

# What Happened in CLEF... For a While?

Nicola Ferro

Department of Information Engineering, University of Padua, Italy  
ferro@dei.unipd.it

**Abstract.** 2019 marks the 20<sup>th</sup> birthday for CLEF, an evaluation campaign activity which has applied the Cranfield evaluation paradigm to the testing of multilingual and multimodal information access systems in Europe. This paper provides a summary of the motivations which led to the establishment of CLEF, and a description of how it has evolved over the years, the major achievements, and what we see as the next challenges.

## 1 Introduction

Performance measuring is a key to scientific progress. This is particularly true for research concerning complex systems, whether natural or human-built. Multilingual and multimedia information systems are particularly complex: they need to satisfy diverse user needs and support challenging tasks. Their development calls for proper evaluation methodologies to ensure that they meet the expected user requirements and provide the desired effectiveness.

Large-scale worldwide experimental evaluations provide fundamental contributions to the advancement of state-of-the-art techniques through the establishment of common evaluation procedures, the organisation of regular and systematic evaluation cycles, the comparison and benchmarking of proposed approaches, and the spreading of knowledge.

The *Conference and Labs of the Evaluation Forum (CLEF)*<sup>1</sup> is a large-scale *Information Retrieval (IR)* evaluation initiative organised in Europe but involving researchers world-wide. CLEF shares the stage and coordinates with the other major evaluation initiatives in the field, namely: the *Text REtrieval Conference (TREC)*<sup>2</sup> [182], the first large-scale evaluation activity in the field of IR, which began in 1992; the *NII Testbeds and Community for Information access Research (NTCIR)*<sup>3</sup>, which promotes research in information access technologies with a special focus on East Asian languages and English; and the *Forum for Information Retrieval Evaluation (FIRE)*<sup>4</sup>, whose aim is to encourage research in Indian languages by creating a platform similar to CLEF, providing data and a common forum for comparing models and techniques applied to these languages.

---

<sup>1</sup> <http://www.clef-initiative.eu/>

<sup>2</sup> <http://trec.nist.gov/>

<sup>3</sup> <http://research.nii.ac.jp/ntcir/>

<sup>4</sup> <http://fire.irsir.res.in/>

This year marks the 20<sup>th</sup> birthday of CLEF, which began as an independent activity in 2000. The goal of this report is to provide a short overview of what motivated the setting up of CLEF, what has happened in CLEF during these years, and how CLEF has evolved to keep pace with emerging challenges.

I was a bit reluctant to write this paper since Carol Peters, who established CLEF and coordinated it for the first ten years, wrote a series of “What Happened in CLEF” papers [347–352] where she was used to welcome readers and introduce them to the main outcomes and news of a CLEF edition. So, putting myself within the wake of this tradition seemed a bit bold but a 20th anniversary is a so special event that it is worth the risk and the attempt.

The paper is organized as follows: Section 2 describes the beginning and the first period of CLEF, the so-called “CLEF Classic” period; Section 3 introduces the second (and current) period of CLEF, known as the “CLEF Initiative” period; Sections 4 and 5 give an idea of the spread and extension of CLEF activities by providing a short account of the topics addressed in the conference, tracks and labs over the years together with pointers to papers providing more details; Section 6 attempts to provide an assessment of the status of CLEF in the IR community; Section 7 presents the book we have prepared for celebrating the 20th anniversary of CLEF; Section 8 remembers friends and colleagues who have greatly contributed to CLEF but, sadly, are no more with us; Section 9 presents the CLEF Association, the no-profit legal entity committed to sustaining and running CLEF; finally, Section 10 discusses some news in the CLEF eco-system and how to look for future research directions.

## 2 CLEF “Classic”: 2000–2009

The *Cross-Language Evaluation Forum* (CLEF) began as a cross-lingual track at TREC in 1997 [421], moving to an independent activity in 2000 [345], since Europe was felt as a more suitable environment than USA for fully empowering multilinguality.

The underlying motivation for CLEF was the “Grand Challenge” formulated at the *Association for the Advancement of Artificial Intelligence (AAAI)* 1997 Spring Symposium on Cross-Language and Speech Retrieval [189]. The ambitious goal was the development of fully multilingual and multimodal information access systems capable of:

- processing a query in any medium and any language;
- finding relevant information from a multilingual multimedia collection containing documents in any language and form;
- presenting it in the style most likely to be useful to the user.

The main objective of CLEF has thus been to promote research and stimulate development of multilingual and multimodal IR systems for European (and non-European) languages [346], through:

- the creation of an evaluation infrastructure and the organisation of regular evaluation campaigns for system testing;

- the building of a multidisciplinary research community;
- the construction of publicly available test-suites.

CLEF has pursued this objective by attempting to anticipate the emerging needs of the R&D community and to promote the development of multilingual and multimodal systems that fulfil the demands of the AAAI 1997 Grand Challenge. However, while the first three editions of CLEF were dedicated to mono- and multilingual ad-hoc text retrieval, gradually the scope of activity was extended to include other kinds of text retrieval across languages (i.e., not just document retrieval but question answering and geographic IR as well) and on other media (i.e., collections containing images and speech).

During what is jokingly referred to as the “classic” period of CLEF (2000–2009), several important results were achieved: research activities in previously unexplored areas were stimulated, permitting the growth of IR for languages other than English; evaluation methodologies for different types of *Cross Language Information Retrieval (CLIR)* as well as *MultiLingual Information Access (MLIA)* systems, operating in diverse domains, were studied and implemented; a large set of empirical data about multilingual information access from the user perspective was created; quantitative and qualitative evidence with respect to best practices in cross-language system development was collected; reusable test collections for system benchmarking were developed; language resources for a wide range of European languages, some of which had been little studied, were built. CLEF activities have resulted in the creation of a considerable amount of valuable resources, also for under-represented languages, extremely useful for many types of text processing and benchmarking activities in the IR domain. Perhaps, most important, a strong, multidisciplinary, and active research community focussed mainly, but not only, on IR for European languages came into being.

If we had to summarize the major outcome of CLEF in this period with just one sentence, we could safely say that CLEF has made multilingual IR for European languages a reality, with performances as satisfactory as monolingual ones.

### 3 The CLEF Initiative: 2010 Onwards

#### 3.1 Scope

The second period of CLEF started with a clear and compelling question: after a successful decade studying multilinguality for European languages, what were the main unresolved issues currently facing us? To answer this question, CLEF turned to the CLEF community to identify the most pressing challenges and to list the steps to be taken to meet them.

The discussion led to the definition and establishment of the *CLEF Initiative*, whose main mission is to promote research, innovation, and the development of information access systems with an emphasis on multilingual and *multimodal* information with various levels of structure.

In the CLEF Initiative an increased focus is on the *multimodal* aspect, 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. These different modalities should, ideally, be addressed in an integrated way; rather than building vertical search systems for each domain/modality the interaction between the different modalities, languages, and user tasks needs to be exploited to provide comprehensive and aggregated search systems.

The continuity with the first period of CLEF on multilinguality and this increased attention for multimodality has led to the definition of a set of action lines for the CLEF Initiative:

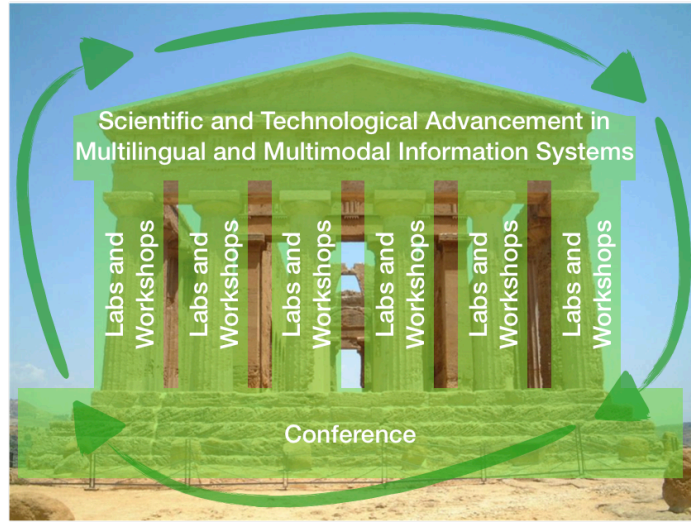
- multilingual and multimodal system testing, tuning and evaluation;
- investigation of the use of unstructured, semi-structured, highly-structured, and semantically enriched data in information access;
- creation of reusable test collections for benchmarking;
- exploration of new evaluation methodologies and innovative ways of using experimental data;
- discussion of results, comparison of approaches, exchange of ideas, and transfer of knowledge.

The new challenges and the new organizational structure, described below, have motivated a change of name for CLEF: from the *Cross-Language Evaluation Forum*, of the “classic” period, to *Conference and Labs of the Evaluation Forum*, which now reflects the widened scope.

### 3.2 Structure

The new challenges for CLEF also called for a renewal of its structure and organization. The annual CLEF meeting is no longer a Workshop, held in conjunction with the European Conference on Digital Libraries (ECDL, now TPD – Theory and Practice of Digital Libraries), but has become an independent event, held over 3.5-4 days and made up of two interrelated activities: the *Conference* and the *Labs*. As shown in Figure 1, the *Conference* and the *Labs* are expected to interact together and mutually reinforce each other, bringing new interest and new expertise into CLEF.

More in detail, the *Conference* is a peer-reviewed conference, open to the IR community as a whole and not just to *Lab* participants, and aims at stimulating discussion on innovative evaluation methodologies, fostering a deeper analysis and understanding of experimental results, and promoting multilingual and multimodal information access at large. The *Labs* are the core of the evaluation activities; they are selected on the basis of topical relevance, novelty, potential research impact, the existence of clear real-world use cases, a likely number of participants, and the experience of the organizing consortium. We allow also for a special case of pilot lab activities, called *Workshops*, whose goal is to explore new and “risky” evaluation activities, which are not ready yet for



**Fig. 1.** Structure of the CLEF Initiative.

being shipped as full-fledged *Labs* and benefit from an incubation and discussion period to better tune them.

The *Conference* and the *Labs* originate two streams of peer-reviewed publications. The *CLEF Proceedings*<sup>5</sup> are published in the Springer Lecture Notes in Computer Science (LNCS) series and contain full and short papers submitted to the *Conference*, condensed overviews of the *Lab* activities, and revised and selected “best of labs” papers from labs in the previous edition of CLEF. The *CLEF Working Notes* are published in the CEUR Workshop Proceedings (CEUR-WS.org)<sup>6</sup> series and contain extended lab overviews and detailed papers from the participants in the lab activities. The peer-review process for the *CLEF Proceedings* is ensured by a Programme Committee, which is established for each CLEF edition; the review process for the *CLEF Working Notes* is ensured by dedicated Programme Committees, which are setup separately for each lab of each CLEF edition.

### 3.3 Organization

In order to favour participation and the introduction of new perspectives, CLEF has introduced a new open-bid process which allows research groups and institutions to bid to host the annual CLEF event and to propose themes. Initially, the bidding process followed a one-year ahead cycle but now, thanks to the interest in and the engagement with CLEF, it follows a three-years ahead cycle, i.e. we are now managing the bids for hosting CLEF 2022.

<sup>5</sup> <https://link.springer.com/conference/clef>

<sup>6</sup> <http://ceur-ws.org/>

While in the CLEF “Classic” period the governing body of CLEF was the *Steering Committee*, which was in charge of the overall coordination of CLEF, of selecting the evaluation activities to be carried out in each edition, and of looking ahead for future research directions, the new participatory approach called for a more articulated organization and for a better separation of concerns.

Each edition of CLEF appoints its own *General Chairs*, *Programme Chairs*, and *Lab Chairs*. The General Chairs are responsible for the overall running of the annual CLEF event, i.e. *Conference* and *Labs* meetings, and serve as the chairs of the organizing committee. The Program Chairs are responsible for planning and implementing the technical program of the *Conference*, and therefore their main responsibility is to ensure that the scientific quality of the *Conference* is at the highest possible level. The Labs Chairs are responsible for selecting, planning, and implementing the focused benchmarking activities, and therefore their main responsibility is to ensure that the scientific and technical quality of the *Labs* is at the highest possible level.

Finally, as before, the *Steering Committee* is in charge of the overall coordination of CLEF: it assists in the appointment of and approves the General Chairs, the Program Chairs and the Labs Chairs for the annual CLEF edition; it devises improvements to the CLEF structure and organization; it manages the bidding process; and, it looks ahead for future research directions to be pursued.

