CLEF 2014

Information Access Evaluation meets Multilinguality, Multimodality, and Interaction

Linda Cappellato
University of Padua, Italy
cappellato@dei.unipd.it

Paul Clough
University of Sheffield, UK
p.d.clough@sheffield.ac.uk

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

Mark Hall
Edge Hill University, UK
mark.hall@edgehill.ac.uk

Martin Halvey
University of Strathclyde, UK
martin.halvey@strath.ac.uk

Allan Hanbury
Vienna University of Technology, Austria
hanbury@ifs.tuwien.ac.at

Evangelos Kanoulas
University of Amsterdam, The Netherlands
ekanoulas@gmail.com

Wessel Kraaij
Radboud University, The Netherlands
w.kraaij@cs.ru.nl

Mihai Lupu
Vienna University of Technology, Austria
lupu@ifs.tuwien.ac.at

Mark Sanderson
RMIT University, Australia
mark.sanderson@rmit.edu.au

Elaine Toms
University of Sheffield, UK
e.toms@sheffield.ac.uk

Robert Villa
University of Sheffield, UK
r.villa@sheffield.ac.uk

1 Introduction

The 2014 Conference and Labs of the Evaluation Forum[1] (CLEF) event was held at the University of Sheffield, UK, on September 15–18, 2014. The conference was entitled "Information Access Evaluation meets Multilinguality, Multimodality and Interaction" and addressed issues around multilingual and multimodal information access, information interaction, as well as the evaluation of search systems.

CLEF celebrated its 15th anniversary this year having been conceived in 2000 as the Cross-Language Evaluation Forum aimed at stimulting research and innovation in multimodal and multilingual information access and retrieval. Over the years, it has actively nurtured and engaged a vibrant, multidisciplinary research community in the study, design, and implementation of evaluation methods for multiple tasks using diverse data sets and in many languages.

In the first 10 years, CLEF conducted a series of experimental labs that were reported annually at workshops held in conjunction with the European Conference on Digital Libraries (ECDL). In 2010, now a mature and well-respected evaluation forum, it expanded to include a complementary peer-reviewed conference for reporting the evaluation of information access and retrieval systems regardless of data type, format, language, etc. Since then CLEF has continued that format with keynotes, contributed papers, lab sessions, and poster sessions, including reports from other benchmarking initiatives from around the world.

The CLEF 2014 conference had more than 150 participants from different academic institutions and industrial organisations. Though the majority (114) of participants come from Europe, there was also considerable interest in CLEF worldwide, with 20 participants from the Americas, 16 from Asia, 1 from Australia, and 2 from the Middle East.

2 The CLEF Conference

For the conference part 12 long papers and 4 short papers were accepted (an acceptance rate of 53%). The papers were presented within three sessions dedicated to (1) evaluation, (2) domain specific approaches, and (3) alternative search tasks. CLEF aims to encourage the submission of papers that address methodological and evaluation questions in order to provide a scientific framework that is meant to be inspiring for participants across all labs.

Three eminent scholars in the field headlined each day of the conference and all took a human-centered focus to their talks. *Ann Blandford* (Professor of Human-Computer Interaction at University College London) put the user at the center of an information seeking and retrieval process which she referred to as the *information journey*. In her talk she examined ways in which the wider user experience process could be evaluated to support this information journey. *Susan Dumais* (Distinguished Scientist at Microsoft and Deputy Managing Director of the Microsoft Research Lab in Redmond) focused on query understanding. Focusing on a short search query in isolation is a very difficult task; query understanding is much easier if the "context" in which the query arises, e.g., previous queries the searcher has issued, the current location and time, etc., is taken into account. Susan described a framework for capturing query context and provided several examples of how context could be represented and used to improve search quality. *Fabio*

---

[1] http://clef2014.clef-initiative.eu/

Ciravegna (Professor of Language and Knowledge Technologies at the Department of Computer Science at the University of Sheffield) discussed his work with emergency services and organisers of very large events involving hundreds of thousands of participants to help identify planned and unplanned situations through social media. To support this required not only development of real-time large-scale text and data analysis tools for social media mining, but also applying visual analytics and an understanding of the human context. Fabio ended by considering ethical issues that must be considered when dealing with social media.

The conference hosted also a panel session on *other evaluation initiatives*, which informed delegates of current and upcoming activities within evaluation initiatives worldwide. The panel session included *Ian Soboroff* from the National Institute of Standards and Technology (USA) who presented TREC[2] (Text REtrieval Conference), the first large-scale evaluation activity orgainsed in the field of IR, which began in 1992. *Hideo Joho* from the University of Tsukuba (Japan) presented NTCIR[3] (NII Testsbeds and Community for Information access Research), which promotes research in information access technologies with a special focus on East Asian languages and English. *Gareth Jones* from Dublin City University (Ireland) presented MediaEval[4], which is dedicated to evaluating new algorithms for multimedia access and retrieva. Finally, *Prasenjit Majumder* from DAIICT (Gandhinagar, India) presented the FIRE[5] (Forum for Information Retrieval Evaluation) initative, whose aim is encouraging research in Indian languages by creating a platform similar to CLEF for Indian languages to provide data and provide a common forum for comparing models and techniques.

