

CONFERENCE REPORT - Report on CLEF 2018: Experimental IR Meets Multilinguality, Multimodality, and Interaction

Patrice Bellot, Linda Cappellato, Nicola Ferro, Josiane Mothe, Fionn Murtagh, Jian-Yun Nie, Eric Sanjuan, Laure Soulier, Chiraz Trabelsi

▶ **To cite this version:**

Patrice Bellot, Linda Cappellato, Nicola Ferro, Josiane Mothe, Fionn Murtagh, et al.. CONFERENCE REPORT - Report on CLEF 2018: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Sigir Forum, 2019, 52 (2), pp.72-82. 10.1145/3308774.3308785 . hal-02442736

HAL Id: hal-02442736
https://hal.science/hal-02442736

Submitted on 17 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Report on CLEF 2018: Experimental IR Meets Multilinguality, Multimodality, and Interaction

Patrice Bellot
Aix-Marseille Université - CNRS LIS, France
patrice.bellot@univ-amu.fr

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

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

Josiane Mothe
Université de Toulouse, IRIT, France
josiane.mothe@irit.fr

Fionn Murtagh
University of Huddersfield, UK
fionn.murtagh@cs.rhul.ac.uk

Jian-Yun Nie
DIRO, Université de Montréal, Canada
nie@iro.umontreal.ca

Eric SanJuan
LIA, UAPV, France
eric.sanjuan@univ-avignon.fr

Laure Soulier
LIP6, UPMC, France
laure.soulier@lip6.fr

Chiraz Trabelsi
University of Tunis El Manar, Tunis
chiraz.latiri3@gmail.com

Abstract

This is a report on the ninth edition of the *Conference and Labs of the Evaluation Forum* (CLEF 2018), held in early September 2018, in Avignon, France. CLEF was a four day event combining a Conference and an Evaluation Forum. The Conference featured keynotes by Nicholas Belkin, Julio Gonzalo, and Gabriella Pasi, and presentation of 29 peer reviewed research papers covering a wide range of topics in addition to many posters. The Evaluation Forum consisted to ten Labs: CENTRE, CheckThat, DynSe, eRisk, eHealth, ImageCLEF, LifeCLEF, Cultural Microblog Contextualization, PAN, and PIR-CLEF, addressing a wide range of tasks, media, languages, and ways to go beyond standard test collections.

1 Introduction

The 2018 edition of the *Conference and Labs of the Evaluation Forum*[1] (CLEF) was jointly organized by Avignon, Marseille, and Toulon Universities and was hosted by the University of Avignon from the 10th to 14th September 2018.

CLEF was established almost 20 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 sizable improvements in the performance of monolingual, bilingual, and multilingual information access systems [5], and achieved a substantial scholarly impact [16, 17]

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 [3].

CLEF 2018 was the 9th year of the CLEF Conference and the 19th year of the CLEF initiative as a forum for Information Retrieval (IR) Evaluation. CLEF 2018 was attended by 175 participants out of which 97 (55%) were students, denoting a young and vibrant community, from different academic institutions and industrial organizations. Although the majority (68%) of the participants came from different European countries and Russia, there was also considerable interest in CLEF worldwide, with 6% participants from the Americas, 11% from Asia and Australia, and 5% from Africa while 10% did not specify their country. Thanks to SIGIR and ARIA sponsoring, under the coordination of the sponsor chair Malek Hajjem (University of Tunis El Manar), 23 overseas students got a grant that covered conference fees, accommodation and meals for 5 days.

2 The CLEF Conference

CLEF 2018 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

[1] `http://clef2018.clef-initiative.eu/`

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 [1].

Keynotes and Other Evaluation Initiatives Three eminent scholars in the field delivered keynote speeches, addressing different areas of evaluation and information retrieval. *Gabriella Pasi* (University of Milano-Bicocca, Italia) talked about evaluation of (personalized) search engines and recommender systems. *Nicholas Belkin* (Rutgers University, NJ, USA) talked about evaluation of personalization of information interaction in an era of information ubiquity. Finally, *Julio Gonzalo* (UNED, Spain) discussed about bias in system evaluation. Talks were streamed live and recorded on CanalU[2].

CLEF 2018 also includes presentations on other evaluation initiatives. *Noriko Kando* from the National Institute of Informatics (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.

