



Science Meeting – Scientific Report

Proposal Title: CLEF 2013 organizers meeting

Application Reference N°: 5100

1) Summary (up to one page)

The CLEF conference is a yearly edition of the popular CLEF lab and workshop series which has run since 2000 contributing to the systematic evaluation of information access systems, primarily through experimentation on shared tasks. After ten years of successful evaluation campaigns, in 2010 CLEF was launched in a new format, as a conference with research presentations, panels, poster and demo sessions and laboratory evaluation workshops interleaved during 3,5 days of intense and stimulating research activities.

The CLEF initiative (www.clef-initiative.eu) is a self-organized body whose main mission is to promote research, innovation, and development of information access systems with an emphasis on multilingual information in different modalities with various levels of structure.

In this CLEF organizers meeting all lab organizers came together to discuss the organization of the labs at CLEF2013, that were to be organized from 23 to 26 September 2013 in Valencia, Spain. This organizers meeting was organized at the European Conference on Information Retrieval (ECIR 2013) in Moscow, from 24 to 27 March 2013.

2) Description of the scientific content of and discussions at the event (up to four pages)

The CLEF organizers meeting was organized to discuss and organize the labs that were to be offered at CLEF 2013. In this meeting the format of the lab, the problem space, the practicalities of the shared task were discussed, as well as the expectation that lab sessions contain ample time for general discussion and engagement by all participants, not just those presenting campaign results and papers. Organizers should plan time for panels, demos, poster sessions etc. where applicable.

This concerns the following labs:

Nine labs follow a "campaign-style" evaluation practice for specific information access problems in the tradition of past CLEF campaign tracks:

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. [CLEFeHealth - CLEF eHealth Evaluation Lab](#) a benchmarking activity aiming at developing processing methods and resources to enrich difficult-to-understand health text as well as their evaluation setting
3. [CLEF-IP - Retrieval in the Intellectual Property Domain](#) a benchmarking activity to investigate IR techniques in the patent domain
4. [ImageCLEF - Cross Language Image Annotation and Retrieval](#) a benchmarking activity on the experimental evaluation of image classification and retrieval, focusing on the combination of textual and visual evidence
5. [INEX - Initiative for the Evaluation of XML retrieval](#) builds evaluation benchmarks for search with rich structure - such as document structure, semantic metadata, entities, or genre/topical structure - as of increasing importance on the web and in professional search.
6. [PAN - Uncovering Plagiarism, Authorship, and Social Software Misuse](#) a benchmarking activity on uncovering plagiarism, authorship and social software misuse
7. [QA4MRE - Question Answering for Machine Reading Evaluation](#) a benchmarking activity on the evaluation of machine reading systems through question answering and reading comprehension tests
8. [QALD-3 - Question Answering over Linked Data](#) a benchmarking activity on question answering over linked data
9. [RepLab 2013](#) second CLEF lab on Online Reputation Management

One lab would run as a workshop/discussion session to explore issues of evaluation methodology, metrics, and processes in information access and closely related fields:

1. [CLEF-ER - Entity Recognition @ CLEF](#) workshop on multilingual annotation of named entities and terminology resources acquisition

3) Assessment of the results and impact of the event on the future directions of the field (up to two pages)

The CLEF labs have a great importance for the research community . The labs at CLEF bring together 185 research teams from all over the world. The labs have an impact on the researchers of all these groups.

CLEF promotes research and development by providing an infrastructure for:

- Independent evaluation of information access systems;

- Investigation of the use of unstructured, semi-structured, highly-structured, and semantically enriched data in information access;
- Creation of reusable test collections for benchmarking;
- 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 Lab Sessions

(The following report on the CLEF 2013 Labs was published in *SIGIR Forum* 47(2), December 2013, pages, 15–20.)

CLEF 2013 hosted ten labs, nine of which followed the campaign-style evaluation practice for specific information access problems, and one lab was organized as a workshop. Campaign-style evaluations are organized during the year preceding the conference and follow the tradition of past CLEF campaign tracks. Lab workshops are organized as speaking and discussion sessions to explore challenges of evaluation methodology, metrics, and processes in information access; they can be a first step towards an evaluation lab.

The call for lab proposals was published in October 2012. Lab proposals were requested to include a detailed description of the topics and goals of the lab, the targeted audience, potential opportunities for future versions of the lab, as well as details about the tasks and data collections. By August 2013, 185 unique research groups had submitted experimental results in a benchmarking activity, and 163 participants registered to attend one of the lab sessions at CLEF. In the following a brief overview of the labs is given.

CHiC – Cultural Heritage in CLEF The lab aims at moving towards a systematic and large scale evaluation of cultural heritage digital libraries and information access systems. After a workshop in 2011 and a pilot lab in 2012, where a standard ad-hoc information retrieval scenario was tested together with two use-case-based scenarios, the 2013 lab strived to diversify more of the tasks and to become more realistic in its tasks organization.

Number of participating teams: 7

CLEFeHealth - CLEF eHealth The goal is to develop methods and resources that make discharge documents easier to understand from a nurses and patient perspective and address their differing queries and information needs when searching further details on matters mentioned in the discharge summaries. It contained three related sub-tasks: (1) identification of disorders from clinical reports and mapping of the SNOMED CT disorders to UMLS codes, (2) mapping abbreviations and acronyms in clinical reports to UMLS codes, and (3) information retrieval to address questions that patients may have when reading clinical reports based on a collection of certified health web pages.

