

Report on CLEF 2015: Experimental IR Meets Multilinguality, Multimodality, and Interaction

Linda Cappellato¹ Nicola Ferro¹ Gareth J. F. Jones²
Jaap Kamps³ Josiane Mothe⁴ Karen Pinel-Sauvagnat⁴
Eric San Juan⁵ Jacques Savoy⁶

¹ University of Padua, Italy, {cappellato,ferro}@dei.unipd.it

² Dublin City University, Ireland, gareth.jones@computing.dcu.ie

³ University of Amsterdam, The Netherlands, kamps@uva.nl

⁴ IRIT, Université de Toulouse, France, {josiane.mothe,sauvagna}@irit.fr

⁵ Université d'Avignon et des Pays du Vaucluse, France,
eric.sanjuan@univ-avignon.fr

⁶ University of Neuchâtel, Switzerland, jacques.savoy@unine.ch

Abstract

This is a report on the sixth edition of the *Conference and Labs of the Evaluation Forum* (CLEF 2015), held in early September 2015, in Toulouse, France. CLEF was a four day event combining a Conference and an Evaluation Forum. The focus of the Conference is “Experimental IR” as carried out in the CLEF Labs and other evaluation forums, it featured keynotes by Greg Greffentette, Mounia Lalmas, and Doug Oard, and 43 papers have been presented, covering a wide range of topics. There were a total of eight Labs: eHealth, ImageCLEF, LifeCLEF, Living Labs for IR, NEWSREEL, PAN, QA, and Social Book Search, addressing a wide range of tasks, media, languages, and ways to go beyond standard test collections.

1 Introduction

The 2015 edition of the *Conference and Labs of the Evaluation Forum*¹ (CLEF) was hosted by the *Institut de Recherche en Informatique de Toulouse* (IRIT), University of Toulouse, France, from 8-11 September 2015. For its sixth edition, the conference, was entitled “Experimental IR meets Multilinguality, Multimodality, and Interaction” and addressed issues around multilingual and multimodal information access, information interaction, as well as the evaluation of search systems.

¹<http://clef2015.clef-initiative.eu/>

CLEF was established over 15 years ago with a specific focus on stimulating research and innovation in multimodal and multilingual information access and retrieval. Over the years CLEF has fostered the creation of language resources in many European and non-European languages, promoted the growth of a vibrant and multidisciplinary research community, provided sizeable improvements in the performance of monolingual, bilingual, and multilingual information access systems [6], and achieved a substantial scholarly impact [11, 12].

In its first 10 years, CLEF hosted a series of experimental labs that reported their results at an annual workshop held in conjunction with the European Conference on Digital Libraries (ECDL). In 2010, now a mature and well-respected evaluation forum, CLEF expanded to include a complementary peer-reviewed conference for discussion of advancing evaluation methodologies and reporting the evaluation of information access and retrieval systems regardless of data type, format, language, etc. Moreover, the scope of the evaluation labs was broadened, to comprise not only multilinguality but also multimodality in information access. Multimodality here is intended not only as the ability to deal with information coming in multiple media but also in different modalities, e.g. the Web, social media, news streams, specific domains and so on. Since 2010 the CLEF conference has established a format with keynotes, contributed papers, lab sessions, and poster sessions, including reports from other benchmarking initiatives from around the world. Since 2013, CLEF has been supported by an association, a lightweight not-for-profit legal entity, which thanks to the financial support of the CLEF community, takes care of the small central coordination needed to operate CLEF on an ongoing basis and makes it a self-sustaining activity [5].

CLEF 2015 was attended by 181 participants, out of which 72 (40%) were students, denoting a young and vibrant community, from different academic institutions and industrial organisations. Although the majority (140) of the participants came from 24 different European countries, there was also considerable interest in CLEF worldwide, with 17 participants from the Americas, 19 from Asia, and 5 from Africa.

2 The CLEF Conference

CLEF 2015 redefined the focus of the CLEF conference to be on “experimental IR” as carried out at evaluation forums (CLEF Labs, TREC, NTCIR, FIRE, MediaEval, RomIP, TAC, ...) with special attention to the challenges of multimodality, multilinguality, and interactive search. We invited submissions on significant new insights demonstrated on the resulting IR test collections, on analysis of IR test collections and evaluation measures, as well as on concrete proposals to push the boundaries of the Cranfield/TREC/CLEF paradigm [9].

