

Report on CLEF 2017: Experimental IR Meets Multilinguality, Multimodality, and Interaction

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

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

Lorraine Goeuriot
Université Grenoble Alpes, France
lorraine.goeuriot@univ-grenoble-alpes.fr

Julio Gonzalo
National Distance Education University (UNED), Spain
julio@lsi.uned.es

Gareth J. F. Jones
Dublin City University, Ireland
gareth.jones@computing.dcu.ie

Liadh Kelly
Maynooth University, Ireland
liadh.kelly@mu.ie

Séamus Lawless
Trinity College Dublin, Ireland
seamus.lawless@scss.tcd.ie

Thomas Mandl
University of Hildesheim, Germany
mandl@uni-hildesheim.de

Abstract

This is a report on the eighth edition of the *Conference and Labs of the Evaluation Forum* (CLEF 2017), held in early September 2017, in Dublin, Ireland. CLEF was a four day event combining a Conference and an Evaluation Forum. The Conference featured keynotes by Leif Azzopardi and Vincent Wade, and presentation of 32 peer reviewed research papers covering a wide range of topics. The Evaluation Forum consisted to eight Labs and two workshops: eHealth, ImageCLEF, LifeCLEF, NEWSREEL, PAN, Cultural Microblog Contextualization, Early Risk Prediction on the Internet, Dynamic Search for Complex Tasks, Personalised Information Retrieval at CLEF, and Multimodal Spatial Role Labelling addressing a wide range of tasks, media, languages, and ways to go beyond standard test collections.

1 Introduction

The 2017 edition of the *Conference and Labs of the Evaluation Forum*¹ (CLEF) was hosted by the ADAPT Centre², Dublin City University and Trinity College Dublin from the 11th to 14th September 2017.

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 [3], and achieved a substantial scholarly impact [7, 8]

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 [2].

CLEF 2017 was the 8th year of the CLEF Conference and the 18th year of the CLEF initiative as a forum for Information Retrieval (IR) Evaluation. CLEF 2017 was attended by 157 participants (including 13 (8%) joint registrations for both CLEF and MediaEval) out of which 59 (38%) were students, denoting a young and vibrant community, from different academic institutions and industrial organisations. Although the majority (78%) of the participants came from different European countries and Russia, there was also considerable interest in CLEF worldwide, with 11% participants from the Americas, 6% from Asia and Australia, and 1% from Africa while 4% did not specify their country.

2 The CLEF Conference

CLEF 2017 continued the focus of the CLEF conference 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,

¹<http://clef2017.clef-initiative.eu/>

²<http://adaptcentre.ie/>

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 [4].

Keynotes Two eminent scholars in the field delivered keynote speeches, addressing different areas of evaluation and information retrieval. *Leif Azzopardi* (University of Stratclyde, UK) talked about how to leverage simulation of user interaction in experimental evaluation. *Vincent Wade* (Trinity College Dublin, Ireland) addressed the topic of personalised information systems and their evaluation.

Joint CLEF/MediaEval Panel on Horizons of Evaluation This year’s conference was also co-located with MediaEval³ and the program included joint sessions between both MediaEval and CLEF to allow for cross fertilisation. In particular a joint CLEF/MediaEval panel discussed horizons of evaluation.

Other Evaluation Initiatives The conference also hosted a session on *other evaluation initiatives*, which informed delegates of current and upcoming activities within evaluation initiatives worldwide. The session included *Ian Soboroff* from the National Institute of Standards and Technology (USA) who presented TREC⁴ (Text REtrieval Conference), the first large-scale evaluation activity organised in the field of IR, which began in 1992, and TRECVID⁵. *Noriko Kando* from the National Institute of Informatics (Japan) presented NTCIR⁶ (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. *Paolo Rosso* from Universitat Politècnica de València (Spain) presented FIRE (Forum for Information Retrieval Evaluation)⁷, which is dedicated to evaluating information access with a focus on the Indian sub-continent.