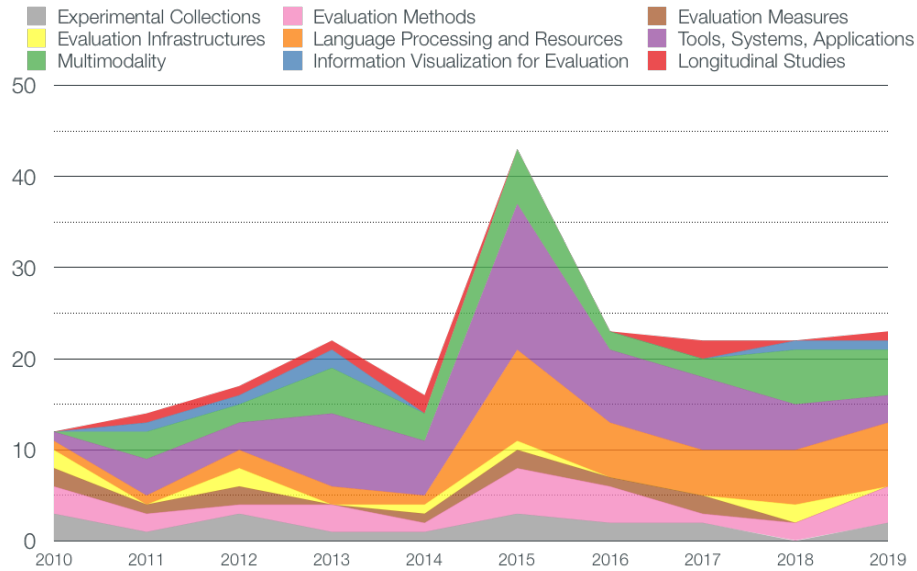
## 4 The Conference

Figure 2 gives an overview of the topics addressed by the CLEF conference over the years, together with the number of submissions for each topic. Figure 2 clearly shows there is a constant stream of papers in the two core areas of CLEF, namely *evaluation* – broken down into “Experimental Collections”, “Evaluation Methods”, “Evaluation Measures”, and “Evaluation Infrastructures” – and *multilinguality and multimodality* – broken down into “Language Processing and Resources”, “Tools, Systems, Applications”, and “Multimodality”. Moreover, we also have a third focus on less mainstream topics – broken down into “Information Visualization for Evaluation” and “Longitudinal Studies”.

We briefly summarize below these topics with pointers to the main references:

**Experimental Collections** explored different issues concerning experimental collections such as: the creation of collections for Persian and Arabic languages; resource-effective creation of pseudo-test collections for specialised tasks; log-based experimental collections; collections for specific domains, e.g. question answering, plagiarism detection, social image tagging; gamification for relevance judgments; early risk detection, such as depression prediction; collections of query features; social media cross-domain corpora [37, 44, 45, 145, 147, 155, 186, 192, 222, 254, 275, 290, 298, 312, 377, 420, 461, 472];

**Evaluation Methods** studied core problems related to evaluation methodologies and proposed new methods, such as: the reliability of relevance assessments; living labs for product search tasks; evaluation of information extraction and entity profiles; semantic-oriented evaluation of machine trans-



**Fig. 2.** Topics addressed by the CLEF conference over the years and number of submissions for each topic.

lation and summarization; search snippet evaluation and query simulators; news recommendation; teaching; study of long tails in relevance judgments; crowdsourcing; definition of transactional tasks; component-based evaluation; methodologies for authorship verification; accounting for bias; evaluation of user models; creation of ground-truth in text classification and question answering; impact of gold standards on evaluation [31, 34, 43, 91, 118, 133, 176, 191, 218, 221, 236, 266, 295, 297, 328–330, 399, 416, 418, 419, 426, 471, 480, 487, 490];

**Evaluation Measures** dealt with the analysis of the features of the evaluation measures and the proposal of new measures such as: formal properties of measures for document filtering; robustness of metrics for patent retrieval; problems with ties in evaluation measures; effort-based measures and measures for speech retrieval; extension of measures to graded relevance; click models; text interestingness and diversity; measures for real-life categorisation and hierarchical clustering [19, 22, 32, 66, 122, 146, 171, 225, 280, 300, 331];

**Evaluation Infrastructures** investigated how to design and develop shared infrastructures to support different aspects of IR evaluation such as: automating component-based evaluation; managing and providing access to the experimental outcomes and the related literature; using cloud-base approaches to offer evaluation services in specialised domains; developing proper ontologies to describe the experimental results; and exploiting map-reduce techniques for effective IR evaluation; frameworks for question-answering; tools for replicability and reproducibility [7, 39, 110, 179, 180, 185, 269];

**Language Processing and Resources** continued the CLEF interest in multilinguality by dealing with tools, algorithm, and resources for multiple languages such as: lemmatizers, decomposers and normalizers for underrepresented resources using statistical approaches; statistical stemmers; named entity extraction, linking and clustering in cross-lingual settings; exploitation of multiple translation resources; language-independent generation of document snippets; language variety identification; gender identification; text alignment; Web genre identification; sentiment analysis and opinion mining; personality and author profiling; mixed-code script analyzers; text clustering; language and terminology analysis, also for query suggestion; microblog contextualization; early depression detection; personality recognition; stance detection in social media; readability; fact checking [1, 25, 33, 50, 52, 65, 74, 79, 89, 90, 93, 95, 96, 141, 149, 153, 217, 248, 253, 257, 268, 270, 274, 294, 327, 336, 391, 394, 395, 408, 410, 412, 415, 427, 440, 449, 450, 455, 477, 482, 491];

**Tools, Systems, and Applications** covered the design and development of various kinds of algorithms, systems, and applications focused on multilinguality and specialised domains such as: semantic discovery of resources in cloud-based systems; Arabic question answering; cross-language similarity search using thesauri; automatic annotation of bibliographic references; exploitation of visual context in multimedia translation; sub-topic mining in Web documents and query interpretation; exploiting relevance feedback for building tag-clouds in image search; query expansion for image retrieval; transcript-based video retrieval; *Peer-To-Peer (P2P)* information retrieval; event detection in microblogs; medical information retrieval; citation for scientific publication; news recommendation; image decomposition and captioning; ranking products in e-commerce; conversational search; mathematical retrieval; systematic reviews; data fusion [2, 3, 11–13, 29, 38, 47, 51, 76, 88, 92, 98, 104, 114, 116, 119, 135, 148, 150, 156, 157, 164, 170, 173, 177, 183, 197, 229, 235, 240, 245, 247, 255, 256, 263, 264, 267, 273, 292, 293, 296, 302, 323, 382, 396, 398, 402, 411, 422, 425, 428, 429, 432, 446, 448, 475, 476, 481, 483–485, 487, 489];

**Multimodality** explored multimodality in the sense described in Section 3 above, i.e. the aggregation and integration of information in multiple languages, media, and coming from different domains, such as: semantic annotation and question answering in the biomedical domain; selecting success criteria in an academic library catalogue; finding similar content in different scenarios on the Web; interactive information retrieval and formative evaluation for medical professionals; microblog summarization, disambiguation and expansion; multimodal music tagging; multi-faceted IR in multimodal domains; ranking in faceted search; domain adaptation; cross-domain vertical search; prediction of venues in social media; query expansion for speech retrieval; neural networks for medical image classification; vocal assistant-mediated search; lifelogging; voice question answering [21, 46, 63, 64, 78, 80, 82, 100, 111, 174, 175, 178, 184, 211, 219, 226, 231, 232, 241, 258, 262, 271, 279, 288, 324, 325, 406, 424, 434, 437, 438, 447, 488];

**Information Visualization for Evaluation** opened up a brand new area concerned with exploiting information visualization and visual analytics tech-

niques not only for presenting the results of a search system but also for improving interaction with and exploration of experimental outcomes such as exploiting visual analytics for failure analysis; comparing the relative performances of IR systems; and visualization for sentiment analysis; visualization for patterns; data analytics and visualization for system settings [24, 99, 105, 252, 409, 465];

**Longitudinal Studies** conducted various kinds of medium and long term analyses such as: the scholarly impact of evaluation initiatives; lessons learned in running evaluation activities and in specific domains; performance trends over the years for multilingual information access; and, component-level analysis across different system configurations [112, 113, 123, 132, 291, 316, 456, 459, 479].

## 5 Tracks and Labs

Figure 3 provides an overview of the tracks and labs offered by CLEF over the years; these are briefly summarized below together with some pointers to relevant literature.

**Multilingual Text Retrieval (Ad-hoc, 2000–2009)** focused on multilingual information retrieval on news corpora, offering monolingual, bilingual and multilingual tasks, and developed a huge collection in 14 European languages [4, 5, 56–60, 106–108, 128];

**Domain Specific Cross-Language IR (DS, 2000–2008)** dealt with multilingual information retrieval on structured scientific data from the social sciences domain [57–59, 237–239, 369, 370, 439];

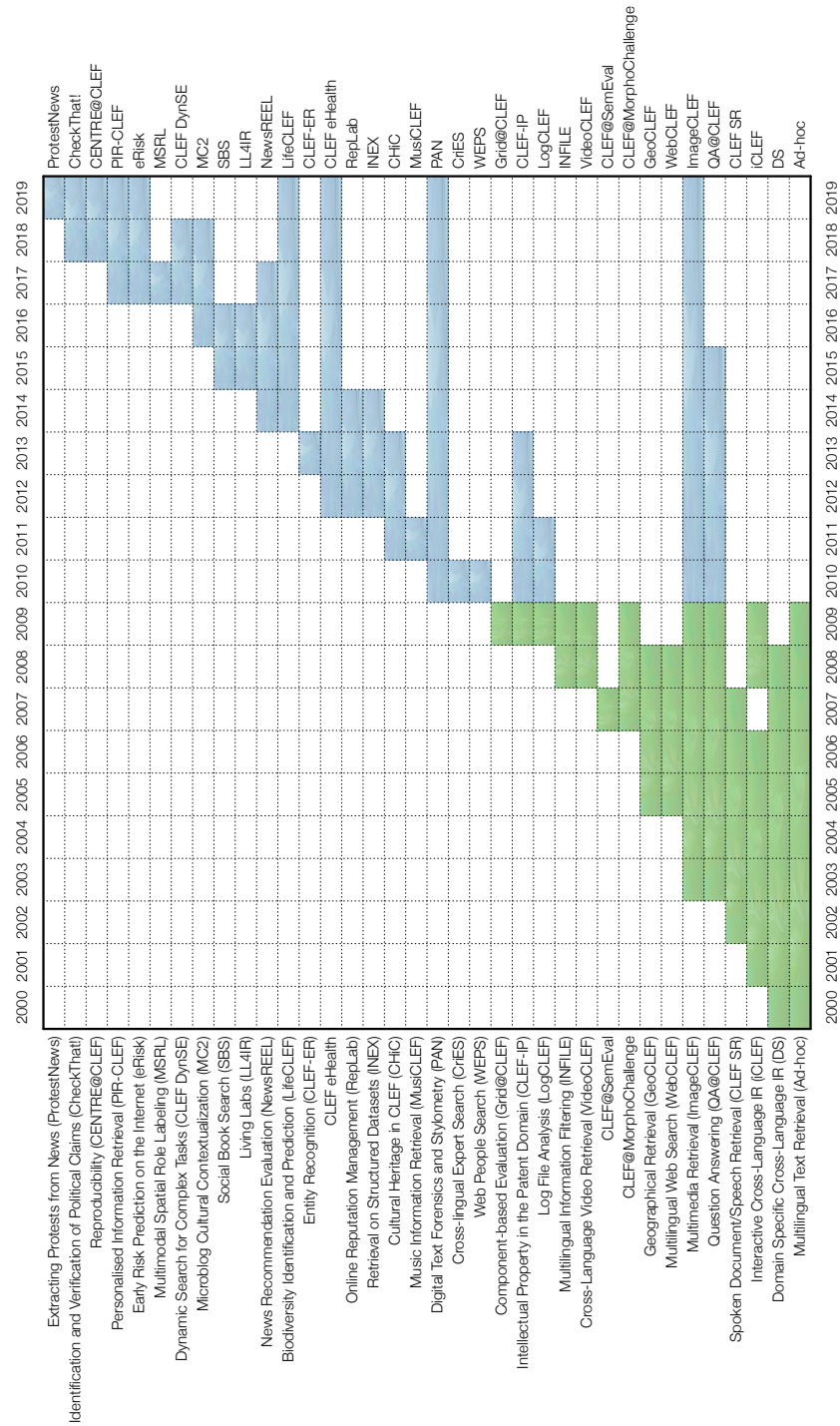
**Interactive Cross-Language IR (iCLEF, 2001–2006, 2008–2009)** explored different aspects of interactive information retrieval on multilingual and multimedia collections, also using gamification techniques [165–169, 223, 320, 321];