Keynotes Three eminent scholars in the field headlined each day of the conference, addressing different areas of evaluation and information retrieval. *Gregory Grefenstette* (INRIA Saclay, France) talked about personal information systems and personal semantics. *Mounia Lalmas* (Yahoo Labs, London, UK) addressed the topic of user engagement evaluation. *Douglas W. Oard* (University of Maryland, USA) examined issues in privacy and ethics when searching among secrets.

Table 1: Statistics of the technical program

Category	Submitted	Accepted (%)
Best of the Labs	8	8 100%
Full papers (12p) + short presentation	24	8 33% 7 29%
Short papers (6p)	36	20 55%
Total	68	43 63%

Other Evaluation Initiatives 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 TREC² (Text REtrieval Conference), the first large-scale evaluation activity organised in the field of IR, which began in 1992. *Noriko Kando* from the National Institute of Informatics (Japan) presented NTCIR³ (NII Testbeds 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 MediaEval⁴, which is dedicated to evaluating new algorithms for multimedia access and retrieval.

Technical program The clear focus on experimental IR helped the CLEF conference grow in terms of the submissions, see the details in Table 1, making the labs and conference parts of CLEF equal in size. CLEF 2015 received a total of 68 submissions, a dramatic increase over previous years. Each submission was reviewed by at least three PC members, and the two program chairs oversaw the reviewing and often extensive follow-up discussion. A novel feature of the CLEF 2015 conference was to invite CLEF 2014 lab organizers to nominate a “best of the labs” paper that was reviewed as a full paper submission to the CLEF 2015 conference according to the same review criteria and PC. This resulted in 8 full papers accepted corresponding to each to the CLEF 2014 labs. We received 24 regular full paper submissions, of which 8 (33%) full papers were accepted for regular oral presentation, and additional 7 full paper submissions (29%, making a total of 63%) was accepted with short oral and a poster presentation. We received 36 short paper submissions, and accepted 20 (55%).

Awards CLEF 2015 introduced three kinds of awards for the papers presented at the conference: a Best Short Paper Award, a Best Paper Award, a Best Student Paper Award (given only in case the best paper award has no student main author).

The Best Short Paper Award went to Wakeling and Clough [13], “*Integrating Mixed-Methods for Evaluating Information Access Systems.*” According to the jury, “I found the paper to be fascinating, extremely well written and possibly one of those papers that leads to much future work. I have no real suggestions, other than I hope that the results from this work will be presented

²<http://trec.nist.gov/>

³<http://research.nii.ac.jp/ntcir/>

⁴<http://www.multimediaeval.org/>

in the future. It is a perfect paper for CLEF.” There were honorable mentions for Abnar et al. [1], “*Meta Text Aligner: Text Alignment Based on Predicted Plagiarism Relation*,” and Shen and Nie [10], “*Is Concept Mapping useful for Biomedical Information Retrieval?*”

The Best Paper Award went to Ermakova [4], “*A Method for Short Message Contextualization: Experiments at CLEF/INEX*.” The jury commented: “This paper makes a substantial contribution to the research of complex evaluation procedures combining NLP and IR. I will read this very carefully and cite this in future work.” There were honorable mentions for four more papers: Budíková et al. [2], “*Search-based Image Annotation: Extracting Semantics from Similar Images*,” Grotov et al. [7], “*A Comparative Study of Click Models for Web Search*,” McMinn and Jose [8], “*Real-Time Entity-Based Event Detection for Twitter*,” and Xu et al. [14], “*Answering Natural Language Questions via Phrasal Semantic Parsing*.” As the best paper award was won by a student author, no separate Best Student Paper Award was handed out.

Social Program A rich social program was organised in conjunction with the conference, starting with a welcome reception with local food and wine specialities, continuing with the city hall reception, local band “*La mal Coiffée*” (bad hair) just before the social dinner in a famous organic restaurant named “*Saveur Bio*”, and master classes in (1) Traditional polyphonic singing with Bastien Zaoui from the famous Vox Bigerri group and (2) Wine and Food Pairing with Yves Cinotti.

3 The CLEF Lab Sessions

Eight laboratories were selected and run during CLEF 2015. To identify the best proposals, besides well-established criteria from previous years’ editions of CLEF such as topical relevance, novelty, potential impact on future world affairs, likely number of participants, and the quality of the organizing consortium. This year we further stressed the connection to real-life usage scenarios and we tried to avoid as much as possible overlaps among labs in order to promote synergies and integration. This was possible also thanks to a new activity introduced two years ago and supported by the ELIAS network⁵: a lab organizers and proposers meeting, co-located with the European Conference on Information Retrieval (ECIR) and held in Vienna on March 31st 2015, in order to improve the synergies and interactions among different labs and to set the stage for next year proposals.