The conference also hosted a rich social programme. On Monday evening a welcome reception was held at the conference venue (The Edge, Endcliffe Village) in conjunction with the conference poster session. This allowed people to socialise and to discuss research activities in a warm and relaxed way. On Tuesday the social dinner was held at Cutlers' Hall, one of Sheffield's most prestigious venues set in the heart of the city centre. During the dinner an award was given for best paper to Roland Roller and Mark Stevenson for their paper entitled *Self-Supervised Relation Extraction using UMLS*. Finally, on Wednesday evening, a party was held at the conference venue to celebrate the 15th birthday of CLEF. This included tasting the culinary delights of South Yorkshire in the form of "Pie and Peas".

3 The CLEF Lab Sessions

Out of 10 received lab proposals, 8 were selected and run during CLEF 2014. Proposals were judged on various criteria, including their topical relevance, novelty, potential research impact, likely number of participants, and the quality of the organizing consortium. This year we further stressed the connection to real-life usage scenarios and avoided overlaps amongst labs in order to promote synergies and integration. This was possible also thanks to a new activity introduced in 2013, which is a lab organizers and proposers meeting co-located with the European Conference on Information Retrieval (ECIR), held in Moscow on March 26th, and carried on in 2014 with a

---

[2] http://trec.nist.gov/
[3] http://research.nii.ac.jp/ntcir/
[4] http://www.multimediaeval.org/
[5] http://www.isical.ac.in/~clia/

meeting on April 14th co-located with ECIR 2014 in Amsterdam. Both meetings were supported by the ELIAS network[6].

Each Lab, building on previous experience, demonstrated maturity coming with new tasks, new and larger data sets, new ways of evaluation or more languages. Details about the activities and the outcomes of each lab are reported in the lab proceedings and here we provide just a brief overview of them.

CLEF eHealth – ShARe/CLEF eHealth Evaluation Lab The usage scenario of the CLEF eHealth lab is to ease patients and next-of-kins' ease in understanding eHealth information. The lab contains three tasks: Visual-Interactive Search and Exploration of eHealth Data, Information extraction from clinical text, User-centered health in-formation retrieval.

ImageCLEF ImageCLEF aims at providing benchmarks for the challenging task of image annotation for a wide range of source images and annotation objective. The tasks address different aspects of the annotation problem and are aimed at supporting and promoting the cutting-edge research addressing the key challenges in the field, such as multi-modal image annotation, domain adaptation and ontology driven image annotation. The Lab tasks are: Robot Vision, Scalable concept Image Annotation, Liver CT Annotation and Domain Adaptation.

INEX - Initiative for the Evaluation of XML retrieval INEX builds evaluation benchmarks for search in the context of rich structure such as document structure, semantic metadata, entities, or genre/topical structure. INEX 2014 runs the following tasks: Social Book Search Task, Interactive Social Book Search Task, Tweet Contextualization Task.

LifeCLEF LifeCLEF aims at evaluating multimedia analysis and retrieval techniques on biodiversity data for species identification. The three Tasks regard respectively birds, plants and fishes: BirdCLEF (a bird songs identification task based on Xeno-Canto audio recordings), PlantCLEF (an image-based plant identification task based on the data of Tela Botanica social network) and FishCLEF (a fish video surveillance task based on the data of the Fish4Knowledge network).

NEWSREEL - News Recommendation Evaluation Lab Newsreel Lab is the first news recommendation evaluation Lab and it is divided in two different tasks. One task is about predicting the items a user will click in the next 10 Minutes based on the offline dataset. The other task is about predicting the articles users will click; in this task, the prediction algorithms are evaluated in an online scenario based on live user-interactions.

PAN Lab on Uncovering Plagiarism, Authorship, and Social Software Misuse PAN centers around the topics of plagiarism, authorship, and social software misuse. The goal is to foster research on automatic detection and uncovering. People increasingly share their work online, contribute to open projects and engage in web-based social interactions. The ease and anonymity with which this can be done raises concerns about verifiability and trust: Is a given text an original? Is the author the one who she claims to be? Does

---

[6]http://www.elias-network.eu/

a piece of information come from a trusted source? There are three Lab Tasks: Author Identification, Author Profiling and Plagiarism Detection.

QA Track - CLEF Question Answering Track In the current general scenario for the CLEF QA Track, the starting point is always a Natural Language question. How-ever, answering some questions may need to query Linked Data; whereas some questions may need textual inferences and querying free-text. Answering some queries may need both. The tasks are: QALD - Question Answering over Linked Data, Bio-ASQ - Biomedical semantic indexing and question answering, Entrance Exams.