**Spoken Document/Speech Retrieval (CLEF SR, 2002–2007)** investigated speech retrieval and spoken document retrieval in a monolingual and bilingual setting on automatic speech recognition transcripts [120, 121, 208, 322, 335, 478];

**Question Answering (QA@CLEF, 2003–2015)** examined several aspects of question answering in a multilingual setting on document collections ranging from news, legal documents, medical documents, linked data [81, 140, 154, 281–283, 299, 337–341, 343, 344, 401, 414, 462–464];

**Multimedia Retrieval (ImageCLEF, 2003–2019)** studied the cross-language annotation and retrieval of images to support the advancement of the field of visual media analysis, indexing, classification, and retrieval [28, 72, 73, 83–86, 101–103, 158, 159, 172, 194–196, 212, 265, 289, 304, 305, 307–309, 311, 317–319, 381, 392, 393, 451, 453, 457, 458, 460, 466–468, 486];

**Multilingual Web Search (WebCLEF, 2005–2008)** addressed multilingual Web search, exploring different faces of navigational queries and known-item search [35, 198, 199, 430];



**Fig. 3.** Labs offered by CLEF over the years (CLEF “Classic” period in green; the CLEF Initiative period in blue).

**Geographical Retrieval (GeoCLEF, 2005–2008)** evaluated cross-language *Geographic Information Retrieval (GIR)* against search tasks involving both spatial and multilingual aspects [151, 152, 285, 287];

**CLEF@SemEval (2007)** explored the impact of *Word Sense Disambiguation (WSD)* on multilingual information retrieval [6]; it continued as a sub-task of the Ad Hoc lab in 2008 and 2009;

**CLEF@MorphoChallenge (2007–2009)** assessed unsupervised morpheme analysis algorithms using information retrieval experiments with the goal of designing statistical machine learning algorithms that discover which morphemes make up words [249–251];

**Cross-Language Video Retrieval (VideoCLEF, 2008–2009)** aimed at developing and evaluating tasks related to the analysis of and access to multilingual and multimedia content with a special focus on video retrieval [260, 261]; it went on to become the *MediaEval Benchmarking Initiative for Multimedia Evaluation*<sup>7</sup> successful series, dedicated to evaluating new algorithms for multimedia access and retrieval;

**Multilingual Information Filtering (INFILE, 2008–2009)** experimented with cross-language adaptive filtering systems on news corpora [48, 49];

**Log File Analysis (LogCLEF, 2009–2011)** investigated the analysis and classification of queries in order to understand search behavior in multilingual contexts and ultimately to improve search systems by offering openly-accessible query logs from search engines and digital libraries [109, 284, 286];

**Intellectual Property in the Patent Domain (CLEF-IP, 2009–2013)** focused on various aspects of patent search and intellectual property search in a multilingual set using the MAREC collection of patents, gathered from the European Patent Office [375, 378–380, 400];

**Component-based Evaluation (Grid@CLEF, 2009)** piloted component-based evaluation by allowing participants to exchange the intermediate state of their systems in order to asynchronously compose components coming from different systems and experiment with a larger grid of possibilities [126];

**Web People Search (WEPS, 2010)** focused on person name ambiguity and person attribute extraction on Web pages and on online reputation management for organizations [17, 30]; the activity continued in the RepLab lab;

**Cross-lingual Expert Search (CriES, 2010)** was run as a brainstorming workshop and addressed the problem of multi-lingual expert search in social media environments [433];

**Digital Text Forensics and Stylometry (PAN, 2010–2019)** studied plagiarism, authorship attribution, social software misuse, different types of profiling [20, 26, 97, 163, 193, 210, 383–386, 388–390, 404, 435, 436];

**Music Information Retrieval (MusiCLEF, 2011)** was run as a brainstorming workshop to aid the development of novel methodologies for both content-based and contextual-based (e.g. tags, comments, reviews, etc.) access and retrieval of music [326]; this activity has continued as part of MediaEval;

---

<sup>7</sup> <http://www.multimediaeval.org/>

**Cultural Heritage in CLEF (CHiC, 2011–2013)** promoted systematic and large-scale evaluation of digital libraries and, more in general, cultural heritage information access systems, using the huge Europeana dataset, aggregating information from libraries, museums, and archives [144, 371, 372];

**Retrieval on Structured Datasets (INEX, 2012–2014)** was a stand-alone initiative pioneering structured and XML retrieval from 2002<sup>8</sup>; it joined forces with CLEF in 2012 to further promote the evaluation of focused retrieval by providing large test collections of structured documents [40, 41, 77, 244, 413, 454, 474];

**Online Reputation Management (RepLab, 2012–2014)** has been a competitive evaluation exercise for online reputation management systems; the lab focused on the task of monitoring the reputation of entities (companies, organizations, celebrities) on Twitter [15, 16, 18];

**CLEF eHealth (2012–2019)** focused on *Natural Language Processing (NLP)* and IR for clinical care, such as annotation of entities in a set of narrative clinical reports or retrieval of web pages based on queries generated when reading the clinical reports [160, 161, 227, 228, 441, 443–445];

**Entity Recognition (CLEF-ER, 2013)** was a brainstorming workshop on the multilingual annotation of named entities and terminology resource acquisition with a focus on entity recognition in biomedical text, in different languages and on a large scale [397];

**Biodiversity Identification and Prediction (LifeCLEF, 2014–2019)** aimed at evaluating multimedia analysis and retrieval techniques on biodiversity data for species identification, namely images for plants, audio for birds, and video for fishes [200–202, 204–206];

**News Recommendation Evaluation (NewsREEL, 2014–2017)** focused on evaluation of news recommender systems in real-time by offering access to the APIs of a commercial system [188, 233, 234, 272];

**Living Labs (LL4IR, 2015–2016)** dealt with evaluation of ranking systems in a live setting with real users in their natural task environments, acting as a proxy between commercial organizations (live environments) and lab participants (experimental systems) [423];

**Social Book Search (SBS, 2015–2016)** investigated techniques to support users in complex book search tasks that involve more than just a query and results list [242, 243].

**Microblog Cultural Contextualization (MC2, 2016–2017)** investigated techniques to support users in complex book search tasks that involve more than just a query and results list [117, 162].

**Dynamic Search for Complex Tasks (CLEF DynSE, 2017–2018)** promoted the development of both algorithms which interact dynamically with user (or other algorithms) towards solving a task and of evaluation methodologies to quantify their effectiveness [214, 215].

**Multimodal Spatial Role Labeling (MSRL, 2017)** explored the extraction of spatial information from two information resources that is image

---

<sup>8</sup> <https://inex.mmci.uni-saarland.de/>

and text, which is important in 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 [246].

**Early Risk Prediction on the Internet (eRisk, 2017–2019)** explored the evaluation methodology, effectiveness metrics and practical applications (particularly those related to health and safety) of early risk detection on the Internet [276–278].

**Personalised Information Retrieval (PIR-CLEF, 2017–2019)** provided a framework for evaluation of *Personalized Information Retrieval (PIR)* by developing a methodology for evaluation PIR which enables repeatable experiments to enable the detailed exploration of personal models and their exploitation in IR [332–334].

**Reproducibility (CENTRE@CLEF, 2018–2019)** run a joint task across CLEF, NTCIR, and TREC on challenging participants to reproduce best results of the most interesting systems submitted in previous editions of CLEF/NTCIR/TREC and to contribute back to the community the additional components and resources developed to reproduce the results [125, 127].

**Identification and Verification of Political Claims (CheckThat!, 2018–2019)** aimed to foster the development of technology capable of spotting check-worthy claims in English political debates in addition to providing evidence-supported verification of Arabic claims [115, 313].

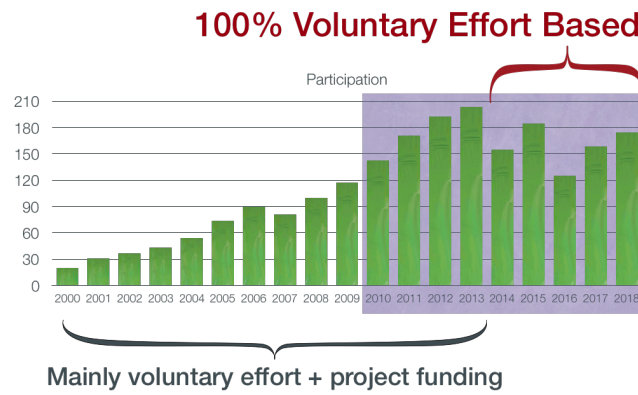
**Extracting Protests from News (ProtestNews, 2019)** aimed to test and improve state-of-the-art generalizable machine learning and natural language processing methods for text classification and information extraction on English news from multiple countries such as India and China for creating comparative 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 [190].

## 6 Trends

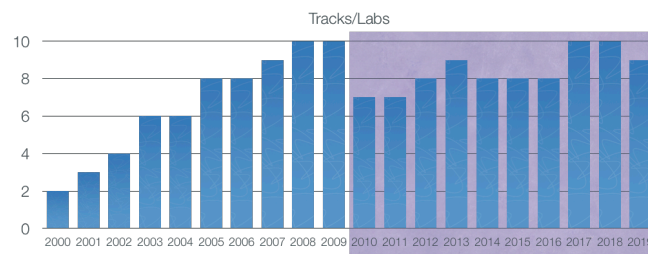
We present here some data on CLEF; the aim is to attempt an informal assessment of its impact on the research community.

Figure 4 shows the participation in CLEF over the years. A positive growth trend is exhibited, a possible consequence of the capacity of CLEF to renew itself and to attract new communities and expertise in addition to core information retrieval activities. In particular, since 2014, CLEF is no more backed by any direct project funding and runs on a completely voluntary-effort basis, still keeping levels of participation comparable to the previous ones.

Figure 5 shows the number of Labs offered by CLEF over the years. It can be noted how the new mechanism introduced for selecting labs is proving effective in restricting the number of Labs run annually, with an average of about 8 Labs per year which allows CLEF to continue successful activities for more than one



**Fig. 4.** Participation in CLEF over the years (CLEF “Classic” period un-shaded; CLEF Initiative period shaded).



**Fig. 5.** Number of labs offered by CLEF over the years (CLEF “Classic” period un-shaded; the CLEF Initiative period shaded).



**Fig. 6.** Number of papers submitted and accepted in the CLEF conference over the years (CLEF “Classic” period un-shaded; the CLEF Initiative period shaded).

cycle, typically three years, but also to introduce new activities every year. Also note that we put a cap on a maximum of 10 labs per edition, in order to avoid dispersion into too many activities.

Figure 6 shows the number of papers submitted and accepted in the CLEF Conference over the years. We see that the number of accepted papers has changed slightly over the years, almost stabilizing in the last two years, while the number of submitted papers has grown, allowing us to increase the selectivity and quality of the Conference.

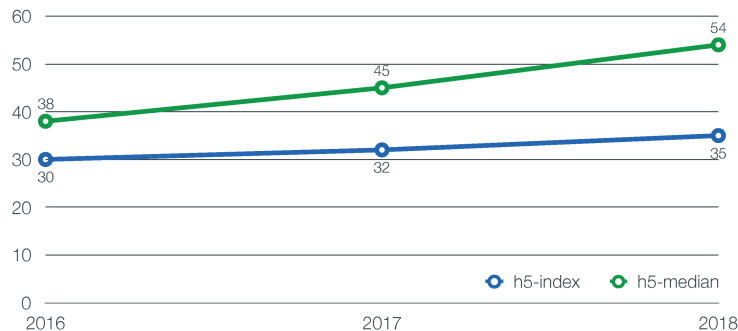
The Conference part of CLEF still needs to be improved and strengthened. The challenge is to define its scope clearly so as to guarantee high quality but to avoid useless overlap with both the major venues in the field, like SIGIR, ECIR and CIKM, and also the fast growing ones, like ICTIR. However, a problem we are still facing is related to communication: CLEF is still mostly associated with its core evaluation activities and therefore, when information is circulated about the conference, it is often viewed as just concerning the evaluation labs even though it actually represents a wider opportunity.

Assessing the impact of an evaluation activity is a very demanding task and it can be done from multiple points of view, e.g. economic impact, industrial impact, scholarly impact, and so on.

In 2010, TREC conducted a deep study on its economic impact [405]. One of goals of CLEF has been to impact not only academia but also industrial research and society in a broader sense. Indeed, IR research can never be considered only at the theoretical level, clearly the overriding factors are the requirements of society at large. An important step in this direction, which began in “CLEF Classic” with ImageCLEF medical retrieval experiments but has certainly been increasingly reinforced in the “CLEF Initiative”, is the involvement of real world user communities. Thus, just to cite a few examples, we have seen collaborations with the intellectual property and patent search domain in CLEF-IP, with health specialists in E-Health, with news portals in the NewsREEL project, until the very recent developments for early risk detection in social media as well as fact checking and trustworthiness.