Technical Program CLEF 2017 received a total of 38 submissions, of which a total of 22 papers were accepted. Each submission was reviewed by PC members, and the program chairs oversaw the reviewing and follow-up discussions. CLEF 2017 continued a novel track introduced at CLEF 2015, that of inviting CLEF lab organizers to nominate a team from their 2016 lab to submit a “best of the labs” paper that was reviewed as a full paper submission to the CLEF 2017 conference according to the same review criteria and PC. In total, 15 long papers were received, of which 7 were accepted. 17 short papers were received, of which 9 were accepted. 6 Best of Labs track papers were received, all of which were accepted.

Awards CLEF 2017 continued the tradition introduced at CLEF 2015 of awarding the papers presented at the conference with both a Best Short Paper Award and a Best Paper Award.

The Best Short Paper Award went to “*Evaluation of Hierarchical Clustering via Markov Decision Processes for Efficient Navigation and Search*” by Raul Moreno, Weipeng Huang, Arjumand

³<http://www.multimediaeval.org/>

⁴<http://trec.nist.gov/>

⁵<http://trecvid.nist.gov/>

⁶<http://research.nii.ac.jp/ntcir/>

⁷<http://fire.irsilab.com/>

Younus, Michael O’Mahony and Neil J. Hurley [6]. The Best Paper Award went to “*An Analysis of Cross-Genre and In-Genre Performance for Author Profiling in Social Media*” by Maria Medvedeva, Hessel Haagsma and Malvina Nissim [5].

Social Program CLEF 2017 was accompanied by a social programme encompassing some of Dublin’s most popular locations. The Welcome Reception took place at the Guinness Storehouse, Ireland’s most popular tourist attraction, and included an introduction to the brewing of Guinness, an exhibition of the famous cartoon advertising campaigns, and the main reception in the Gravity Bar with panoramic views across the city. The conference dinner was held jointly with MediaEval in the Dining Hall at Trinity College Dublin. Participants were also able to join a Literary Pub Crawl exploring Dublin’s historic literary tradition and its social settings.

3 The CLEF Lab Sessions

Eight laboratories and two workshops were selected and ran during CLEF 2017. To identify the best proposals, well established criteria from previous editions of CLEF were applied like, for example, 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.

The Labs at CLEF 2017 [1], 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 their proceedings, here we just provide a brief overview of each one.

News Recommendation Evaluation Lab (NEWSREEL)⁸ provides a vehicle for the IR/recommender system communities to move from conventional offline evaluation to online evaluation. The lab addresses the following information access challenge: Whenever a visitor of an online news portal reads a news article on their side, the task is to recommend other news articles that the user might be interested in. Many of the recommendations of the participants enter into the live system and the team receives immediate feedback about click data of the user.

LifeCLEF⁹ aims at boosting research on the identification of living organisms and on the production of biodiversity data in general. Through its biodiversity informatics related challenges, LifeCLEF aims to push the boundaries of the state-of-the-art in several research directions at the frontier of multimedia information retrieval, machine learning and knowledge engineering. For a task with over 10.000 classes, the best systems reached an excellent performance with over 0.9 Mean Reciprocal Rank.

Uncovering Plagiarism, Authorship and Social Software Misuse (PAN)¹⁰ provides evaluation of uncovering plagiarism, authorship, and social software misuse. PAN offered three tasks at CLEF 2017 with new evaluation resources consisting of large-scale corpora, performance

⁸<http://clef-newsreel.org/>

⁹<http://www.lifeclef.org/>

¹⁰<http://pan.webis.de/>

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: author identification; author profiling; and, author obfuscation. The obfuscation task measures how well systems can reformulate a text to make author identification impossible.

CLEFeHealth¹¹ provides 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 tasks are: multilingual Information extraction; technologically assisted reviews in empirical medicine; and, patient-centred information retrieval.

Cultural Microblog Contextualization (CMC) Workshop¹² deals with how cultural context of a microblog affects its social impact at large. This involves microblog search, classification, filtering, language recognition, localization, entity extraction, linking open data and summarization. Regular lab participants have access to the private massive multilingual microblog stream of the festival galleries project. One of the goals was the illustration of a timeline for an event.

ImageCLEF¹³ organises 3 main tasks with a global objective of benchmarking lifelogging retrieval and summarization, tuberculosis type prediction from CT images and bio-medical image caption prediction; and a pilot task on remote sensing image analysis. This task required the estimation of population density based on satellite images for some African regions.

