

Introduction to the CLEF 2011 Labs

Vivien Petras¹ and Paul Clough²

¹Berlin School of Library and Information Science, Humboldt-Universität zu Berlin
vivien.petras@ibi.hu-berlin.de

²Information School, University of Sheffield, United Kingdom
p.d.clough@sheffield.ac.uk

1 Introduction

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:

- Definition of the task by the lab organisers;
- Preparation and release of data collections and test data by lab organisers;
- Release of queries, topics or other task descriptions by lab organisers;
- Submission of experimental results by participants;
- Relevance assessments and/or results analysis by lab organisers;
- Release of relevance assessments and/or results analysis by lab organisers;
- Submission of working notes on experiments and results analysis by participants;
- Presentation and discussion at the CLEF conference.

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. Seven labs were offered in 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 labs are discussed in the following sections.

2 Lab Selection Process

A call for lab proposals was distributed in October 2010 and a lab selection committee 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 organisers and steering committee members, etc.).

The proposals were submitted by 8th November 2010 and the selection committee had 2 weeks to decide on the labs that would feature at CLEF 2011. Selection committee members reviewed all proposals and submitted both an individual ranking for each proposal as well as their comments on the lab. All reviews were aggregated by the Lab Chairs and a general ranking of labs was produced. Reviewing criteria for the benchmarking labs included (from the call for proposals¹):

- Soundness of methodology and feasibility of the task;
- Potential use and business cases for the lab, e.g. a description of the underlying problems, industrial stakeholders, the potential for market;
- Estimated number of potential participants / critical mass;
- Clear movement along a growth path and development of the field.

Reviewing criteria for the workshop-style labs included (from call for proposals):

- Appropriateness to the overall information access agenda pursued by CLEF;
- Estimated number of potential participants / critical mass;
- Whether outcomes of the workshop would make significant contributions to the field.

Additional factors, such as innovation, minimal overlap with other evaluation initiatives and events, the vision for potential continuation, and the interdisciplinary nature of the lab were also considered. Of the 9 submitted labs, 6 were accepted as benchmarking labs and 1 accepted as a workshop, resulting in an acceptance rate of 7/9 (=77%).

3 Descriptions of the Labs

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 organised: 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.

¹ http://clef2011.org/resources/call_for_labs/CLEF2011-Call_for_Lab_proposals.pdf

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 is to analyse transaction logs and classify queries in order to help with understanding user's search behavior in multilingual contexts. Three tasks were organised: 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 is 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 organisers of the previous Question Answering (QA) track of CLEF. The aim of this lab is to evaluate machine reading ability through question answering and reading comprehension tests. The task involves answering multiple choice questions based on reading a single document and with some inference from previously-acquired background knowledge.

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 seeks to support users of Wikipedia in their search for 'vandalised' pages.

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

CHiC 2011 – Cultural Heritage in CLEF: From Use Cases to Evaluation in Practice for Multilingual Information Access to Cultural Heritage: 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.

Acknowledgements

We would like to thank the lab selection committee for their assistance with reviewing proposals and in the selection process:

- Martin Braschler, Zurich University of Applied Sciences, Switzerland;
- Gregory Grefenstette, Exalead, France;
- Jussi Karlgren, Swedish Institute of Computer Science, Sweden;
- Doug Oard, College of Information Studies, University of Maryland;
- Mark Sanderson, Royal Melbourne Institute of Technology University, Australia;
- Ryen White, Microsoft, USA;

Last, but not least, we would like to thank everyone involved in the organisation, running and presentation of the labs, especially the lab organisers. Without their tireless and mostly voluntary efforts, CLEF Labs would not be possible.