When it comes to the scientific and scholarly impact, we enter the realm of bibliometrics: *TREC Video Retrieval Evaluation (TRECVID)* conducted a study on its scholarly impact [452] and some steps in this direction have been performed for CLEF as well [23, 456, 459]. However, analysing the impact of evaluation activities on system performances longitudinally over the years is still a research challenge, even if some attempts have been made for both TREC [27, 230] and CLEF [132, 134].

Such rigorous studies are beyond the scope of the present report, here we concentrate on identifying rough indicators with respect to the maturity and liveliness of the scientific production originated by CLEF. Therefore, as proxy for a more rigorous scholarly impact study, we can look at some statistics gathered from Google Scholar.



**Fig. 7.** Google Scholar Metrics for “Cross-Language Evaluation Forum”: h5-index (blue) and h5-median (green).

The query “CLEF evaluation” returns 53,300 hits which is comparable to the 53,000 hits of the query “TREC evaluation” (run in June 2019), suggesting a lively research community.

Figure 7 shows the trends of the h5-index, i.e. the h-index of the papers published in the last 5 years, and the h5-median, i.e. the the median number of citations for the articles that make up the h5-index, from 2016 to 2018 taken from Google Scholar Metrics. We can observe a steady increase trend for both indicators, suggesting a positive scholarly impact for the research outcomes of the CLEF community.

As far as maturity is concerned, an indicator might be found in publications critically analysing, systematizing, and digesting the achievements, outcomes and experience; this has been done both for TREC [181, 182, 469] and CLEF [62, 303, 353], with a special publication for the 20th anniversary of CLEF, as discussed in the next section.

## 7 The CLEF Book

To celebrate the 20th anniversary of CLEF, we have prepared a book [130] which accounts for the evolution of CLEF over the years, its contribution to the advancement of research in multilingual and multimodal information access, and its perspectives for the future.

In order to do this, the volume is divided into six parts. The first three chapters in Part I “Experimental Evaluation and CLEF” explain what is intended by experimental evaluation and the underlying theory [470], describing how this has been interpreted in CLEF and in other internationally recognized evaluation initiatives [407]. In addition, the introductory chapter illustrates the activity and results of CLEF over the years in some detail [129]. Part II “Evaluation Infrastructure” presents research architectures and infrastructures that have been developed to manage experimental data [8] and to provide evaluation services in CLEF and elsewhere [306, 387].

Parts III, IV and V represent the core of the volume, consisting of a series of chapters presenting some of the most significant evaluation activities in CLEF, ranging from the early multilingual text processing exercises to the later, more sophisticated experiments on multimodal collections in diverse genre and media. In all cases, the focus has not only been on describing “what has been achieved” but most of all on “what has been learnt”. Part III “Multilingual and Multimedia Information Retrieval” focuses on multilinguality [417] and the impact of languages on information access [224]; it then addresses multimodality from the perspective of both images [87,310,374,473] and sound and vision [207]. Part IV “Retrieval in New Domains” deals with the medical domain [442], the intellectual property and patent domain [376], the biodiversity domain [203], and the structured data and semantic search domains [213]. Part V “Beyond Retrieval” covers information access tasks other than pure retrieval, namely question answering [342], digital text forensics [403], online reputation management [10], and continuous evaluation and living labs [187].

The final Part VI “Impact and Future Challenges” is dedicated to examining the impact CLEF has had on the research world and to discussing current and future challenges, both academic and industrial. We conduct a proper scholarly impact analysis [259] and we discuss open issues and areas for future development, such as reproducibility and validity [142] and *Visual Analytics* (VA) for experimental evaluation [131]. In particular, the concluding chapter discusses the relevance of IR benchmarking in an industrial setting [220].

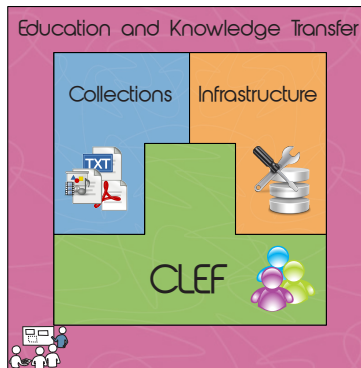
## 8 In Memoriam

This 20th anniversary is not only an opportunity to celebrate the achievements of CLEF but also, sadly, an occasion for remembering friends and colleagues who greatly contributed to what CLEF is today and inspired all of us.

In 2013 Emanuele Pianta, Fondazione Bruno Kessler (FBK) and Centre for the Evaluation of Language and Communication Technologies (CELCT), Trento, Italy, passed away in a car accident close to home. Emanuele, member of the CLEF Steering Committee, greatly contributed to the transition from “CLEF Classic” to the “CLEF Initiative”, putting a lot of effort, ideas, and enthusiasm in this new project.

In 2017 Ray Larson, UC Berkeley, USA, passed away following a struggle with cancer. Ray was a dedicated scholar and innovator for CLEF since the early days, bringing in new activities and new research horizons, such as the geographic information retrieval tasks.

Very recently, in 2019, Séamus Lawless, Trinity College Dublin, Ireland, fell during a descent of Mount Everest, after having achieved his dream of reaching the peak. Séamus, member of the CLEF Steering Committee, greatly contributed to CLEF with his research talent and generosity of spirit by opening new research collaborations with the adaptivity and personalisation communities.



**Fig. 8.** Pillar activities of the CLEF Association.

## 9 The CLEF Association

The CLEF Association<sup>9</sup> is an independent no-profit legal entity, established in October 2013 as a result of activity of the PROMISE<sup>10</sup> Network of Excellence which backed CLEF from 2010 to 2013.

The CLEF Association has scientific, cultural and educational objectives and operates in the field of information access systems and their evaluation. Its mission is:

- to promote access to information and use evaluation;
- to foster critical thinking about advancing information access and use from a technical, economic and societal perspective.

Within these two areas of interest, the CLEF Association aims at a better understanding of the use and access to information and how to improve this. The two areas of interest stated in the the above mission translate into the following objectives:

- *clustering stakeholders* with multidisciplinary competences and different needs, including academia, industry, education and other societal institutions;
- *facilitating medium/long-term research* in information access and use and its evaluation;
- increasing, *transferring* and applying *expertise*.

As Figure 8 shows, the CLEF Association pursues its mission and objectives via four pillar activities:

- *CLEF*: sustains and promotes the popular CLEF evaluation series as well as providing support for its coordination, organisation, and running;

<sup>9</sup> <http://www.clef-initiative.eu/association>

<sup>10</sup> <http://www.promise-noe.eu/>

- *Collections and Experimental Data*: fosters the adoption and exploitation of large-scale shared experimental collections, makes them available under appropriate conditions and trusted channels, and shares experimental results and scientific data for comparison with state-of-the-art and for reuse;
- *Infrastructure*: supports the adoption and deployment of software and hardware infrastructures which facilitate the experimental evaluation process, the sharing of experimental collections and results, and interaction with and understanding of experimental data;
- *Education and knowledge transfer*: organises educational events, such as summer schools, and knowledge transfer activities, such as workshops, aimed not only at spreading know-how about information access and use but also at raising awareness and stimulating alternative viewpoints about the technical, economic, and societal implications.

In its initial phase, the CLEF Association has been focused mainly on the first pillar, i.e. ensuring the continuity and self-sustainability of CLEF. CLEF 2014 was the first edition of CLEF not supported by a main European project, but run on a totally volunteer basis with only the support of the CLEF association membership fees paid by its multidisciplinary research community.

Moreover, the CLEF association plans to continue the already initiated activities for promoting and developing shared infrastructures and formats in IR evaluation [431] by also joining forces with relevant stakeholders in the fields as well as stimulating and contributing critical thinking about large-scale evaluation initiative and IR evaluation more in general.

## 10 Beyond CLEF

As it should have emerged across this paper, we constantly strive for improving CLEF and opening up new perspectives and collaborations.

In this respect, CLEF 2019 will host for the first time an Industry Day, jointly organized with the Swiss Alliance for Data-Intensive Services. The goal is to further open CLEF to a wider, industrial community through demo sessions, panels and special keynotes where the very best and most pertinent work of CLEF participants will be made more publicly accessible.

An aspect of CLEF of which we are particularly proud is the consolidation of a strong community of European researchers in the multidisciplinary context of IR. This year, for the first time, the *European Conference for Information Retrieval (ECIR)* and CLEF have joined forces: ECIR 2019 hosted a session dedicated to CLEF Labs where lab organizers present the major outcomes of their Labs and plans for ongoing activities, followed by a poster session in order to favour discussion during the conference. This is reflected in the ECIR 2019 proceedings, where CLEF Lab activities and results are reported as short papers. Similar plans are already in place for CLEF 2020 and ECIR 2020. The goal is not only to engage the ECIR community in CLEF activities, but also to disseminate the research results achieved during CLEF evaluation cycles at ECIR. This collaboration will of course strengthen European IR research even more.

However, this European community should not be seen in isolation. CLEF is part of a global community; we have always maintained close links with our peer initiatives in the Americas and Asia. There is a strong bond connecting TREC, NTCIR, CLEF and FIRE, and a continual, mutually beneficial exchange of ideas, experiences and results.

On a closing note, there have been very recent brainstorming workshops [14, 124] to devise a future research agenda in IR for the coming years and they also touched upon evaluation topics. However, after 20 years of activities, the time may be ripe also for CLEF to organize a brainstorming event to envision the next decade of evaluation challenges.

## **Support for the Central Coordination of CLEF**

CLEF 2000 and 2001 were supported by the European Commission under the Information Society Technologies programme and within the framework of the DELOS Network of Excellence for Digital Libraries (contract no. IST-1999-12262).

CLEF 2002 and 2003 were funded as an independent project (contract no. IST-2000-31002) under the 5th Framework Programme of the European Commission.

CLEF 2004 to 2007 were sponsored by the DELOS Network of Excellence for Digital Libraries (contract no. G038-507618) under the 6th Framework Programme of the European Commission.

Under the 7th Framework Programme of the European Commission, CLEF 2008 and 2009 were supported by TrebleCLEF Coordination Action (contract no. 215231) and CLEF 2010 to 2013 were funded by the PROMISE Network of Excellence (contract no. 258191).

CLEF 2011 to 2014 also received support from the ELIAS network (contract no. 09-RNP-085) of the European Science Foundation (ESF).

Over the years CLEF has also attracted industrial sponsorship: from 2010 onwards, CLEF has received the support of Google, Microsoft, Yandex, Xerox, Celi as well as publishers in the field such as Springer and Now Publishers.

Note that, beyond receiving the support of all the volunteer work of its community, CLEF tracks and labs have often received the support of many other projects and organisations; unfortunately, it is impossible to list them all here.

## **Acknowledgements**

CLEF would not be possible without all the effort, enthusiasm, and passion of its community: lab organizers, lab participants, and attendees are the core and the real success of CLEF.

Many friends and colleagues – too many to mention them all but I sincerely thank all of them – have shared with me this journey through CLEF and their work, passion, ideas, expertise and wisdom have shaped what CLEF is today.

However, all of this would have not even been possible without Carol Peters, who established CLEF back in 2000, made it grown over the years with constant care, and put into CLEF her secret ingredient which makes it so special: a very friendly environment where everybody feels to be welcome and comfortable in sharing ideas and contributions. Carol has had the generosity of sharing her experience with me and teaching me a lot about how to run an evaluation initiative and grow a healthy research community: I will never thank her enough for this.