Early risk prediction on the Internet (eRisk)¹⁴ explores issues of evaluation methodology, effectiveness metrics and other processes related to early risk detection. Early detection technologies can be employed in different areas, particularly those related to health and safety. For instance, early alerts could be sent when a predator starts interacting with a child for sexual purposes, or when a potential offender starts publishing antisocial threats on a blog, forum or social network. The main goal is to pioneer a new interdisciplinary research area that would be potentially applicable to a wide variety of situations and to many different personal profiles. The task in 2017 consisted of identifying depressed people based on their texts in Social Media.

Personalised Information Retrieval at CLEF (PIR-CLEF)¹⁵ provides a framework for evaluation of Personalised Information Retrieval (PIR). Current approaches to the evaluation of PIR are user-centered, i.e., they rely on experiments that involve real users in a supervised environment, PIR-CLEF aims to develop and demonstrate a methodology for evaluation PIR which enables repeatable experiments to enable the detailed exploration of personal models and their exploitation in IR.

Dynamic Search for Complex Tasks¹⁶ Information Retrieval research has traditionally focused on serving the best results for a single query – so-called ad hoc retrieval. However,

¹¹<https://sites.google.com/site/clefehealth2017/>

¹²<https://mc2.talne.eu/>

¹³<http://imageclef.org/2017>

¹⁴<http://early.irlab.org/>

¹⁵<http://www.ir.disco.unimib.it/pirclef2017/>

¹⁶<https://ekanou.github.io/dynamicsearch/>

users typically search iteratively, refining and reformulating their queries during a session. A key challenge in the study of this interaction is the creation of suitable evaluation resources to assess the effectiveness of IR systems over sessions. The goal of the CLEF Dynamic Search lab is to propose and standardize an evaluation methodology that can lead to reusable resources and evaluation metrics able to assess retrieval performance over an entire session, keeping the “user” in the loop.

Multimodal Spatial Role Labeling¹⁷ explores the extraction of spatial information from two information resources that is image and text. This is important for various applications such as semantic search, question answering, geographical information systems and even in robotics for machine understanding of navigational instructions or instructions for grabbing and manipulating objects.

More information on the CLEF 2017 conference, the CLEF initiative and the CLEF Association is provided on the Web:

- CLEF 2017: <http://clef2017.clef-initiative.eu/>
- CLEF initiative: <http://www.clef-initiative.eu/>
- CLEF Association: <http://www.clef-initiative.eu/association>

4 CLEF 2018 and Beyond

CLEF 2018 will be hosted by the University of Avignon, France, 10-14 September 2018, and it is jointly organized by UAPV, AMU and Toulon University.

The call for papers for the CLEF 2018 Conference¹⁸ has been released and the expected deadlines are:

- Submission of abstracts / intentions: 20 April 2018
- Submission of Long Papers: 7 May 2018
- Submission of Short Papers: 14 May 2018
- Notification of Acceptance: 8 June 2018
- Camera Ready Copy due: 22 June 2018

As far as labs are concerned, CLEF 2018 will run ten evaluation activities: 8 will be a continuation of the labs running during CLEF 2017 and two will be new pilot labs.

The continued activities are:

- **CLEF eHealth**¹⁹: medical content is available electronically in a variety of forms ranging from patient records and medical dossiers, scientific publications and health-related websites

¹⁷http://www.cs.tulane.edu/~pkordjam/mSpRL_CLEF_lab.htm/

¹⁸<http://clef2018.clef-initiative.eu/index.php?page=Pages/cfp.html>

¹⁹<https://sites.google.com/site/clefehealth2018/>

to medical-related topics shared across social networks. This lab aims to support the development of techniques to aid laypeople, clinicians and policy-makers in easily retrieving and making sense of medical content to support their decision making. *Task 1 - Multilingual Information Extraction*: Participants will be required to extract the causes of death from death certificates, authored by physicians in European languages. This can be seen as a named entity recognition, normalization, and/or text classification task. *Task 2 - Technologically Assisted Reviews in Empirical Medicine*: Participants will be challenged to retrieve medical studies relevant to conducting a systematic review on a given topic. This can be seen as a total recall problem and is addressed by both query generation and document ranking. *Task 3 - Patient-centred Information Retrieval*: Participants must retrieve web pages that fulfil a given patients personalised information need. This needs to fulfil the following criteria: information reliability, quality, and suitability. The task also has a multilingual querying track.