The Labs at CLEF 2015 [3], building on previous experience, demonstrate the maturity of the CLEF evaluation environment via the incorporation of new tasks, new and larger data sets, new ways of evaluation or more languages. Details of the individual Labs are described by the Lab organizers in these proceedings, here we just provide brief comment on each one.

CLEF eHealth – ShARe/CLEF eHealth Evaluation Lab CLEFeHealth explores scenarios which aim to ease patients and nurses understanding and accessing of 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

⁵<http://www.elias-network.eu/>

lab contains two tasks: Information Extraction from Clinical Data (with two subtasks: Clinical speech recognition and Named entity recognition from clinical narratives in European languages) and User-centered Health Information Retrieval (with two subtasks: Monolingual IR and Multilingual IR).

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.

LifeCLEF The LifeCLEF lab continues image-based plant identification task which has originally run within ImageCLEF since 2011, with the same tasks of last year (BirdCLEF, PlantCLEF and FishCLEF). However, the LifeCLEF tasks radically enlarges the evaluated challenge towards multimodal data by (i) considering birds and fish in addition to plants, (ii) considering audio and video content in addition to images, (iii) scaling-up the evaluation data to hundreds of thousands of life media records and thousands of living species.

LL4IR - Living Labs for IR CLEF 2015 sees the first edition of this new lab, which features one task (Product search and web search). The main goal of the Lab is to provide a benchmarking platform for researchers to evaluate their ranking systems in a live setting with real users in their natural task environments. The lab acts as a proxy between commercial organizations (live environments) and lab participants (experimental systems), facilitates data exchange, and makes comparison between the participating systems.

NEWSREEL - News Recommendation Evaluation Lab CLEF 2015 is the second iteration of this lab. Participants can: 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. NEWSREEL provides two tasks designed to address the challenge of real-time news recommendation: Benchmark News Recommendations in a Living Lab and Benchmarking News Recommendations in a Simulated Environment.

PAN Lab on Uncovering Plagiarism, Authorship, and Social Software Misuse This is the 12th edition of the PAN lab on evaluation of uncovering plagiarism, authorship, and social software misuse. PAN offers three tasks at CLEF 2015 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, Author Profiling.

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. 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), BioASQ (Large-Scale Biomedical Semantic Indexing), and BioASQ (Biomedical Question answering).

SBS - Social Book Search The Social Book Search Lab was previously part of the INEX evaluation benchmark (since 2007). 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. SBS runs two tasks: Suggestion Track and Interactive Track.

4 CLEF 2016 and Beyond

More information on the CLEF 2015 conference, the CLEF initiative and the CLEF Association is provided on the Web:

- CLEF 2015: <http://clef2015.clef-initiative.eu/>
- CLEF initiative: <http://www.clef-initiative.eu/>
- CLEF Association: <http://www.clef-initiative.eu/association>

CLEF 2016⁶ will be hosted by the Computer Science Department of the School of Sciences and Technology of the University of Évora, Portugal, 5-8 September 2016.

The call for papers for the CLEF 2016 Conference will be released around November 2015, and the expected deadline for the submission of papers will be around April 2016.

As far as labs are concerned, CLEF 2016 will run seven evaluation activities:

CLEF eHealth Goals of the lab are to develop processing methods and resources (e.g., dictionaries, abbreviation mappings, and data with model solutions for method development and evaluation) in a multilingual setting to enrich difficult-to-understand eHealth texts and provide personalised reliable access to medical information. In addition, CLEF eHealth will provide valuable documentation; develop an evaluation setting and release evaluation results for these methods and resources. The lab will contribute to the participants and organisers' professional networks and interaction with all interdisciplinary actors of the ecosystem for producing, processing, and consuming eHealth information.

ImageCLEF Following the ImageCLEF 2003–2015 evaluation campaigns, the ImageCLEF 2016 lab aims to provide an evaluation forum for the language independent annotation and retrieval of images. Motivated by the need to support potentially multilingual users from a global community accessing the ever growing body of visual information, the main goal of ImageCLEF 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 systems operating in monolingual, language-independent and multimodal contexts,

⁶<http://clef2016.clef-initiative.eu/>

providing reusable resources for benchmarking. To meet its objectives, ImageCLEF organises tasks that benchmark the annotation and retrieval of diverse images such as general photographic and medical images, a domain with a particularly active community.