Technical Program CLEF 2018 received a total of 39 submissions, of which a total of 13 papers (9 long, 4 short) were accepted. Each submission was reviewed by three program committee members, and the program chairs oversaw the reviewing and follow-up discussions. Thirteen different countries are represented in the accepted papers. Many contributions this year tackle the medical e-health multimedia retrieval challenges in different ways: from medical image analysis to query suggestion. However, there are many other topics of research in the accepted papers such as document clustering, social biases in IR, social book search, personality profiling, to cite a few. Like in previous editions since 2015, CLEF 2018 continued inviting CLEF lab organizers to nominate a "best of the labs" paper that was reviewed as a full paper submission to the CLEF 2018 conference according to the same review criteria and PC. Among the 9 invited papers, 6 were accepted as long and 3 as short. Finally 8 posters have also been accepted. although they are not included in the LNCS volume, posters give the opportunity to their authors to discuss their research during the conference and are accessible through the web pages of the conference.

Master Classes Master classes were organized on Friday by the ARIA just after CLEF closure. They were opened to CLEF participants and ARIA members. Julio Gonzalo gave a lecture about *the art of evaluation* and Nicholas Belkin about *Personalization of Information Interaction*. Lectures were followed by practical experiments on microblog data from CLEF 2018 Labs under the supervision of the chairs Philippe Mulhem (LIG - Grenoble) and Vincent Labatut (LIA - Avignon).

Social Program Avignon is famous for its medieval architecture and its international theatre festival. The social program of CLEF 2018 set up a Science & Music festival in medieval downtown at Théâtre des Halles[4] and surrounding gardens from Tuesday to Thursday. Music is a very

[2] `https://www.canal-u.tv/`

[3] `http://research.nii.ac.jp/ntcir/`

[4] `http://www.theatredeshalles.com/`

popular hobby among members of the scientific community. Evenings have been a mix of music and participatory science around PlantNet. Tuesday was specially devoted to welcoming students at CLEF, with a theater workshop and a student concert. On Wednesday the focus was on IR scientific societies around the world mixing all CLEF languages in one evening, starting with a tango workshop followed by musical improvisations around new labs joining CLEF for 2019. Finally, science outreach activities were carried out on Thursday, where local musicians and students looking for a good time, were invited to come and meet the participants of the CLEF conference. Dinner tickets for two places close to the theater were offered to all participants. Times and places were randomly distributed to encourage unexpected encounters and discussions.

3 The CLEF Lab Sessions

Ten laboratories and two workshops were selected and ran during CLEF 2018. 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 2018, 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 the CLEF Working Notes [2]. We only provide a brief overview of them here.

CENTRE@CLEF 2018 - CLEF/NTCIR/TREC Reproducibility[5] aims 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 [4].

CheckThat![6] aims to foster the development of technology capable of both spotting and verifying check-worthy claims in political debates in English and Arabic [12].

Dynamic Search for Complex Tasks[7] The lab strives to answer one key question: how can we evaluate, and consequently build, dynamic search algorithms? The 2018 Lab focuses on the development of an evaluation framework, where participants submit "querying agents" that generate queries to be submitted to a static retrieval system. Effective "querying agents" can then simulate users towards developing dynamic search systems [10].

CLEFeHealth[8] targets scenarios which aim to ease patients' and nurses' understanding and access 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

[5] http://www.centre-eval.org/clef2018/

[6] http://alt.qcri.org/clef2018-factcheck/

[7] https://ekanou.github.io/dynamicsearch/

[8] https://sites.google.com/view/clef-ehealth-2018/

in empirical medicine; and, patient-centred information retrieval [15].

ImageCLEF[9] organizes three main tasks and a pilot task: (i) a caption prediction task that aims at predicting the caption of a figure from the biomedical literature based only on the figure image; (ii) a tuberculosis task that aims at detecting the tuberculosis type, severity and drug resistance from CT (Computed Tomography) volumes of the lung; (iii) a lifelog task (videos, images and other sources) about daily activities understanding and moment retrieval, and (iv) a pilot task on visual question answering where systems are asked to answer medical questions [7].

LifeCLEF[10] 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 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 [8].

MC2[11] mainly focuses on developing processing methods and resources to mine the social media (SM) sphere surrounding cultural events such as festivals, music, books, movies and museums. Following previous editions (CMC 2016 and MC2 2017), the 2018 edition focused on argumentative mining and multilingual cross SM search [6].

PAN[12] is a networking initiative for the digital text forensics, where researchers and practitioners study technologies that analyze texts with regard to originality, authorship, and trustworthiness. PAN offers three tasks at CLEF 2018 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: author identification; author profiling; and, author obfuscation [14].

Early risk prediction on the Internet (eRisk)[13] 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. Our 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. eRisk 2018 had two campaign-style tasks: early detection of signs of depression and early detection of signs of anorexia [11].