- **ImageCLEF**²⁰: provides an evaluation forum for the language independent annotation and retrieval of images, a domain for which tools are by far not as advanced as for text analysis and retrieval. *Task 1 - ImageCLEFlifelog*: An increasingly wide range of personal devices, such as smartphones, video cameras as well as wearable devices that allow capturing pictures, videos, and audio clips in every moment of our life are becoming available. The task addresses the problems of lifelogging data understanding, summarization and retrieval. *Task 2 - ImageCLEFcaption*: Interpreting and summarizing the insights gained from medical images such as radiology output is a time-consuming task that involves highly trained experts and often represents a bottleneck in clinical diagnosis pipelines. The task addresses the problem of bio-medical image concept detection and caption prediction from large amounts of training data. *Task 3 - ImageCLEFtuberculosis*: The objective of this task is to determine tuberculosis subtypes and drug resistances, as far as possible automatically, from the volumetric image information in computed tomography (CT) volumes (mainly texture analysis) and based on clinical information (e.g., age, gender, etc). *Task 4 - VisualQuestionAnswering*: With the ongoing drive for improved patient engagement and access to the electronic medical records via patient portals, patients can now review structured and unstructured data from labs and images to text reports associated with their healthcare utilization. Given a medical image accompanied with a set of clinically relevant questions, participating systems are tasked with answering the questions based on the visual image content.
- **LifeCLEF**²¹: aims at boosting research on the identification of living organisms and on the production of biodiversity data. Through its biodiversity informatics related challenges, LifeCLEF is intended to push the boundaries of the state-of-the-art in several research directions at the frontier of multimedia information retrieval, machine learning and knowledge engineering. The lab is organized around three tasks: *Task 1 - GeoLifeCLEF*: location-based species recommendation. *Task 2 - BirdCLEF*: bird species identification from bird calls and songs. *Task 3 - ExpertLifeCLEF*: experts vs. machines identification quality.

²⁰<http://www.imageclef.org/2018/>

²¹<http://lifeclef.org/>

-
- **PAN Lab on Digital Text Forensics**²²: is a series of scientific events and shared tasks on digital text forensics. *Task 1 - Author Identification*: cross-domain authorship attribution. More specifically, cases where the topic of texts varies significantly will be examined. In addition, we will continue the pilot task of style change detection, focusing on finding switches of authors within documents based on an intrinsic style analysis. *Task 2 - Author Obfuscation*: while the goal of author identification and author profiling is to model author style so as to deanonymize authors, the goal of author obfuscation technology is to prevent that by disguising the authors. We will study author masking vs. authorship verification. *Task 3 - Author Profiling*: the goal is to identify an author's traits based on their writing style. The focus will be on age and gender, whereas text and image will be used as information sources, offering tweets in English, Spanish and Arabic.
 - **Multilingual Cultural Mining and Retrieval (MC2)**²³: developing processing methods and resources to mine the social media sphere surrounding cultural events such as festivals. This requires participants to deal with diverse languages and dialects as well as informal expressions. There are three tasks: *Task 1 - Cross Language Cultural Retrieval over MicroBlogs*: a) Small Microblogs Multilingual Information Retrieval in Arabic, English, French and Latin languages; b) Microblogs Bilingual Information Retrieval for tuning systems running on language pairs; c) Microblog Monolingual Information Retrieval based on 2017 language identification. *Task 2 - Mining Opinion Argumentation*: a) Polarity detection in microblogs; b) Automatic identification of argumentation elements over Microblogs and Wikipedia; c) Classification and summarization of arguments in texts. *Task 3 - Dialectal Focus Retrieval*: a) Arabic dialects in Blogs, MicroBlogs and Video News transcriptions; b) Spanish language variations in Blogs, MicroBlog and Journals.
 - **Early risk prediction on the Internet (eRisk)**²⁴: explores the evaluation methodology, effectiveness metrics and practical applications (particularly those related to health and safety) of early risk detection on the Internet. *Task 1 - Early Detection of Signs of Depression*: the challenge consists of sequentially processing pieces of evidence (Social Media entries) and detect early traces of depression as soon as possible. *Task 2 - Early Detection of Signs of Anorexia*: the challenge consists of sequentially processing pieces of evidence (Social Media entries) and detect early traces of anorexia as soon as possible. Both tasks are mainly concerned about evaluating Text Mining solutions and, thus, we concentrate on texts written in Social Media. Texts should be processed in the order in which they were posted. In this way, systems that effectively perform this task could be applied to sequentially monitor user interactions in blogs, social networks, or other types of online media.
 - **Evaluation of Personalised Information Retrieval (PIR-CLEF)**²⁵: the primary aim of the PIR-CLEF 2018 laboratory is: 1) to facilitate comparative evaluation of PIR by offering participating research groups a mechanism for evaluation of their personalisation algorithms; 2) to give the participating groups the means to formally define and evaluate their