LifeCLEF This lab 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 second edition of the lab focuses on three use-cases and one specific notion of what a living lab is, with a view to expanding to other use-cases and other interpretations of living labs in subsequent years. Use-cases for the second lab are: product search (through REGIO Jatek, an e-commerce site), web search (through Seznam, a large scale commercial web search engine), and academic search (through SSOAR a Social Science Open Access Repository). The first two use-cases are a continuation of last year (see Section 6 for details), the last use-case is new this year.

News Recommendation Evaluation Lab (NEWSREEL) In the third iteration of the lab, we provide two tasks that address the challenge of real-time news recommendation. Participants could: a) Benchmark News Recommendations in a Living Lab, i.e., researchers gain access to the resources of a company to evaluate different recommendation techniques using A/B testing; b) Benchmark News Recommendations in a simulated environment, where we will provide a large data set comprising interactions between users and various news portals in a two-month time span. Participants in this task have to predict users' clicks on recommended news articles in simulated realtime.

Uncovering Plagiarism, Authorship and Social Software Misuse (PAN) This lab will focus on authorship tasks from the fields of author identification, author profiling, and author obfuscation. More specifically, the tasks will include two variants per field, namely author clustering and linking, age and gender prediction, and author masking and imitation.

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 Task, Interactive Task, and Mining Task.

and two workshops/lab incubators:

CLEF Cultural Microblog Contextualization (CMC) The CMC task aims at focusing on different ways to gather, organize, and deliver relevant social data related to events that generate large numbers of microblog messages. Such structuration for gathering broader views of the social impact of an event. The goal of this CLEF Cultural Microblog Lab is to develop processing methods and resources to mine the social media sphere surrounding cultural events such as festivals. Tweets linked to an event makes a dense, rich but very noisy

corpus: informal language, out of the language phrases and symbols, hashtags, hyperlinks
The information is also often imprecise, duplicate, or non-informative. The interest of mining such data is to extract relevant, and informative content, as well as to potentially discover new information.

Multi-Perspective Health Information Search (MPHIS) Term “Multi-Perspective Health Information Search” (MPHIS) denotes queries for which there is “No Single Best Correct Answer”. Instead multiple and diverse perspectives/points of view (which very often are contradictory in nature) are available on the web regarding the queried information. The MPHIS workshop investigates complex health information search in scenarios where users search for health information with more than just a single correct answer, and look for multiple perspectives from diverse sources both from medical research and from real world patient narratives. The goal of MPHIS is to research and develop techniques to support users in complex multi-perspective health information queries.

More information about CLEF 2016 are available on its homepage at:

<http://clef2016.clef-initiative.eu/>

CLEF 2017 will be jointly hosted by Dublin City University and Trinity College Dublin, Ireland, in early September 2017.

Finally, bids for hosting CLEF 2018 are now open and will close around April 2016. Proposals can be sent to the CLEF Steering Committee Chair at chair@clef-initiative.eu.

Acknowledgments

The success of CLEF 2015 would not have been possible without the huge effort of several people and organizations, including the CLEF Association⁷, the Program Committee, the Lab Organizing Committee, the local organization committee in Toulouse, the reviewers, and the many students and volunteers who contributed along the way.

We would like to acknowledge the *Institut de Recherche en Informatique de Toulouse* UMR 5505 CNRS and its director, Prof. Michel Daydé for the support we got, first for bidding to host the conference, then to organize it. We also received the support from the following universities and schools: Ecole Supérieure du Professorat et de l'Education, Université Toulouse Jean Jaurés, Université Paul Sabatier, Université du Capitole.

We also gratefully acknowledge the support we received from our sponsors. The ESF Research Networking Programme ELIA⁸, the ACM SIGIR⁹, the *Université Toulouse-Jean Jaurés*¹⁰, and the *Région Midi-Pyrénées*¹¹ for their strong financial support; but also: Springer¹², the *Université*

⁷<http://www.clef-initiative.eu/association>

⁸<http://elias-network.eu/>

⁹<http://sigir.org/>

¹⁰<http://www.univ-tlse2.fr/>

¹¹<http://www.midipyrenees.fr/Midi-Pyrenees,3650>

¹²<http://www.springer.com/>

Paul Sabatier, Institut de Recherche en Informatique de Toulouse UMR 5505 CNRS, INFORSID, Université Toulouse Capitole, EGC, and ARIA.

The level of sponsorship allowed us to offer 20 grants for students in addition to a free registration for the 25 volunteers including 11 further students.

Last but not least without the important and tireless effort of the enthusiastic and creative authors, the organizers of the selected labs, the colleagues and friends involved in running them, and the participants who contribute their time to making the labs and the conference a success, as well as financially supporting them through the CLEF Association, CLEF would not be possible.