Personalised Information Retrieval at CLEF (PIR-CLEF)[14] provides a framework for the evaluation of Personalised Information Retrieval (PIR). Current approaches to the evaluation of PIR are user-centric, mostly based on user studies, i.e., they rely on experiments that involve real users in a supervised environment. PIR-CLEF aims to develop and demonstrate a methodology for the evaluation of personalised search that enables repeatable experiments. The main aim is

[9] http://www.imageclef.org/2018
[10] http://www.lifeclef.org/
[11] https://mc2.talne.eu/
[12] http://pan.webis.de/
[13] http://early.irlab.org/
[14] http://www.ir.disco.unimib.it/pir-clef2018/

to enable research groups working on PIR to both experiment with and provide feedback on the proposed PIR evaluation methodology [13].

More information on the CLEF 2018 conference, the CLEF initiative and the CLEF Association is provided on the Web:

- CLEF 2018: `http://clef2018.clef-initiative.eu/`
- CLEF initiative: `http://www.clef-initiative.eu/`
- CLEF Association: `http://www.clef-initiative.eu/association`

4 CLEF 2019 and Beyond

CLEF 2019 will be hosted by the University of Lugano, Switzerland, 9-12 September 2019, and it will mark the 20th anniversary of CLEF.

The call for papers for the CLEF 2019 Conference[15] has been released and the expected deadlines are:

- Submission of abstracts / intentions: 3 May 2019
- Submission of Long and Short Papers: 10 May 2019
- Notification of Acceptance: 7 June 2019
- Camera Ready Copy due: 21 June 2019

As far as labs are concerned, CLEF 2019 will run nine evaluation activities: eight will be a continuation of the labs running during CLEF 2018 and one will be a new pilot lab. For the first time, we introduced a mentorship program to support the preparation of lab proposals for newcomers to CLEF.

The continued activities are:

- **CENTRE@CLEF 2019 - CLEF/NTCIR/TREC REproducibiliry**[16]: the goal of CENTRE@CLEF 2018 is to run a joint CLEF/NTCIR/TREC task 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; 3) to explore the generalizability of experimental findings. *Task 1 - Replicability*: the task will focus on the replicability of selected methods on the same experimental collections. *Task 2 - Reproducibility*: the task will focus on the reproducibility of selected methods on different experimental collections. *Task 3 - Generalizability*: the task will focus on collection performance prediction and the goal is to rank (sub-)collections on the basis of the expected performance over them.

[15] `http://clef2019.clef-initiative.eu/index.php?page=Pages/cfp.html`
[16] `http://www.centre-eval.org/clef2019/`

- **CheckThat! – Automatic Identification and Verification of Political Claims**[17]: aims to foster the development of technology capable of spotting check-worthy claims in English political debates in addition to providing evidence-supported verification of both Arabic and English claims. *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 - Evidence and factuality*: given a sentence-long check-worthy claim and a set of retrieved Web pages that are potentially related to the claim, the goal is for the system to identify which of the Web pages can be useful in assisting a human in fact-checking the claim. Moreover, the system is required to estimate the veracity of the claim based on the given pages."

- **CLEF eHealth**[18]: 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*: the task will explore the automatic assignment of ICD-10 codes to German health-related documents, non-technical summaries (NTSs) of animal experiments. Participants will be challenged with the semantic indexing of NTSs using codes from the German version of the International Classification of Diseases (ICD-10). *Task 2 - Technology 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 - Consumer Health Search*: Participants must retrieve web pages that fulfil a given patient's personalised information need. This needs to fulfil the following criteria: information reliability, quality, and suitability. Spoken queries and automatic speech-to-text transcripts are also provided.

- **eRisk – Early risk prediction on the Internet**[19]: 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 anorexia*: the challenge consists in performing a task on early risk detection of anorexia. The challenge consists of sequentially processing pieces of evidence and detect early traces of anorexia as soon as possible. *Task 2 - self-harm*: the task focuses on self-harm data and it has a similar format as T1 (but we will provide no training data). *Task 3 - depression level estimation*: this is a ranking task designed to estimate the level of depression from a thread of user submissions. For each user, the participants will be given a history of postings and the participants will have to rank users by decreasing level of estimated depression (the gold standard will be based on depression questionnaires associated to the users).