## References

1. Abnar, S., Dehghani, M., Shakery, A.: Meta Text Aligner: Text Alignment Based on Predicted Plagiarism Relation. In: Mothe et al. [301], pp. 193–199
2. Adams, J., Bedrick, S.: Automatic Indexing of Journal Abstracts with Latent Semantic Analysis. In: Mothe et al. [301], pp. 200–208
3. Afzal, Z., Akhondi, S.A., van Haagen, H.H.H.B.M., van Mulligen, E.M., Kors, J.A.: Concept Recognition in French Biomedical Text Using Automatic Translation. In: Fuhr et al. [143], pp. 162–173
4. Agirre, E., Di Nunzio, G.M., Ferro, N., Mandl, T., Peters, C.: CLEF 2008: Ad Hoc Track Overview. In: Peters et al. [359], pp. 15–37
5. Agirre, E., Di Nunzio, G.M., Mandl, T., Otegi, A.: CLEF 2009 Ad Hoc Track Overview: Robust-WSD Task. In: Peters et al. [360], pp. 36–49
6. Agirre, E., de Lacalle, O.L., Magnini, B., Otegi, A., Rigau, G., Vossen, P.: SemEval-2007 Task 01: Evaluating WSD on Cross-Language Information Retrieval. In: Peters et al. [366], pp. 908–917
7. Agosti, M., Di Buccio, E., Ferro, N., Masiero, I., Peruzzo, S., Silvello, G.: DIRECTIONS: Design and Specification of an IR Evaluation Infrastructure. In: Catarci et al. [75], pp. 88–99
8. Agosti, M., Di Nunzio, G.M., Ferro, N., Silvello, G.: An Innovative Approach to Data Management and Curation of Experimental Data Generated through IR Test Collections. In: Ferro and Peters [130]
9. Agosti, M., Ferro, N., Peters, C., de Rijke, M., Smeaton, A. (eds.): Multilingual and Multimodal Information Access Evaluation. Proceedings of the International Conference of the Cross-Language Evaluation Forum (CLEF 2010). Lecture Notes in Computer Science (LNCS) 6360, Springer, Heidelberg, Germany (2010)
10. Carrillo-de Albornoz, J., Gonzalo, J., Amigó, E.: RepLab: an Evaluation Campaign for Online Monitoring Systems. In: Ferro and Peters [130]
11. Alfalahi, A., Eriksson, G., Sneider, E.: Shadow Answers as an Intermediary in Email Answer Retrieval. In: Mothe et al. [301], pp. 209–214
12. Alharbi, A., Stevenson, M.: Improving Ranking for Systematic Reviews Using Query Adaptation. In: Crestani et al. [94]
13. Alkhawaldeh, R.S., Jose, J.M.: Experimental Study on Semi-structured Peer-to-Peer Information Retrieval Network. In: Mothe et al. [301], pp. 3–14
14. Allan, J., Arguello, J., Azzopardi, L., Bailey, P., Baldwin, T., Balog, K., Bast, H., Belkin, N., Berberich, K., von Billerbeck, B., Callan, J., Capra, R., Carman, M., Carterette, B., Clarke, C.L.A., Collins-Thompson, K., Craswell, N., Croft, W.B., Culpepper, J.S., Dalton, J., Demartini, G., Diaz, F., Dietz, L., Dumais, S., Eickhoff, C., Ferro, N., Fuhr, N., Geva, S., Hauff, C., Hawking, D., Joho, H., Jones, G.J.F., Kamps, J., Kando, N., Kelly, D., Kim, J., Kiseleva, J., Liu, Y., Lu, X.,

- Mizzaro, S., Moffat, A., Nie, J.Y., Olteanu, A., Ounis, I., Radlinski, F., de Rijke, M., Sanderson, M., Scholer, F., Sitbon, L., Smucker, M.D., Soboroff, I., Spina, D., Suel, T., Thom, J., Thomas, P., Trotman, A., Voorhees, E.M., de Vries, A.P., Yilmaz, E., Zuccon, G.: Research Frontiers in Information Retrieval – Report from the Third Strategic Workshop on Information Retrieval in Lorne (SWIRL 2018). SIGIR Forum 52(1), 34–90 (June 2018)
15. Amigó, E., Carrillo de Albornoz, J., Chugur, I., Corujo, A., Gonzalo, J., Martín-Wanton, T., Meij, E., de Rijke, M., Spina, D.: Overview of RepLab 2013: Evaluating Online Reputation Monitoring Systems. In: Forner et al. [138], pp. 333–352
  16. Amigó, E., Carrillo de Albornoz, J., Chugur, I., Corujo, A., Gonzalo, J., Meij, E., de Rijke, M., Spina, D.: Overview of RepLab 2014: Author Profiling and Reputation Dimensions for Online Reputation Management. In: Kanoulas et al. [216], pp. 307–322
  17. Amigó, E., Artiles, J., Gonzalo, J., Spina, D., Liu, B., Corujo, A.: WePS3 Evaluation Campaign: Overview of the On-line Reputation Management Task. In: Braschler et al. [61]
  18. Amigó, E., Corujo, A., Gonzalo, J., Meij, E., de Rijke, M.: Overview of RepLab 2012: Evaluating Online Reputation Management Systems. In: Forner et al. [137]
  19. Amigó, E., Gonzalo, J., Verdejo, M.F.: A Comparison of Evaluation Metrics for Document Filtering. In: Forner et al. [136], pp. 38–49
  20. Anderka, M., Stein, B.: Overview of the 1th International Competition on Quality Flaw Prediction in Wikipedia. In: Forner et al. [137]
  21. Andriarczyk, V., Müller, H.: Deep Multimodal Classification of Image Types in Biomedical Journal Figures. In: Bellot et al. [42], pp. 3–14
  22. Angelini, M., Ferro, N., Järvelin, K., Keskustalo, H., Pirkola, A., Santucci, G., Silvello, G.: Cumulated Relative Position: A Metric for Ranking Evaluation. In: Catarci et al. [75], pp. 112–123
  23. Angelini, M., Ferro, N., Larsen, B., Müller, H., Santucci, G., Silvello, G., Tsikrika, T.: Measuring and Analyzing the Scholarly Impact of Experimental Evaluation Initiatives. In: Agosti, M., Catarci, T., Esposito, F. (eds.) Proc. 10th Italian Research Conference on Digital Libraries (IRCDL 2014). vol. 38, pp. 133–137. *Procedia Computer Science*, Vol. 38 (2014)
  24. Angelini, M., Ferro, N., Santucci, G., Silvello, G.: Improving Ranking Evaluation Employing Visual Analytics. In: Forner et al. [138], pp. 29–40
  25. Antunes, H., Lopes, C.: Analyzing the adequacy of readability indicators to a non-English language. In: Crestani et al. [94]
  26. Argamon, S., Juola, P.: Overview of the International Authorship Identification Competition at PAN-2011. In: Petras et al. [373]
  27. Armstrong, T.G., Moffat, A., Webber, W., Zobel, J.: Improvements That Don’t Add Up: Ad-Hoc Retrieval Results Since 1998. In: Cheung, D.W.L., Song, I.Y., Chu, W.W., Hu, X., Lin, J.J. (eds.) Proc. 18th International Conference on Information and Knowledge Management (CIKM 2009). pp. 601–610. ACM Press, New York, USA (2009)
  28. Arni, T., Clough, P., Sanderson, M., Grubinger, M.: Overview of the ImageCLEF-photo 2008 Photographic Retrieval Task. In: Peters et al. [359], pp. 500–511
  29. Arora, P., Foster, J., Jones, G.J.F.: Query Expansion for Sentence Retrieval Using Pseudo Relevance Feedback and Word Embedding. In: Jones et al. [209], pp. 97–103
  30. Artiles, J., Borthwick, A., Gonzalo, J., Sekine, S., Amigó, E.: WePS-3 Evaluation Campaign: Overview of the Web People Search Clustering and Attribute Extraction Tasks. In: Braschler et al. [61]

31. Azarbondyad, H., Marx, M.: How many labels? Determining the Number of Labels in Multi-Label Text Classification. In: Crestani et al. [94]
32. Azarbondyad, H., Saan, F., Dehghani, D., Marx, M., Kamps, J.: Are Topically Diverse Documents Also Interesting? In: Mothe et al. [301], pp. 215–221
33. Azarbondyad, H., Shakery, A., Faili, H.: Exploiting Multiple Translation Resources for English-Persian Cross Language Information Retrieval. In: Forner et al. [138], pp. 93–99
34. Azzopardi, L., Balog, K.: Towards a Living Lab for Information Retrieval Research and Development - A Proposal for a Living Lab for Product Search Tasks. In: Forner et al. [136], pp. 26–37
35. Balog, K., Azzopardi, L., Kamps, J., de Rijke, M.: Overview of WebCLEF 2006. In: Peters et al. [357], pp. 803–819
36. Balog, K., Cappellato, L., Ferro, N., Macdonald, C. (eds.): CLEF 2016 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-1609/> (2016)
37. Baradaran Hashemi, H., Shakery, A., Feili, H.: Creating a Persian-English Comparable Corpus. In: Agosti et al. [9], pp. 27–39
38. Basile, A., Dwyer, G., Medvedeva, M., Rawee, J., Haagsma, H., Nissim, M.: Simply the Best: Minimalist System Trumps Complex Models in Author Profiling. In: Bellot et al. [42], pp. 143–156
39. Baudis, P., Sedivý, J.: Modeling of the Question Answering Task in the YodaQA System. In: Mothe et al. [301], pp. 222–228
40. Bellot, P., Bogers, T., Geva, S., Hall, M.A., Huurdeman, H.C., Kamps, J., Kazai, G., Koolen, M., Moriceau, V., Mothe, J., Preminger, M., SanJuan, E., Schenkel, R., Skov, M., Tannier, X., Walsh, D.: Overview of INEX 2014. In: Kanoulas et al. [216], pp. 212–228
41. Bellot, P., Doucet, A., Geva, S., Gurajada, S., Kamps, J., Kazai, G., Koolen, M., Mishra, A., Moriceau, V., Mothe, J., Preminger, M., SanJuan, E., Schenkel, R., Tannier, X., Theobald, M., Trappett, M., Wang, Q.: Overview of INEX 2013. In: Forner et al. [138], pp. 269–281
42. Bellot, P., Trabelsi, C., Mothe, J., Murtagh, F., Nie, J.Y., Soulier, L., SanJuan, E., Cappellato, L., Ferro, N. (eds.): 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)
43. Beloborodov, A., Braslavski, P., Driker, M.: Towards Automatic Evaluation of Health-Related CQA Data. In: Kanoulas et al. [216], pp. 7–18
44. Bensalem, I., Rosso, P., Chikhi, S.: A New Corpus for the Evaluation of Arabic Intrinsic Plagiarism Detection. In: Forner et al. [138], pp. 53–58
45. Berendsen, R., Tsagkias, M., de Rijke, M., Meij, E.: Generating Pseudo Test Collections for Learning to Rank Scientific Articles. In: Catarci et al. [75], pp. 42–53
46. Berlanga Llavori, R., Jimeno-Yepes, A., Pérez Catalán, M., Rebholz-Schuhmann, D.: Context-Dependent Semantic Annotation in Cross-Lingual Biomedical Resources. In: Forner et al. [138], pp. 120–123
47. Berlanga Llavori, R., Pérez Catalán, M., Museros Cabedo, L., Forcada, R.: Semantic Discovery of Resources in Cloud-Based PACS/RIS Systems. In: Forner et al. [138], pp. 167–178
48. Besançon, R., Chaudiron, S., Mostefa, D., Hamon, O., I., T., Choukri, K.: Overview of CLEF 2008 INFILE Pilot Track. In: Peters et al. [359], pp. 939–946

49. Besançon, R., Chaudiron, S., Mostefa, D., I., T., Choukri, K., Laïb, M.: Information Filtering Evaluation: Overview of CLEF 2009 INFILE Track. In: Peters et al. [360], pp. 342–353
50. Bhaskar, P., Bandyopadhyay, S.: Language Independent Query Focused Snippet Generation. In: Catarci et al. [75], pp. 138–140
51. Blinov, V., Mishchenko, K., Bolotova, V., Braslavski, P.: A Pinch of Humor for Short-Text Conversation: An Information Retrieval Approach. In: Jones et al. [209], pp. 3–15
52. Bordea, G., Thiessard, F., Hamon, T., Mougin, F.: Automatic Query Selection for Acquisition and Discovery of Food-Drug Interactions. In: Bellot et al. [42], pp. 115–120
53. Borri, F., Nardi, A., Peters, C., Ferro, N. (eds.): CLEF 2008 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-1174/> (2008)
54. Borri, F., Nardi, A., Peters, C., Ferro, N. (eds.): CLEF 2009 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-1175/> (2009)
55. Borri, F., Peters, C., Ferro, N. (eds.): CLEF 2004 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-1170/> (2004)
56. Braschler, M.: CLEF 2000 – Overview of Results. In: Peters [345], pp. 89–101
57. Braschler, M.: CLEF 2001 – Overview of Results. In: Peters et al. [354], pp. 9–26
58. Braschler, M.: CLEF 2002 – Overview of Results. In: Peters et al. [355], pp. 9–27
59. Braschler, M.: CLEF 2003 – Overview of Results. In: Peters et al. [356], pp. 44–63
60. Braschler, M., Di Nunzio, G.M., Ferro, N., Peters, C.: CLEF 2004: Ad Hoc Track Overview and Results Analysis. In: Peters et al. [358], pp. 10–26
61. Braschler, M., Harman, D.K., Pianta, E., Ferro, N. (eds.): CLEF 2010 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-1176/> (2010)
62. Braschler, M., Peters, C.: Cross-Language Evaluation Forum: Objectives, Results, Achievements. *Information Retrieval* 7(1–2), 7–31 (2004)
63. Braslavski, P., Verberne, S., Talipov, R.: Show Me How to Tie a Tie: Evaluation of Cross-Lingual Video Retrieval. In: Fuhr et al. [143], pp. 3–15
64. Budíková, P., Batko, M., Botorek, J., Zezula, P.: Search-Based Image Annotation: Extracting Semantics from Similar Images. In: Mothe et al. [301], pp. 327–339
65. Buraya, K., Farseev, A., Filchenkov, A.: Multi-view Personality Profiling Based on Longitudinal Data. In: Bellot et al. [42], pp. 15–27
66. Cabanac, G., Hubert, G., Boughanem, M., Chrismont, C.: Tie-Breaking Bias: Effect of an Uncontrolled Parameter on Information Retrieval Evaluation. In: Agosti et al. [9], pp. 112–123
67. Cappellato, L., Ferro, N., Goeuriot, L., Mandl, T. (eds.): CLEF 2017 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-1866/> (2017)
68. Cappellato, L., Ferro, N., Halvey, M., Kraaij, W. (eds.): CLEF 2014 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-1180/> (2014)
69. Cappellato, L., Ferro, N., Jones, G.J.F., SanJuan, E. (eds.): CLEF 2015 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-1391/> (2015)