²²<http://pan.webis.de/>

²³<https://mc2.talne.eu/>

²⁴<http://early.irlab.org/>

²⁵<http://www.ir.disco.unimib.it/pir-clef2018/>

own and novel user profiling approaches for PIR. *Task 1 - Personalized Search*: we will provide a bag-of-words profile gathered during the query sessions performed by real searchers, the set of queries formulated by each user, together with the corresponding document relevance, and the the search logs of each user. Task participants will be expected to compute search results obtained by applying their personalization algorithms on these queries. The search will be carried out on the ClueWeb12 collection, by using the API provided by DCU. *Task 2 - User Profile Models*: participants will be required to develop their own user profile models using the information gathered about the real user during her interactions with the system. The same information have been used for creating the baseline (keyword-based user profiles), which is provided in the benchmark.

- **Dynamic Search for Complex Tasks (DynSe)**²⁶: the primary aim of the CLEF Dynamic Search Lab is to develop algorithms which interact dynamically with user (or other algorithms) towards solving a task, and evaluation methodologies to quantify their effectiveness. The lab is organized along two tasks: *Task 1 - Query Suggestion*: given a verbose topic description participants will generate and submit a sequence of queries and a ranking of the collection for each query. Queries will be evaluated over their effectiveness (query agent) and/or resemblance to user queries (user simulation). Query suggestion will be performed iteratively. *Task 2 - Result Composition*: given the obtained results from the aforementioned queries obtain a single ranked list by merging the individual rankings.

The two new labs are:

- **CLEF/NTCIR/TREC REproducibility (CENTRE@CLEF 2018)**²⁷: the goal of CENTRE@CLEF 2018 is to run a joint CLEF/NTCIR/TREC task on challenging participants: 1) to reproduce best results of best/most interesting systems in previous editions of CLEF/NTCIR/TREC by using standard open source IR systems; 2) to contribute back to the community the additional components and resources developed to reproduce the results in order to improve existing open source systems. *Task 1 - Replicability*: replicability of selected methods on the same experimental collections. *Task 2 - Reproducibility*: reproducibility of selected methods on the different experimental collections. *Task 3 - Re-reproducibility*: using the components developed in T1 and T2 and made available by the other participants to replicate/reproduce their results.
- **Automatic Identification and Verification of Political Claims (CheckThat!)**²⁸: aims to foster the development of technology capable of both spotting and verifying check-worthy claims in political debates in English and Arabic. *Task 1 - Check-Worthiness*: given a political debate, which is segmented into sentences with speakers annotated, identify which statements (claims) should be prioritized for fact-checking. This will be a ranking problem, and systems will be asked to produce a score, according to which the ranking will be performed. *Task 2 - Factuality*: given a list of already-extracted claims, classify them with factuality labels (e.g., true, half-true, false). This task will be run in an open mode. We will

²⁶<https://ekanou.github.io/dynamicsearch/>

²⁷<http://www.centre-eval.org/>

²⁸<http://alt.qcri.org/clef2018-factcheck/>

not provide any pre-selected set of documents to support the veracity labels. Participants will be free to use whatever resources they have and the Web in general, with the exception of the websites used by the organizers to collect the data.