RepLab - Online Reputation Management The aim of RepLab is to bring together the Information Access research community with representatives from the Online Reputation Management industry, with the ultimate goals of (i) establishing a roadmap on the topic that includes a description of the language technologies required in terms of resources, algorithms, and applications; (ii) specifying suitable evaluation methodologies and metrics to measure scientific progress; and (iii) developing of test collections that enable systematic comparison of algorithms and reliable benchmarking of commercial systems. Lab Tasks are Reputation Dimensions and Author Profiling.

4 CLEF 2015 and Beyond

More information on the CLEF initiative and the CLEF 2014 conference including detailed online working notes is provided on the Web:

- CLEF initiative: `http://www.clef-initiative.eu/`

- CLEF 2014: `http://clef2014.clef-initiative.eu/`

CLEF 2015 will be hosted by the Institut de Recherche en Informatique de Toulouse (IRIT), France, 8-11 September 2015.

The call for papers for the CLEF 2015 Conference will be released around November 2014, and the expected deadline for the submission of papers will be around April 2015.

As far as labs are concerned, CLEF 2015 will run eight evaluation activities:

CLEF eHealth Usage scenario is to ease patients and nurses ease in understanding and accessing eHealth information. The goals of the lab are to develop processing methods and resources in a multilingual setting to enrich difficult-to-understand eHealth texts, and provide valuable documentation. The lab contains two tasks: Information extraction from Clinical Data; and, User-centered health information retrieval.

ImageCLEF In 2015, ImageCLEF will organize four main tasks with a global objective of benchmarking automatic annotation and indexing of images. The tasks tackle different aspects of the annotation problem and are aimed at supporting and promoting cutting-edge research addressing the key challenges in the field: Image Annotation; Medical Classification; Medical Clustering; and, Liver CT Annotation.

LifeCLEF LifeCLEF aims at evaluating multimedia analysis and retrieval techniques on biodiversity data for species identification. It will continue the three tasks offered in the previous year, namely BirdCLEF, PlantCLEF and FishCLEF, but scaling-up the evaluation data to hundreds of thousands of life media records and thousands of living species.

Living Lab for IR (LL4IR) The main goal is to provide a benchmarking platform for researchers to evaluate their ranking systems in a live setting with real users in their natural task environment. This first edition of the lab features three tasks, each corresponding to a single use-case: (i) Local domain search using a university search engine; (ii) Product search on a medium-sized online retailer; (iii) Web search using a commercial search engine.

News Recommendation Evaluation Lab (NEWSREEL) In the second iteration of the lab, we provide two tasks that address the challenge of real-time news recommendation. Participants could: a) develop news recommendation algorithms and b) have them tested by millions of users over the period of a few weeks in a living lab.

Uncovering Plagiarism, Authorship and Social Software Misuse (PAN) This is the 12th evaluation lab on uncovering plagiarism, authorship, and social software misuse. PAN offers three tasks with new evaluation resources consisting of large-scale corpora, performance measures, and web services that allow for meaningful evaluations. The main goal is to provide for sustainable and reproducible evaluations, to get a clear view of the capabilities of state-of-the-art-algorithms. The tasks are: Plagiarism Detection; Author Identification; and, Author Profiling.

Question answering (QA) In the current general scenario for the CLEF QA Track, the starting point is always a Natural Language question. However, answering some questions may need to query Linked Data (especially if aggregations or logical inferences are required); whereas some questions may need textual inferences and querying free-text. Answering some queries may need both. The tasks are: QALD, Question Answering over Linked Data; Entrance Exams, Questions from reading tests; Large-Scale Biomedical Semantic Indexing; and, BioASQ, Biomedical Question answering.

Social Book Search (SBS) Real-world information needs are generally complex, yet almost all research focuses instead on either relatively simple search based on queries or recommendation based on profiles. The goal of the Social Book Search Lab is to investigate techniques to support users in complex book search tasks that involve more than just a query and results list. Tasks: Suggestion track, and Interactive track.

More information about CLEF 2015 are available on its homepage at:

<div align="center">

`http://clef2015.clef-initiative.eu/`

</div>

CLEF 2016 will be hosted by the University of Èvora, Portugal, in early September 2016.

Finally, bids for hosting CLEF 2017 are now open and will close on 10th April 2015. Proposals can be sent to the CLEF Steering Committee Chair at `chair@clef-initiative.eu`.

Acknowledgments

References

[1] Cappellato, L., Ferro, N., Halvey, M., and Kraaij, W., editors (2014). *CLEF 2014 Labs and Workshops, Notebook Papers.* CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, `http://ceur-ws.org/Vol-1180/`.

[2] Kanoulas, E., Lupu, M., Clough, P., Sanderson, M., Hall, M., Hanbury, A., and Toms, E., editors (2014). *Information Access Evaluation – Multilinguality, Multimodality, and Interaction. Proceedings of the Fifth International Conference of the CLEF Initiative (CLEF 2014).* Lecture Notes in Computer Science (LNCS) 8685, Springer, Heidelberg, Germany.

---

[7]`http://www.elias-network.eu/`

[8]`http://www.esf.org/`

[9]`http://www.springer.com/`

[10]`http://www.nowpublishers.com/`

[11]`http://www.khresmoi.eu/project`