- **ImageCLEF – Multimedia Retrieval in CLEF**[20]: following the ImageCLEF 2003-2018 editions, the ImageCLEF 2019 lab provides an evaluation forum for visual media analysis, indexing, classification/learning, and retrieval in medical, nature, security and lifelogging

[17] `http://alt.qcri.org/clef2019-checkthat/`
[18] `https://sites.google.com/site/clefehealth2019/`
[19] `http://early.irlab.org/`
[20] `https://www.imageclef.org/2019/`

applications. *Task 1 - ImageCLEFlifelog*: the task addresses the problems of lifelogging data retrieval and summarization from real-world multi-modal data. *Task 2 - ImageCLEF-security*: the task addresses the problems of automatically identifying forged content and retrieve hidden information from images. *Task 3 - ImageCLEFmedical*: the task combines several medical image analysis scenarios, e.g., automatically predicting tuberculosis type from 3D chest CT scans and mapping of visual information to textual descriptions, with the objective to foster collaborations. *Task 4 - ImageCLEFcoralReef*: The task addresses the problem of automatically segmenting and labeling a collection of images that can be used in combination to create 3D models of an underwater environment for the monitoring of coral reefs.

- **LifeCLEF – Biodiversity identification and prediction**[21]: aims at boosting research on the identification and prediction of living organisms in order to solve the taxonomic gap and improve our knowledge of biodiversity. *PlantCLEF* aims at evaluating image-based plant identification on 10K species. *BirdCLEF* aims at evaluating bird species detection in audio soundscapes. *GeoLifeCLEF* aims at evaluating location-based prediction of species based on environmental and occurrence data.

- **PAN Lab on Digital Text Forensics and Stylometry**[22]: is a series of scientific events and shared tasks on digital text forensics and stylometry. *Task 1 - Bots and Gender Profiling*: detect if a social media account is operated by a human or a bot. If it is indeed a human, determine the authors' gender. *Task 2 - Celebrity Profiling*: determine the personal demographics of a public figure via style analysis of its social media account. *Task 3 - Cross-domain Authorship Attribution*: attribute the author of a text between different domains and within open-set authorship conditions. *Task 4 - Style Change Detection*: detect if a document was written by a single author, or by multiple ones. "

- **PIR-CLEF – Evaluation of Personalised Information Retrieval**[23]: the core objective of PIR-CLEF 2019 is to develop laboratory-based evaluation methods to enable comparative evaluation of personalization in web search and medical search. *Task 1 – Web Search*: the Web Search task aims to explore personalization in web search sessions based on user profiles and activity with the current and previous search sessions by this user. Task participants will be provided with user profile data and logs of search activity with the objective of improve search effectiveness over that achieved in the logged search sessions. A particular focus in PIR-CLEF 2019 will be on the exploration on methods and metrics for comparative of search sessions. *Task 2 – Medical Search*: Medical search is one of the most common interests of users of search engines. User queries for medical information are often highly ambiguous.The Medical Search task will focus on adaptation and personalization of medical search sessions carried out by volunteer searchers based on defined diverse search scenarios and logged search behaviour. Task participants will be provided with user profiles, queries and search session logs. Similar to the Web Search task, the Medical task will also explore the methods and metrics for comparative evaluation of search sessions.

[21] `http://lifeclef.org/`
[22] `http://pan.webis.de/`
[23] `http://www.ir.disco.unimib.it/pir-clef2019/`

The new pilot lab is:

- **ProtestNews – Extracting Protests from News**[24]: aims to test and improve state-of-the-art generalizable machine learning and natural language processing methods for text classification and information extraction on English news from multiple countries such as India and China for creating comparative databases of contentious politics events (riots, social movements), i.e. the repertoire of contention that can enable large scale comparative social and political science studies. *Task 1 - News article classification as protest vs. non-protest*: given a random news article, to what extent can we predict whether it is reporting a contentious politics event that was happened or is happening? *Task 2 - Event sentence detection*: given a news article that is classified as positive in Task 1, to what extent can we identify the sentence(s) that contain the event information? *Task 3 - Event extraction*: given the event sentence that is identified in Task 2, to what extent can we extract key event information such as place, time, participants, etc.? *Task 4 - Violent protest detection*: given a news article that is classified as positive in Task 1, to what extent can we determine whether it is a violent event? ”

The registration to the labs is open[25] and the overall schedule for the CLEF 2019 evaluation cycle is:

- Registration opens: 5 November 2018
- Registration closes: 26 April 2019
- End Evaluation Cycle: 10 May 2019
- Submission of Participant Papers: 24 May 2019

More information about CLEF 2019 is available at:

`http://clef2019.clef-initiative.eu/`

CLEF 2020 will be hosted by the Centre for Research and Technology Hellas (CERTH) and its Information Technologies Institute (ITI), Greece, in early September 2020.

CLEF 2021 will be hosted by University Politehnica of Bucharest, Romania, in early September 2021.