The registration to the labs is open²⁹ and the overall schedule for the CLEF 2018 evaluation cycle is:

- Registration opens: 8 November 2017
- Registration closes: 27 April 2018
- End Evaluation Cycle: 11 May 2018
- Submission of Participant Papers: 31 May 2018

More information about CLEF 2018 is available at:

<http://clef2018.clef-initiative.eu/>

CLEF 2019 will be hosted by University of Lugano, Switzerland, in early September 2019.

Finally, bids for hosting CLEF 2020 are now open and will close around May 2018. Proposals can be sent to the CLEF Steering Committee Chair at chair@clef-initiative.eu.

Acknowledgments

The success of CLEF 2017 would not have been possible without the huge effort of several people and organizations, including the CLEF Association³⁰, the Program Committee, the Lab Organizing Committee, the local organization committee in Dublin, the reviewers, and the many students and volunteers who contributed along the way.

We gratefully acknowledge the support we received from our sponsors: Fáilte Ireland³¹, Meet In Ireland³², The ADAPT Centre³³, Science Foundation Ireland³⁴, and ACM SIGIR³⁵.

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!

²⁹<http://clef2018-labs-registration.dei.unipd.it/>

³⁰<http://www.clef-initiative.eu/association>

³¹<http://www.failteireland.ie/>

³²<http://www.meetinireland.com/>

³³<http://www.adaptcentre.ie/>

³⁴<http://www.sfi.ie/>

³⁵<http://sigir.org/>

References

- [1] L. Cappellato, N. Ferro, L. Goeuriot, and T. Mandl, editors. *CLEF 2017 Working Notes*. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-1866/>, 2017.
 - [2] N. Ferro. CLEF 15th Birthday: Past, Present, and Future. *SIGIR Forum*, 48(2):31–55, December 2014.
 - [3] N. Ferro and G. Silvello. 3.5K runs, 5K topics, 3M assessments and 70M measures: What trends in 10 years of Adhoc-ish CLEF? *Information Processing & Management*, 53(1):175–202, January 2017.
 - [4] G. J. F. Jones, S. Lawless, J. Gonzalo, L. Kelly, L. Goeuriot, T. Mandl, L. Cappellato, and N. Ferro, editors. *Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Eighth International Conference of the CLEF Association (CLEF 2017)*. Lecture Notes in Computer Science (LNCS) 10456, Springer, Heidelberg, Germany, 2017.
 - [5] M. Medvedeva, H. Haagsma, and M. Nissim. An Analysis of Cross-Genre and In-Genre Performance for Author Profiling in Social Media. In Jones et al. [4], pages 211–223.
 - [6] R. Moreno, W. Huang, A. Younus, M. O’Mahony, and N. Hurley. Evaluation of Hierarchical Clustering via Markov Decision Processes for Efficient Navigation and Search. In Jones et al. [4], pages 125–131.
 - [7] T. Tsikrika, A. Garcia Seco de Herrera, and H. Müller. Assessing the Scholarly Impact of ImageCLEF. In P. Forner, J. Gonzalo, J. Kekäläinen, M. Lalmas, and M. de Rijke, editors, *Multilingual and Multimodal Information Access Evaluation. Proceedings of the Second International Conference of the Cross-Language Evaluation Forum (CLEF 2011)*, pages 95–106. Lecture Notes in Computer Science (LNCS) 6941, Springer, Heidelberg, Germany, 2011.
 - [8] T. Tsikrika, B. Larsen, H. Müller, S. Endrullis, and E. Rahm. The Scholarly Impact of CLEF (2000–2009). In P. Forner, H. Müller, R. Paredes, P. Rosso, and B. Stein, editors, *Information Access Evaluation meets Multilinguality, Multimodality, and Visualization. Proceedings of the Fourth International Conference of the CLEF Initiative (CLEF 2013)*, pages 1–12. Lecture Notes in Computer Science (LNCS) 8138, Springer, Heidelberg, Germany, 2013.
-