70. Cappellato, L., Ferro, N., Losada, D.E., Müller, H. (eds.): CLEF 2019 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-2380/> (2019)
71. Cappellato, L., Ferro, N., Nie, J.Y., Soulier, L. (eds.): CLEF 2018 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-2125/> (2018)
72. Caputo, B., Müller, H., Martínez-Gómez, J., Villegas, M., Acar, B., Patricia, N., Barzegar Marvasti, N., Üsküdarlı, S., Paredes, R., Cazorla, M., García-Varea, I., Morell, V.: ImageCLEF 2014: Overview and Analysis of the Results. In: Kanoulas et al. [216], pp. 192–211
73. Caputo, B., Müller, H., Thomee, B., Villegas, M., Paredes, R., Zellhöfer, D., Goëau, H., Joly, A., Bonnet, P., Martínez-Gómez, J., García-Varea, I., Cazorla, M.: ImageCLEF 2013: The Vision, the Data and the Open Challenges. In: Forner et al. [138], pp. 250–268
74. Cassidy, T., Ji, H., Deng, H., Zheng, J., Han, J.: Analysis and Refinement of Cross-Lingual Entity Linking. In: Catarci et al. [75], pp. 1–12
75. Catarci, T., Forner, P., Hiemstra, D., Peñas, A., Santucci, G. (eds.): Information Access Evaluation. Multilinguality, Multimodality, and Visual Analytics. Proceedings of the Third International Conference of the CLEF Initiative (CLEF 2012). Lecture Notes in Computer Science (LNCS) 7488, Springer, Heidelberg, Germany (2012)
76. Chaa, M., Nouali, O., Bellot, P.: Combining Tags and Reviews to Improve Social Book Search Performance. In: Bellot et al. [42], pp. 64–75
77. Chappell, T., Geva, S.: Overview of the INEX 2012 Relevance Feedback Track. In: Forner et al. [137]
78. Chidlovskii, B., Csurka, G., Clinchant, S.: Evaluating Stacked Marginalised Denoising Autoencoders Within Domain Adaptation Methods. In: Mothe et al. [301], pp. 15–27
79. Chikka, V.R., Mariyasagayam, N., Niwa, Y., Karlapalem, K.: Information Extraction from Clinical Documents: Towards Disease/Disorder Template Filling. In: Mothe et al. [301], pp. 389–401
80. Chuklin, A., Severyn, A., Trippas, J., Alfonseca, E., Silen, H., Spina, D.: Using Audio Transformations to Improve Comprehension in Voice Question Answering. In: Crestani et al. [94]
81. Cimiano, P., Lopez, V., Unger, C., Cabrio, E., Ngonga Ngomo, A.C., Walter, S.: Multilingual Question Answering over Linked Data (QALD-3): Lab Overview. In: Forner et al. [138], pp. 321–332
82. Clough, P., Goodale, P.: Selecting Success Criteria: Experiences with an Academic Library Catalogue. In: Forner et al. [138], pp. 59–70
83. Clough, P., Grubinger, M., Deselaers, T., Hanbury, A., Müller, H.: Overview of the ImageCLEF 2006 Photographic Retrieval and Object Annotation Tasks. In: Peters et al. [357], pp. 223–256
84. Clough, P., Müller, H., Deselaers, T., Grubinger, M., Lehmann, T.M., Jensen, J.R., Hersh, W.R.: The CLEF 2005 Cross-Language Image Retrieval Track. In: Peters et al. [365], pp. 535–557
85. Clough, P., Müller, H., Sanderson, M.: The CLEF 2004 Cross-Language Image Retrieval Track. In: Peters et al. [358], pp. 597–613
86. Clough, P., Sanderson, M.: The CLEF 2003 Cross Language Image Retrieval Track. In: Peters et al. [356], pp. 581–593
87. Clough, P., Tsikrika, T.: Multi-lingual Retrieval of Pictures in ImageCLEF. In: Ferro and Peters [130]

88. Clough, P., Willett, P., Lim, J.: Unfair Means: Use Cases Beyond Plagiarism. In: Mothe et al. [301], pp. 229–234
89. Coello-Guilarte, D.L., Ortega Mendoza, R.M., Villaseñor-Pineda, L., Montes-y Gómez, M.: Cross-lingual depression detection in Twitter using bilingual word-level alignment. In: Crestani et al. [94]
90. Collovini, S., de Bairoos P. Filho, M., Vieira, R.: Analysing the Role of Representation Choices in Portuguese Relation Extraction. In: Mothe et al. [301], pp. 105–116
91. Conlan, O., Fraser, K., Kelly, L., Yousuf, B.: A User Modeling Shared Challenge Proposal. In: Crestani et al. [94]
92. Corezola Pereira, R., Pereira Moreira, V., Galante, R.: A New Approach for Cross-Language Plagiarism Analysis. In: Agosti et al. [9], pp. 15–26
93. Cossu, J.V., Ferreira, E., Janod, K., Gaillard, J., El-Bèze, M.: NLP-Based Classifiers to Generalize Expert Assessments in E-Reputation. In: Mothe et al. [301], pp. 340–351
94. Crestani, F., Braschler, M., Savoy, J., Rauber, A., Müller, H., Losada, D.E., Heinatz, G., Cappellato, L., Ferro, N. (eds.): Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Tenth International Conference of the CLEF Association (CLEF 2019). Lecture Notes in Computer Science (LNCS) 11696, Springer, Heidelberg, Germany (2019)
95. Custódio, J., Paraboni, I.: An Ensemble Approach to Cross-Domain Authorship Attribution. In: Crestani et al. [94]
96. Dadashkarimi, J., Esfahani, H.N., Faili, H., Shakery, A.: SS4MCT: A Statistical Stemmer for Morphologically Complex Texts. In: Fuhr et al. [143], pp. 201–207
97. Daelemans, W., Kestemont, M., Manjavacas, E., Potthast, M., Rangel Pardo, F.M., Rosso, P., Specht, G., Stamatatos, E., Stein, B., Tschuggnall, M., Wiegmann, M., Zangerle, E.: Overview of PAN 2019: Bots and Gender Profiling, Celebrity Profiling, Cross-domain Authorship Attribution and Style Change Detection. In: Crestani et al. [94]
98. Dehghani, M., Azarbonyad, H., Kamps, J., Marx, M.: Two-Way Parsimonious Classification Models for Evolving Hierarchies. In: Fuhr et al. [143], pp. 69–82
99. Déjean, S., Mothe, J., Ullah, M.Z.: Studying the Variability of System Setting Effectiveness by Data Analytics and Visualization. In: Crestani et al. [94]
100. Deneu, B., Servajean, M., Botella, C., Joly, A.: Evaluation of Deep Species Distribution Models using Environment and Co-occurrences. In: Crestani et al. [94]
101. Deselaers, T., Deserno, T.M.: Medical Image Annotation in ImageCLEF 2008. In: Peters et al. [359], pp. 523–530
102. Deselaers, T., Hanbury, A.: The Visual Concept Detection Task in ImageCLEF 2008. In: Peters et al. [359], pp. 531–538
103. Deselaers, T., Hanbury, A., Viitaniemi, V., Benczúr, A.A., Brendel, M., Daróczy, B., Escalante Balderas, H.J., Gevers, T., Hernández-Gracidas, C.A., Hoi, S.C.H., Laaksonen, J., Li, M., Marín Castro, H.M., Ney, H., Rui, X., Sebe, N., Stöttinger, J., Wu, L.: Overview of the ImageCLEF 2007 Object Retrieval Task. In: Peters et al. [366], pp. 445–471
104. Devezas, J., Nunes, S.: Index-Based Semantic Tagging for Efficient Query Interpretation. In: Fuhr et al. [143], pp. 208–213
105. Di Buccio, E., Dussin, M., Ferro, N., Masiero, I., Santucci, G., Tino, G.: To Re-rank or to Re-query: Can Visual Analytics Solve This Dilemma? In: Forner et al. [136], pp. 119–130
106. Di Nunzio, G.M., Ferro, N., Jones, G.J.F., Peters, C.: CLEF 2005: Ad Hoc Track Overview. In: Peters et al. [365], pp. 11–36

107. Di Nunzio, G.M., Ferro, N., Mandl, T., Peters, C.: CLEF 2006: Ad Hoc Track Overview. In: Peters et al. [357], pp. 21–34
108. Di Nunzio, G.M., Ferro, N., Mandl, T., Peters, C.: CLEF 2007: Ad Hoc Track Overview. In: Peters et al. [366], pp. 13–32
109. Di Nunzio, G.M., Leveling, J., Mandl, T.: LogCLEF 2011 Multilingual Log File Analysis: Language Identification, Query Classification, and Success of a Query. In: Petras et al. [373]
110. Di Nunzio, G.M., Vezzani, F.: Using R Markdown for Replicable Experiments in Evidence Based Medicine. In: Bellot et al. [42], pp. 28–39
111. Dicente Cid, Y., Batmanghelich, K., Müller, H.: Textured Graph-Based Model of the Lungs: Application on Tuberculosis Type Classification and Multi-drug Resistance Detection. In: Bellot et al. [42], pp. 157–168
112. Dietz, F., Petras, V.: A Component-Level Analysis of an Academic Search Test Collection. - Part I: System and Collection Configurations. In: Jones et al. [209], pp. 16–28
113. Dietz, F., Petras, V.: A Component-Level Analysis of an Academic Search Test Collection. - Part II: Query Analysis. In: Jones et al. [209], pp. 29–42
114. Domann, J., Lommatzsch, A.: A Highly Available Real-Time News Recommender Based on Apache Spark. In: Jones et al. [209], pp. 161–172
115. Elsayed, T., Nakov, P., Barrón-Cedeño, A., Hasanain, M., Suwaileh, R., Da San Martino, G., Atanasova, P.: Overview of the CLEF-2019 CheckThat!: Automatic Identification and Verification of Claims. In: Crestani et al. [94]
116. Ermakova, K.: A Method for Short Message Contextualization: Experiments at CLEF/INEX. In: Mothe et al. [301], pp. 352–363
117. Ermakova, L., Goeuriot, L., Mothe, J., Mulhem, P., Nie, J.Y., SanJuan, E.: CLEF 2017 Microblog Cultural Contextualization Lab Overview. In: Jones et al. [209], pp. 304–314
118. Esuli, F., Sebastiani, F.: Evaluating Information Extraction. In: Agosti et al. [9], pp. 100–111
119. Ezzeldin, A.M., Kholief, M.H., El-Sonbaty, Y.: ALQASIM: Arabic Language Question Answer Selection in Machines. In: Forner et al. [138], pp. 100–103
120. Federico, M., Bertoldi, N., Levow, G.A., Jones, G.J.F.: CLEF 2004 Cross-Language Spoken Document Retrieval Track. In: Peters et al. [358], pp. 816–820
121. Federico, M., Jones, G.J.F.: The CLEF 2003 Cross-Language Spoken Document Retrieval Track. In: Peters et al. [356], p. 646
122. Ferrante, M., Ferro, N., Maistro, M.: Rethinking How to Extend Average Precision to Graded Relevance. In: Kanoulas et al. [216], pp. 19–30
123. Ferro, N.: What Happened in CLEF... For a While? In: Crestani et al. [94]
124. Ferro, N., Fuhr, N., Grefenstette, G., Konstan, J.A., Castells, P., Daly, E.M., Declerck, T., Ekstrand, M.D., Geyer, W., Gonzalo, J., Kuflik, T., Lindén, K., Magnini, B., Nie, J.Y., Perego, R., Shapira, B., Soboroff, I., Tintarev, N., Verspoor, K., Willemsen, M.C., Zobel, J.: Manifesto from Dagstuhl Perspectives Workshop 17442 – From Evaluating to Forecasting Performance: How to Turn Information Retrieval, Natural Language Processing and Recommender Systems into Predictive Sciences. Dagstuhl Manifestos, Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Germany 7(1), 96–139 (2018)
125. Ferro, N., Fuhr, N., Maistro, M., Sakai, T., Soboroff, I.: CENTRE@CLEF 2019. In: Azzopardi, L., Stein, B., Fuhr, N., Mayr, P., Hauff, C., Hiemstra, D. (eds.) *Advances in Information Retrieval. Proc. 41st European Conference on IR Research (ECIR 2019) – Part II*. pp. 283–290. Lecture Notes in Computer Science (LNCS) 11438, Springer, Heidelberg, Germany (2019)