Finally, bids for hosting CLEF 2022 are now open and will close around May 2019. Proposals can be sent to the CLEF Steering Committee Chair at `chair@clef-initiative.eu`.

[24] `https://emw.ku.edu.tr/clef-protestnews-2019/`
[25] `http://clef2019-labs-registration.dei.unipd.it/`

Acknowledgments

The success of CLEF 2018 would not have been possible without the huge effort of several people and organizations, including the CLEF Association[26], 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: ACM SIGIR[27], Région Sud[28] and Aix-Marseille Université through the Institute of Language, Communication and the Brain[29] and the OpenEdition infrastructure[30].

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] P. Bellot, C. Trabelsi, J. Mothe, F. Murtagh, J.-Y. Nie, L. Soulier, E. SanJuan, L. Cappellato, and N. Ferro, editors. *Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Ninth International Conference of the CLEF Association (CLEF 2018).* Lecture Notes in Computer Science (LNCS) 11018, Springer, Heidelberg, Germany, 2018.

[2] L. Cappellato, N. Ferro, J.-Y. Nie, and L. Soulier, editors. *CLEF 2018 Working Notes.* CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, `http://ceur-ws.org/Vol-2125/`, 2018.

[3] N. Ferro. CLEF 15th Birthday: Past, Present, and Future. *SIGIR Forum*, 48(2):31–55, December 2014.

[4] N. Ferro, M. Maistro, T. Sakai, and I. Soboroff. Overview of CENTRE@CLEF 2018: a First Tale in the Systematic Reproducibility Realm. In Bellot et al. [1], pages 239–246.

[5] 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.

[6] M. Hajjem, J.-V. Cossu, C. Latiri, and E. SanJuan. CLEF MC2 2018 Lab Overview. In Bellot et al. [1], pages 302–308.

[7] B. Ionescu, H. Müller, M. Villegas, A. Garcia Seco de Herrera, C. Eickhoff, V. Andrearczyk, Y. Dicente Cid, V. Liauchuk, V. Kovalev, S. A. Hasan, Y. Ling, O. Farri, J. Liu, M. Lungren,

[26] `http://www.clef-initiative.eu/association`
[27] `http://sigir.org/`
[28] `https://www.maregionsud.fr/`
[29] `http://http://www.blri.fr/`
[30] `https://www.openedition.org/`

D.-T. Dang-Nguyen, L. Piras, M. Riegler, L. Zhou, M. Lux, and C. Gurrin. Overview of ImageCLEF 2018: Challenges, Datasets and Evaluation. In Bellot et al. [1], pages 309–334.

[8] A. Joly, H. Goëau, C. Botella, H. Glotin, P. Bonnet, W.-P. Vellinga, R. Planqué, and H. Müller. Overview of LifeCLEF 2018: A Large-Scale Evaluation of Species Identification and Recommendation Algorithms in the Era of AI. In Bellot et al. [1], pages 247–266.

[9] 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.

[10] E. Kanoulas, L. Azzopardi, and G. Hui Yang. Overview of the CLEF Dynamic Search Evaluation Lab 2018. In Bellot et al. [1], pages 362–371.

[11] D. E. Losada, F. Crestani, and J. Parapar. Overview of eRisk: Early Risk Prediction on the Internet. In Bellot et al. [1], pages 343–361.

[12] P. Nakov, A. Barron-Cedeno, T. Elsayed, R. Suwaileh, L. Marquez, W. Zaghouani, P. Atanasova, S. Kyuchukov, and G. Da San Martino. Overview of the CLEF-2018 CheckThat! Lab on Automatic Identification and Verification of Political Claims. In Bellot et al. [1], pages 372–387.

[13] G. Pasi, G. J. F. Jones, K. Curtis, S. Marrara, C. Sanvitto, D. Ganguly, and P. Sen. Evaluation of Personalised Information Retrieval at CLEF 2018 (PIR-CLEF). In Bellot et al. [1], pages 335–342.

[14] E. Stamatatos, F. Rangel Pardo, M. Tschuggnall, B. Stein, M. Kestemont, P. Rosso, and M. Potthast. Overview of PAN 2018. In Bellot et al. [1], pages 267–285.

[15] H. Suominen, L. Kelly, L. Goeuriot, A. Névéol, L. Ramadier, A. Robert, E. Kanoulas, R. Spijker, L. Azzopardi, D. Li, Jimmy, J. Palotti, and G. Zuccon. Overview of the CLEF eHealth Evaluation Lab 2018. In Bellot et al. [1], pages 286–301.

[16] 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.

[17] 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.