CLEF 2013

Information Access Evaluation meets Multilinguality, Multimodality, and Visualization

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

Pamela Forner
CELCT, Italy
forner@celct.it

Henning Müller
Affiliation
henning.mueller@hevs.ch

Roberto Navigli
Sapienza University of Rome, Italy
navigli@di.uniroma1.it

Roberto Paredes
Affiliation
rparedes@dsic.upv.es

Paolo Rosso
Universitat Politècnica de València
prosso@dsic.upv.es

Benno Stein
Bauhaus-Universität Weimar, Germany
benno.stein@uni-weimar.de

Dan Tufis
RACAI, Romania
tufis@racai.ro

1 Introduction

The CLEF 2013 Conference on Information Access Evaluation was held at the Universitat Politècnica de València, Spain, September 23–26, 2013.

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 multimodal and multilingual information access. Through the years, CLEF has promoted the study and implementation of evaluation methodologies for diverse tasks, resulting in the creation of a broad, strong and multidisciplinary research community. CLEF's core goals are the benchmarking activities carried out in various tracks. These are complemented with a peer-reviewed conference component that aims at advancing research in the evaluation of complex information access systems in different languages and modalities.

The CLEF 2013 conference had more than 200 participants (190 participants in 2012, Rome) from different academic and industrial institutions. Though the majority of participants comes from Europe (149), there is also a considerable interest in CLEF around the world, above all from

the Americas with 28 participants, Asia with 22 participants, Australia with 3 participants, and Africa with 2 participants.

2 The CLEF Conference

For the conference part 11 long papers, 5 short papers, and 6 posters were accepted (acceptance rate 51%); 15 authors were given a slot for oral presentation.[1] The papers were presented within three sessions dedicated to (1) evaluation and visualization, (2) evaluation and multilinguality, and (3) evaluation and applications. It is the idea of CLEF to encourage the submission of papers that address methodological and evaluation issues in order provide a scientific frame program that is meant to be inspiring for participants across all labs.

The two conference keynotes focused on two central CLEF topics, namely evaluation and multilinguality.

Evangelos Kanoulas (Google Zurich, Switzerland) talked about different aspects of batch experimentation. He distinguishes three evaluation principles for measuring information retrieval effectiveness: batch experiments based on static test collections, lab studies measuring actual users interacting with a system, and online experiments tracking user's interactions with a live system. His observation is that experiments often introduce too many simplifying assumptions and hence cannot predict the usefulness of a system to its users. He suggests to create test collections that better model the variability encountered in real-life search scenarios, including the variation of queries, corpora, and search result interactions of the users. His talk addressed the challenges when dealing with an increased amount of variation.

Rada Mihalcea (University of Michigan) talked about the increasing number of languages in the Web introduced by the online resources such as Wikipedia, Twitter, or Facebook, and the growing need for effective solutions for multilingual natural language processing. Her talk was centered around the hypothesis that a multilingual representation can enrich the feature space for natural language processing and hence lead to improvements compared to traditional solutions that rely on a monolingual representation. To support her hypothesis she reported on experiments for three different tasks: word sense disambiguation, subjectivity analysis, and text semantic similarity.

The conference hosted also three so-called "community sessions", which informed about the evaluation initiatives MediaEval, NTCIR, and FIRE. The MediaEval benchmarking initiative is dedicated to evaluating new algorithms for multimedia access and retrieval. It emphasizes the "multi" in multimedia and focuses on human and social aspects of multimedia tasks. NTCIR is an evaluation workshop series aiming to enhance the research in information access technologies by providing large-scale infrastructure of the experiments. It has run in 18-month-cycle and mainly used East Asian Languages and English. The aim of FIRE, short for Forum for Information Retrieval Evaluation, is to encourage research in Indian language by creating a similar platform for Indian languages that provides the data and a common forum for comparing models and techniques.

[1] CLEF 2012 had 17 papers (9 long, 5 short, and 3 posters) and an acceptance rate of 52%.

3 The CLEF Lab Sessions

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.[2] 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 issues 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 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 nurses and patients' 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.

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

[2]CLEF 2012 hosted seven campaign-style labs and one workshop.

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

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 resources acquisition. It addresses entity recognition in biomedical text, in different languages and at a large scale. The workshop was organized by the EC Mantra (Multilingual Annotation of Named Entities and Terminology Resources Acquisition) project. Mantra will provide multilingual terminologies and semantically annotated multilingual documents in order to improve the accessibility of scientific information from multilingual documents.

4 CLEF 2013 and Beyond

In order to have a single publication that reports the whole breadth of the CLEF activities, it has been decided that the annual Springer LNCS proceedings will report not only the conference papers but will also contain revised overviews of the labs.

More information on the CLEF initiative and the CLEF 2013 conference including detailed online working notes is provided on the Web:

- CLEF initiative. `http://www.clef-initiative.eu`
- Conference 2013 home. `http://www.clef2013.org`

CLEF 2014 will be hosted by the University of Sheffield, United Kingdom, 15-19 September 2014. In particular, eight lab proposals were accepted for CLEF 2014. The call for papers for the CLEF Conference will be released in November 2013, and the expected deadline for the submission of papers will be April 2014.

Acknowledgments

The success of CLEF 2013 would not have been possible without the contributions of the members of the Program Committee, the Organizing Committee, the students, and other volunteers that supported the initiative in its various stages. Thank you all!

We would like to express our gratitude to the following sponsoring organizations for their support: PROMISE Network of Excellence, ELIAS Research Network Programme, European Science Foundation (ESF), Yandex, Universitat Politècnica de València, Departamento Sistemas Informáticos y Computación (DSIC), Escola Técnica Superior d'Enginieria Informática (ETSINF), COREX, and the Web Information Quality Evaluation Initiative (WIQ-EI).