Number of participating teams: 33

CLEF-IP Retrieval in the Intellectual Property Domain The lab provides a large collection of XML documents representing patents and patent images. Based on this collection three tasks were organized. (1) Passage retrieval starting from claims: Given a claim, the participants were asked

to retrieve relevant documents in the collection and mark out the relevant passages in these documents. (2) Text to image and image to text: Given a patent application document as an XML file and the set of images occurring in the application, extract the links between the image labels and the text pointing to the object of the image label. (3) Structure Recognition Task: Extract the information in these images and return it in a predefined textual format.

Number of participating teams: 3

ImageCLEF – Cross Language Image Annotation and Retrieval The lab deals with the cross-language annotation and retrieval of images. Motivated by the need to support multilingual users from a global community accessing the ever growing body of visual information, the main goal is to support the advancement of the field of visual media analysis, indexing, classification, and retrieval, by developing the necessary infrastructure for the evaluation of visual information retrieval systems operating in both monolingual, cross-language and language-independent contexts.

Number of participating teams: 34

INEX – Initiative for the Evaluation of XML Retrieval Main goal is to promote the evaluation of focused retrieval by providing large test collections of structured documents, uniform evaluation measures, and a forum for organizations to compare their results. A search engine is referred to as being focused if it—aside from identifying documents that are relevant to a user’s information need—also locates the relevant information within the document. Focused Retrieval takes different forms: Passage Retrieval from a long document, Element Retrieval from an XML document, Page Retrieval from books, as well as Question Answering.

Number of participating teams: 19

PAN – Uncovering Plagiarism, Authorship, and Social Software Misuse The lab provides three tasks from the field of digital text forensics. (1) Plagiarism Detection: Given a document, analyze whether it is an original. This task is divided into source retrieval (searching for likely sources) and text alignment (matching passages of reused text). (2) Author Identification: Given a document, analyze who wrote it. This task focuses on authorship verification as well as methods to answer the question whether two given documents have the same author or not. (3) Author Profiling: Given a document, analyze particular author characteristics.

Number of participating teams: 46

QA4MRE – Question Answering for Machine Reading Evaluation Main goal of this lab is to develop a methodology for evaluating machine reading systems through question answering and reading comprehension tests. Systems should be able to extract knowledge from large volumes of text and use this knowledge to answer questions. Three tasks were provided. (1) The machine reading task addresses the problem of building a bridge between knowledge encoded as natural text and the formal reasoning systems that need such knowledge. (2) Machine reading of biomedical texts about the Alzheimer’s Disease. (3) Entrance Exams, which aims at evaluating systems under the same conditions humans are evaluated to enter the University.

Number of participating teams: 14

QALD-3 – Question Answering over Linked Data A lab on question answering over linked data,

this time with a strong emphasis on multilinguality. It offered two challenges: (1) Multilingual question answering and (2) Ontology lexicalization. Altogether, the key challenge ACM SIGIR Forum 18 Vol. 47 No. 2 December 2013 lies in translating the users' information needs into a form such that they can be evaluated using standard semantic web query processing and inferencing techniques.

Number of participating teams: 6

RebLab – Online Reputation Management A competitive evaluation exercise for online reputation management systems. The lab focused on the task of monitoring the reputation of entities (companies, organizations, celebrities) on Twitter. The monitoring task for analysts consisted of searching the stream of tweets for potential mentions to the entity, filtering those that do refer to the entity, detecting topics (i.e., clustering tweets by subject) and ranking them based on the degree to which they signal reputation alerts (i.e., issues that may have a substantial impact on the reputation of the entity).

Number of participating teams: 16

CLEF-ER – Entity Recognition A workshop on the multilingual annotation of named entities and terminology resource acquisition. It addresses entity recognition in biomedical text, in different languages and at a large scale. The workshop was organized by the Mantra (Multilingual Annotation of Named Entities and Terminology Resources Acquisition) EUfunded project. Mantra will provide multilingual terminologies and semantically annotated multilingual documents in order to improve the accessibility of scientific information from multilingual documents.

The CLEF 2013 labs had a significant impact on the research. Important progress was made in the evaluation of a broad range of information access tasks. In 2014 lessons learned and further collaborations between the labs will be implemented.

4) Annexes 4a) and 4b): Programme of the meeting and full list of speakers and participants

Annex 4a: Programme of the meeting

Program of the meeting:

CLEF Labs @ ECIR 2013
Digital October Centre, 3rd floor
Moscow
March 26, 2013

09:00 Opening
09:05-11.00 The Labs plus quick feedback and replies
11.00-11.30 Break (aligned with ECIR 2013)
11.30-13.00 Sharing resources, sharing expertise, sharing lessons
13.00-14.30 Lunch on our own (aligned with ECIR 2013)
14.30-16.00 Sharing the future of information access evaluation
16:05 Wrap-up

The first "sharing session" (11.30-13.00) was about where the participants thought their lab could benefit from other labs or vice versa (resources, expertise and lessons learned).

The second "sharing session" (14.30-16.00) was about where the participants thought information access evaluation should be heading next and what the participants thought they, their labs, other labs and CLEF could be doing to get there, starting with CLEF 2014.

Annex 4b: Full list of speakers and participants

List of speakers and participants:

- dr. Nicola Ferro, University of Padova
- dr. Allan Hanbury, Vienna University of Technology
- dr. Jaap Kamps, University of Amsterdam
- prof. dr. Vivien Petras, Humboldt University Berlin
- dr. Florina Piroi, Vienna University of Technology
- dr. Dietrich Rebholz-Schuhmann, University of Zurich
- prof. dr. Maarten de Rijke, University of Amsterdam
- dr. Paolo Rosso, Universitat Politecnica de Valencia