126. Ferro, N., Harman, D.: CLEF 2009: Grid@CLEF Pilot Track Overview. In: Peters et al. [360], pp. 552–565
127. Ferro, N., Maistro, M., Sakai, T., Soboroff, I.: CENTRE@CLEF2018: Overview of the Replicability Task. In: Cappellato et al. [71]
128. Ferro, N., Peters, C.: CLEF 2009 Ad Hoc Track Overview: TEL & Persian Tasks. In: Peters et al. [360], pp. 13–35
129. Ferro, N., Peters, C.: From Multilingual to Multimodal: The Evolution of CLEF over Two Decades. In: Information Retrieval Evaluation in a Changing World – Lessons Learned from 20 Years of CLEF [130]
130. Ferro, N., Peters, C. (eds.): Information Retrieval Evaluation in a Changing World – Lessons Learned from 20 Years of CLEF, The Information Retrieval Series, vol. 41. Springer International Publishing, Germany (2019)
131. Ferro, N., Santucci, G.: Visual Analytics and IR Experimental Evaluation. In: Ferro and Peters [130]
132. Ferro, N., Silvello, G.: CLEF 15th Birthday: What Can We Learn From Ad Hoc Retrieval? In: Kanoulas et al. [216], pp. 31–43
133. Ferro, N., Silvello, G.: The CLEF Monolingual Grid of Points. In: Fuhr et al. [143], pp. 16–27
134. Ferro, N., Silvello, G.: 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)
135. Fontanella, S., Rodríguez-Sánchez, A.J., Piater, J., Szedmak, S.: Kronecker Decomposition for Image Classification. In: Fuhr et al. [143], pp. 137–149
136. Forner, P., Gonzalo, J., Kekäläinen, J., Lalmas, M., de Rijke, M. (eds.): Multilingual and Multimodal Information Access Evaluation. Proceedings of the Second International Conference of the Cross-Language Evaluation Forum (CLEF 2011). Lecture Notes in Computer Science (LNCS) 6941, Springer, Heidelberg, Germany (2011)
137. Forner, P., Karlgren, J., Womser-Hacker, C., Ferro, N. (eds.): CLEF 2012 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-1178/> (2012)
138. Forner, P., Müller, H., Paredes, R., Rosso, P., Stein, B. (eds.): Information Access Evaluation meets Multilinguality, Multimodality, and Visualization. Proceedings of the Fourth International Conference of the CLEF Initiative (CLEF 2013). Lecture Notes in Computer Science (LNCS) 8138, Springer, Heidelberg, Germany (2013)
139. Forner, P., Navigli, R., Tufis, D., Ferro, N. (eds.): CLEF 2013 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-1179/> (2013)
140. Forner, P., Peñas, A., Agirre, E., Alegria, I., Forascu, C., Moreau, N., Osenova, P., Prokopidis, P., Rocha, P., Sacaleanu, B., Sutcliffe, R.F.E., Sang, E.F.T.K.: Overview of the Clef 2008 Multilingual Question Answering Track. In: Peters et al. [359], pp. 262–295
141. Franco-Salvador, M., Rangel Pardo, F.M., Rosso, P., Taulé, M., Martí, M.A.: Language Variety Identification Using Distributed Representations of Words and Documents. In: Mothe et al. [301], pp. 28–40
142. Fuhr, N.: Reproducibility and Validity in CLEF. In: Ferro and Peters [130]
143. Fuhr, N., Quaresma, P., Gonçalves, T., Larsen, B., Balog, K., Macdonald, C., Cappellato, L., Ferro, N. (eds.): Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Seventh International Conference of

- the CLEF Association (CLEF 2016). Lecture Notes in Computer Science (LNCS) 9822, Springer, Heidelberg, Germany (2016)
144. Gäde, M., Ferro, N., Lestari Paramita, M.: CHiC 2011 – Cultural Heritage in CLEF: From Use Cases to Evaluation in Practice for Multilingual Information Access to Cultural Heritage. In: Petras et al. [373]
  145. Gäde, M., Stiller, J., Petras, V.: Which Log for Which Information? Gathering Multilingual Data from Different Log File Types. In: Agosti et al. [9], pp. 70–81
  146. Galuscáková, P., Pecina, P., Hajic, J.: Penalty Functions for Evaluation Measures of Unsegmented Speech Retrieval. In: Catarci et al. [75], pp. 100–111
  147. Ganguly, D., Jones, G.J.F.: A Gamified Approach to Relevance Judgement. In: Fuhr et al. [143], pp. 214–220
  148. Ganguly, D., Leveling, J., Jones, G.J.F.: Simulation of Within-Session Query Variations Using a Text Segmentation Approach. In: Forner et al. [136], pp. 89–94
  149. Ganguly, D., Leveling, J., Jones, G.J.F.: A Case Study in Decomposing for Bengali Information Retrieval. In: Forner et al. [138], pp. 108–119
  150. Gebremeskel, G.G., de Vries, A.P.: Random Performance Differences Between Online Recommender System Algorithms. In: Fuhr et al. [143], pp. 187–200
  151. Gey, F., Larson, R., Sanderson, M., Bischoff, K., Mandl, T., Womser-Hacker, K., Santos, D., Rocha, P., Di Nunzio, G.M., Ferro, N.: GeoCLEF 2006: the CLEF 2006 Cross-Language Geographic Information Retrieval Track Overview. In: Peters et al. [357], pp. 852–876
  152. Gey, F.C., Larson, R.R., Sanderson, M., Joho, H., Clough, P., Petras, V.: Geo-CLEF: The CLEF 2005 Cross-Language Geographic Information Retrieval Track Overview. In: Peters et al. [365], pp. 908–919
  153. Ghosh, S., Singhanian, P., Singh, S., Rudra, K., Ghosh, S.: Stance Detection in Web and Social Media : A Comparative Study. In: Crestani et al. [94]
  154. Giampiccolo, D., Forner, P., Herrera, J., Peñas, A., Ayache, C., Forascu, C., Jijkoun, V., Osenova, P., Rocha, P., Sacaleanu, B., Sutcliffe, R.F.E.: Overview of the CLEF 2007 Multilingual Question Answering Track. In: Peters et al. [366], pp. 200–236
  155. Gînsca, A.L., Popescu, A., Lupu, M., Iftene, A., Kanellos, I.: Evaluating User Image Tagging Credibility. In: Mothe et al. [301], pp. 41–52
  156. Glinos, D.G.: Discovering Similar Passages within Large Text Documents. In: Kanoulas et al. [216], pp. 98–109
  157. Gobeill, J., Gaudinat, A., Ruch, P.: Instance-Based Learning for Tweet Monitoring and Categorization. In: Mothe et al. [301], pp. 235–240
  158. Goëau, H., Bonnet, P., Joly, A., Boujemaa, N., Barthelemy, D., Molino, J.F., Birnbaum, P., Mouysset, E., Picard, M.: The CLEF 2011 Plant Images Classification Task. In: Petras et al. [373]
  159. Goëau, H., Bonnet, P., Joly, A., Yahiaoui, I., Barthelemy, D., Boujemaa, N., Molino, J.F.: The ImageCLEF 2012 Plant Identification Task. In: Forner et al. [137]
  160. Goeuriot, L., Kelly, L., Suominen, H., Hanlen, L., Névéal, A., Grouin, C., Palotti, J.R.M., Zuccon, G.: Overview of the CLEF eHealth Evaluation Lab 2015. In: Mothe et al. [301], pp. 429–443
  161. Goeuriot, L., Kelly, L., Suominen, H., Névéal, A., Robert, A., Kanoulas, E., Spijker, R., Palotti, J.R.M., Zuccon, G.: CLEF 2017 eHealth Evaluation Lab Overview. In: Jones et al. [209], pp. 291–303
  162. Goeuriot, L., Mothe, J., Mulhem, P., Murtagh, F., SanJuan, E.: Overview of the CLEF 2016 Cultural Micro-blog Contextualization Workshop. In: Fuhr et al. [143], pp. 371–378

163. Gollub, T., Potthast, M., Beyer, A., Busse, M., Rangel Pardo, F., Rosso, P., Stamatatos, E., Stenno, B.: Recent Trends in Digital Text Forensics and Its Evaluation - Plagiarism Detection, Author Identification, and Author Profiling. In: Forner et al. [138], pp. 282–302
164. Gómez-Adorno, H., Martín-del Campo-Rodríguez, C., Sidorov, G., Alemán, Y., Vilariño, D., Pinto, D.: Hierarchical Clustering Analysis: The Best-Performing Approach at PAN 2017 Author Clustering Task. In: Bellot et al. [42], pp. 216–223
165. Gonzalo, J., Clough, P., Karlgren, J.: Overview of iCLEF 2008: Search Log Analysis for Multilingual Image Retrieval. In: Peters et al. [359], pp. 227–235
166. Gonzalo, J., Clough, P., Vallin, A.: Overview of the CLEF 2005 Interactive Track. In: Peters et al. [365], pp. 251–262
167. Gonzalo, J., Oard, D.W.: The CLEF 2002 Interactive Track. In: Peters et al. [355], pp. 372–382
168. Gonzalo, J., Oard, D.W.: iCLEF 2004 Track Overview: Pilot Experiments in Interactive Cross-Language Question Answering. In: Peters et al. [358], pp. 310–322
169. Gonzalo, J., Peinado, V., Clough, P., Karlgren, J.: Overview of iCLEF 2009: Exploring Search Behaviour in a Multilingual Folksonomy Environment. In: Peters et al. [368], pp. 13–20
170. Goodwin, T., Harabagiu, S.M.: The Impact of Belief Values on the Identification of Patient Cohorts. In: Forner et al. [138], pp. 155–166
171. Grotov, A., Chuklin, A., Markov, I., Stout, L., Xumara, F., de Rijke, M.: A Comparative Study of Click Models for Web Search. In: Mothe et al. [301], pp. 78–90
172. Grubinger, M., Clough, P., Hanbury, A., Müller, H.: Overview of the ImageCLEF-photo 2007 Photographic Retrieval Task. In: Peters et al. [366], pp. 433–444
173. Gupta, P., Barrón-Cedeño, A., Rosso, P.: Cross-Language High Similarity Search Using a Conceptual Thesaurus. In: Catarci et al. [75], pp. 67–75
174. Hagen, M., Glimm, C.: Supporting More-Like-This Information Needs: Finding Similar Web Content in Different Scenarios. In: Kanoulas et al. [216], pp. 50–61
175. Hall, M., Toms, E.: Building a Common Framework for IIR Evaluation. In: Forner et al. [138], pp. 17–28
176. Halvani, O., Graner, L.: Rethinking the Evaluation Methodology of Authorship Verification Methods. In: Bellot et al. [42], pp. 40–51
177. Hammarström, H.: Automatic Annotation of Bibliographical References for Descriptive Language Materials. In: Forner et al. [136], pp. 62–73
178. Han Lee, S., Loong Chang, Y., Seng Chan, S., Alexis, J., Bonnet, P., Goëau, H.: Plant Classification Based on Gated Recurrent Unit. In: Bellot et al. [42], pp. 169–180
179. Hanbury, A., Müller, H.: Automated Component-Level Evaluation: Present and Future. In: Agosti et al. [9], pp. 124–135
180. Hanbury, A., Müller, H., Langs, G., Weber, M.A., Menze, B.H., Salas Fernandez, T.: Bringing the Algorithms to the Data: Cloud-Based Benchmarking for Medical Image Analysis. In: Catarci et al. [75], pp. 24–29
181. Harman, D.K.: Information Retrieval Evaluation. Morgan & Claypool Publishers, USA (2011)
182. Harman, D.K., Voorhees, E.M. (eds.): TREC. Experiment and Evaluation in Information Retrieval. MIT Press, Cambridge (MA), USA (2005)
183. Harris, C.G., Xu, T.: The Importance of Visual Context Clues in Multimedia Translation. In: Forner et al. [136], pp. 107–118

