 images from articles published in Radiology and Radiographics, Photo Annotation of a MIR Flickr 25,000 database of consumer photos with multiple annotations, Robot Vision Challenge, and 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.

- 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 CLEF2009, 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.

These two exploration workshops ran as labs in CLEF 2010.

- CriES 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.
- LogCLEF investigated the analysis and classification of queries in order to understand search behavior in multilingual contexts and ultimately to improve search systems. The 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.

We would like to thank the reviewers of the lab proposals for their help:

Susan Dumais, Microsoft Research, USA

Gregory Grefenstette, Exalead, France

Jussi Karlgren, Swedish Institute of Computer Science, Sweden

Stephen Robertson, Microsoft Research, UK

We would also like to thank all of the people involved in making these labs happen; they are the heart of CLEF.