Thank you all very much!

References

- [1] S. Abnar, M. Dehghani, and A. Shakery. Meta text aligner: Text alignment based on predicted plagiarism relation. In Mothe et al. [9], pages 193–199. URL http://dx.doi.org/10.1007/978-3-319-24027-5_16.
 - [2] P. Budíková, M. Batko, J. Botorek, and P. Zezula. Search-based image annotation: Extracting semantics from similar images. In Mothe et al. [9], pages 327–339. URL http://dx.doi.org/10.1007/978-3-319-24027-5_36.
 - [3] L. Cappellato, N. Ferro, G. J. F. Jones, and E. SanJuan, editors. *Working Notes of CLEF 2015 - Conference and Labs of the Evaluation forum, Toulouse, France, September 8-11, 2015*, volume 1391 of *CEUR Workshop Proceedings*, 2015. CEUR-WS.org. URL <http://ceur-ws.org/Vol-1391>.
 - [4] L. Ermakova. A method for short message contextualization: Experiments at CLEF/INEX. In Mothe et al. [9], pages 352–363. URL http://dx.doi.org/10.1007/978-3-319-24027-5_38.
 - [5] N. Ferro. CLEF 15th birthday: Past, present, and future. *SIGIR Forum*, 48(2):31–55, 2014. URL <http://doi.acm.org/10.1145/2701583.2701587>.
 - [6] N. Ferro and G. Silvello. CLEF 15th birthday: What can we learn from ad hoc retrieval? In *Information Access Evaluation. Multilinguality, Multimodality, and Interaction - 5th International Conference of the CLEF Initiative, CLEF 2014, Sheffield, UK, September 15-18, 2014. Proceedings*, pages 31–43, 2014. URL http://dx.doi.org/10.1007/978-3-319-11382-1_4.
 - [7] A. Grotov, A. Chuklin, I. Markov, L. Stout, F. Xumara, and M. de Rijke. A comparative study of click models for web search. In Mothe et al. [9], pages 78–90. URL http://dx.doi.org/10.1007/978-3-319-24027-5_7.
 - [8] A. J. McMinn and J. M. Jose. Real-time entity-based event detection for twitter. In Mothe et al. [9], pages 65–77. URL http://dx.doi.org/10.1007/978-3-319-24027-5_6.
 - [9] J. Mothe, J. Savoy, J. Kamps, K. Pinel-Sauvagnat, G. J. F. Jones, E. SanJuan, L. Cappellato, and N. Ferro, editors. *Experimental IR Meets Multilinguality, Multimodality, and Interaction - 6th International Conference of the CLEF Association, CLEF 2015, Toulouse, France, September 8-11, 2015, Proceedings*, volume 9283 of *Lecture Notes in Computer Science*, 2015. Springer. URL <http://dx.doi.org/10.1007/978-3-319-24027-5>.
-

-
- [10] W. Shen and J. Nie. Is concept mapping useful for biomedical information retrieval? In Mothe et al. [9], pages 281–286. URL http://dx.doi.org/10.1007/978-3-319-24027-5_29.
- [11] T. Tsirikika, A. G. S. de Herrera, and H. Müller. Assessing the scholarly impact of imageclef. In *Multilingual and Multimodal Information Access Evaluation - Second International Conference of the Cross-Language Evaluation Forum, CLEF 2011, Amsterdam, The Netherlands, September 19-22, 2011. Proceedings*, pages 95–106, 2011. URL http://dx.doi.org/10.1007/978-3-642-23708-9_12.
- [12] T. Tsirikika, B. Larsen, H. Müller, S. Endrullis, and E. Rahm. The scholarly impact of CLEF (2000-2009). In *Information Access Evaluation. Multilinguality, Multimodality, and Visualization - 4th International Conference of the CLEF Initiative, CLEF 2013, Valencia, Spain, September 23-26, 2013. Proceedings*, pages 1–12, 2013. URL http://dx.doi.org/10.1007/978-3-642-40802-1_1.
- [13] S. Wakeling and P. D. Clough. Integrating mixed-methods for evaluating information access systems. In Mothe et al. [9], pages 306–311. URL http://dx.doi.org/10.1007/978-3-319-24027-5_33.
- [14] K. Xu, Y. Feng, S. Huang, and D. Zhao. Question answering via phrasal semantic parsing. In Mothe et al. [9], pages 414–426. URL http://dx.doi.org/10.1007/978-3-319-24027-5_43.
-