184. Hasan, S.A., Ling, Y., Liu, J., Sreenivasan, R., Anand, S., Raj Arora, T., Datla, V., Lee, K., Qadir, A., Swisher, C., Farri, O.: Attention-Based Medical Caption Generation with Image Modality Classification and Clinical Concept Mapping. In: Bellot et al. [42], pp. 224–230
185. Hiemstra, D., Hauff, C.: MapReduce for Information Retrieval Evaluation: “Let’s Quickly Test This on 12 TB of Data”. In: Agosti et al. [9], pp. 64–69
186. Hoang, T.B.N., Mothe, J., Baillon, M.: TwitCID: a Collection of Data Sets for Studies on Information Diffusion on Social Networks. In: Crestani et al. [94]
187. Hopfgartner, F., Balog, K., Lommatzsch, A., Kelly, L., Kille, B., Schuth, A., Larson, M.: Continuous Evaluation of Large-scale Information Access Systems: A Case for Living Labs. In: Ferro and Peters [130]
188. Hopfgartner, F., Kille, B., Lommatzsch, A., Plumbaum, T., Brodt, T., Heintz, T.: Benchmarking News Recommendations in a Living Lab. In: Kanoulas et al. [216], pp. 250–267
189. Hull, D.A., Oard, D.W.: Cross-Language Text and Speech Retrieval – Papers from the AAAI Spring Symposium. Association for the Advancement of Artificial Intelligence (AAAI), Technical Report SS-97-05, <http://www.aaai.org/Press/Reports/Symposia/Spring/ss-97-05.php> (1997)
190. Hürriyetoglu, A., Yörük, E., Yuret, D., Yoltar, Ç., Gürel, B., Duruşan, F., Mutlu, O., Akdemir, A.: Overview of CLEF 2019 Lab ProtestNews: Extracting Protests from News in a Cross-context Setting. In: Crestani et al. [94]
191. Huurnink, B., Hofmann, K., de Rijke, M., Bron, M.: Validating Query Simulators: An Experiment Using Commercial Searches and Purchases. In: Agosti et al. [9], pp. 40–51
192. Imhof, M., Braschler, M.: Are Test Collections “Real”? Mirroring Real-World Complexity in IR Test Collections. In: Mothe et al. [301], pp. 241–247
193. Inches, G., Crestani, F.: Overview of the International Sexual Predator Identification Competition at PAN-2012. In: Forner et al. [137]
194. Ionescu, B., Müller, H., Péteri, R., Dicente Cid, Y., Liauchuk, V., Kovalev, V., Klimuk, D., Tarasau, A., Ben Abacha, A., Hasan, S.A., Datla, V., Liu, J., Demner-Fushman, D., Dang-Nguyen, D.T., Piras, L., Riegler, M., Tran, M.T., Lux, M., Gurrin, C., Pelka, O., Friedrich, C.M., Garcia Seco de Herrera, A., Garcia, N., Kavallieratou, E., del Blanco, C.R., Cuevas Rodríguez, C., Vasillopoulos, N., Karampidis, K., Chamberlain, J., Clark, A., Campello, A.: ImageCLEF 2019: Multimedia Retrieval in Medicine, Lifelogging, Security and Nature. In: Crestani et al. [94]
195. Ionescu, B., Müller, H., Villegas, M., Arenas, H., Boato, G., Dang-Nguyen, D.T., Dicente Cid, Y., Eickhoff, C., Garcia Seco de Herrera, A., Gurrin, C., Islam, M.B., Kovalev, V., Liauchuk, V., Mothe, J., Piras, L., Riegler, M., Schwall, I.: Overview of ImageCLEF 2017: Information Extraction from Images. In: Jones et al. [209], pp. 315–337
196. Ionescu, B., Müller, H., Villegas, M., Garcia Seco de Herrera, A., Eickhoff, C., Andrearczyk, V., Dicente Cid, Y., Liauchuk, V., Kovalev, V., Hasan, S.A., Ling, Y., Farri, O., Liu, J., Lungren, M., Dang-Nguyen, D.T., Piras, L., Riegler, M., Zhou, L., Lux, M., Gurrin, C.: Overview of ImageCLEF 2018: Challenges, Datasets and Evaluation. In: Bellot et al. [42], pp. 309–334
197. Jabeur, L.B., Soulier, L., Tamine, L., Mousset, P.: A Product Feature-Based User-Centric Ranking Model for E-Commerce Search. In: Fuhr et al. [143], pp. 174–186
198. Jijkoun, V., de Rijke, M.: Overview of WebCLEF 2007. In: Peters et al. [366], pp. 725–731

199. Jijkoun, V., de Rijke, M.: Overview of WebCLEF 2008. In: Peters et al. [359], pp. 787–793
200. Joly, A., Goëau, H., Botella, C., Glotin, H., Bonnet, P., Vellinga, W.P., Planqué, R., Müller, H.: Overview of LifeCLEF 2018: A Large-Scale Evaluation of Species Identification and Recommendation Algorithms in the Era of AI. In: Bellot et al. [42], pp. 247–266
201. Joly, A., Goëau, H., Botella, C., Kahl, S., Servajean, M., Glotin, H., Bonnet, P., Vellinga, W.P., Planqué, R., Stöter, F.R., Müller, H.: Overview of LifeCLEF 2019: Identification of Amazonian Plants, South & North American Birds, and Niche Prediction. In: Crestani et al. [94]
202. Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W.P., Lombardo, J.C., Planqué, R., Palazzo, S., Müller, H.: LifeCLEF 2017 Lab Overview: Multimedia Species Identification Challenges. In: Jones et al. [209], pp. 255–274
203. Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W.P., Lombardo, J.C., Planqué, R., Palazzo, S., Müller, H.: Biodiversity Information Retrieval through Large Scale Content-Based Identification: A Long-Term Evaluation. In: Ferro and Peters [130]
204. Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W.P., Planqué, R., Rauber, A., Fisher, R.B., Müller, H.: LifeCLEF 2014: Multimedia Life Species Identification Challenges. In: Kanoulas et al. [216], pp. 229–249
205. Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W.P., Planqué, R., Rauber, A., Palazzo, S., Fisher, B., Müller, H.: LifeCLEF 2015: Multimedia Life Species Identification Challenges. In: Mothe et al. [301], pp. 462–483
206. Joly, A., Goëau, H., Spampinato, C., Bonnet, P., Vellinga, W.P., Champ, J., Planqué, R., Palazzo, S., Müller, H.: LifeCLEF 2016: Multimedia Life Species Identification Challenges. In: Fuhr et al. [143], pp. 286–310
207. Jones, G.J.F.: 'Bout Sound and Vision: CLEF beyond Text Retrieval Tasks. In: Ferro and Peters [130]
208. Jones, G.J.F., Federico, M.: CLEF 2002 Cross-Language Spoken Document Retrieval Pilot Track Report. In: Peters et al. [355], pp. 446–457
209. Jones, G.J.F., Lawless, S., Gonzalo, J., Kelly, L., Goeuriot, L., Mandl, T., Cappellato, L., Ferro, N. (eds.): 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)
210. Juola, P.: An Overview of the Traditional Authorship Attribution Subtask. In: Forner et al. [137]
211. Jürgens, J., Hansen, P., Womser-Hacker, C.: Going beyond CLEF-IP: The 'Reality' for Patent Searchers. In: Catarci et al. [75], pp. 30–35
212. Kalpathy-Cramer, J., Müller, H., Bedrick, S., Eggel, I., Garcia Seco de Herrera, A., Tsikrika, T.: Overview of the CLEF 2011 Medical Image Classification and Retrieval Tasks. In: Petras et al. [373]
213. Kamps, J., Koolen, M., Geva, S., Schenkel, R., SanJuan, E., Bogers, T.: From XML Retrieval to Semantic Search and Beyond. In: Ferro and Peters [130]
214. Kanoulas, E., Azzopardi, L.: CLEF 2017 Dynamic Search Evaluation Lab Overview. In: Jones et al. [209], pp. 361–366
215. Kanoulas, E., Azzopardi, L., Hui Yang, G.: Overview of the CLEF Dynamic Search Evaluation Lab 2018. In: Bellot et al. [42], pp. 362–371
216. Kanoulas, E., Lupu, M., Clough, P., Sanderson, M., Hall, M., Hanbury, A., Toms, E. (eds.): 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 (2014)
217. Karadzhov, G., Mihaylova, T., Kiproff, Y., Georgiev, G., Koychev, Y., Nakov, P.: The Case for Being Average: A Mediocrity Approach to Style Masking and Author Obfuscation. In: Jones et al. [209], pp. 173–185
  218. Karan, M., Snajder, J.: Evaluation of Manual Query Expansion Rules on a Domain Specific FAQ Collection. In: Mothe et al. [301], pp. 248–253
  219. Karisani, P., Oroumchian, F., Rahgozar, M.: Tweet Expansion Method for Filtering Task in Twitter. In: Mothe et al. [301], pp. 55–64
  220. Karlgren, J.: Adopting Systematic Evaluation Benchmarks in Operational Settings. In: Ferro and Peters [130]
  221. Karlgren, J.: How Lexical Gold Standards Have Effects On The Usefulness Of Text Analysis Tools For Digital Scholarship. In: Crestani et al. [94]
  222. Karlgren, J., Callin, J., Collins-Thompson, K., Cuba Gyllensten, A., Ekgrén, A., Jurgens, D., Korhonen, A., Olsson, F., Sahlgrén, M., Schütze, H.: Evaluating Learning Language Representations. In: Mothe et al. [301], pp. 254–260
  223. Karlgren, J., Gonzalo, J., Clough, P.: iCLEF 2006 Overview: Searching the Flickr WWW Photo-Sharing Repository. In: Peters et al. [357], pp. 186–194
  224. Karlgren, J., Hedlund, T., Järvelin, K., Keskustalo, H., Kettunen, K.: The Challenges of Language Variation in Information Access. In: Ferro and Peters [130]
  225. Karlsson, V., Herman, P., Karlgren, J.: Evaluating Categorisation in Real Life – An Argument Against Simple but Impractical Metrics. In: Fuhr et al. [143], pp. 221–226
  226. Kavallieratou, E., del Blanco, C.R., Cuevas, C., García, N.: Interactive Learning-based Retrieval Technique for Visual Lifelogging. In: Crestani et al. [94]
  227. Kelly, L., Goeuriot, L., Suominen, H., Névél, A., Palotti, J., Zuccon, G.: Overview of the CLEF eHealth Evaluation Lab 2016. In: Fuhr et al. [143], pp. 255–266
  228. Kelly, L., Goeuriot, L., Suominen, H., Schreck, T., Leroy, G., Mowery, D.L., Velupillai, S., Webber Chapman, W., Martínez, D., Zuccon, G., Palotti, J.R.M.: Overview of the ShARe/CLEF eHealth Evaluation Lab 2014. In: Kanoulas et al. [216], pp. 172–191
  229. Keszler, A., Kovács, L., Szirányi, T.: The Appearance of the Giant Component in Descriptor Graphs and Its Application for Descriptor Selection. In: Catarci et al. [75], pp. 76–81
  230. Kharazmi, S., Scholer, F., Vallet, D., Sanderson, M.: Examining Additivity and Weak Baselines. *ACM Transactions on Information Systems (TOIS)* 34(4), 23:1–23:18 (June 2016)
  231. Khwileh, A., Ganguly, D., Jones, G.J.F.: An Investigation of Cross-Language Information Retrieval for User-Generated Internet Video. In: Mothe et al. [301], pp. 117–129
  232. Khwileh, A., Way, A., Jones, G.J.F.: Improving the Reliability of Query Expansion for User-Generated Speech Retrieval Using Query Performance Prediction. In: Jones et al. [209], pp. 43–56
  233. Kille, B., Lommatzsch, A., Gebremeskel, G.G., Hopfgartner, F., Larson, M., Seiler, J., Malagoli, D., Serény, A., Brodt, T., de Vries, A.P.: Overview of NewsREEL’16: Multi-dimensional Evaluation of Real-Time Stream-Recommendation Algorithms. In: Fuhr et al. [143], pp. 311–331

234. Kille, B., Lommatzsch, A., Turrin, R., Serény, A., Larson, M., Brodt, T., Seiler, J., Hopfgartner, F.: Stream-Based Recommendations: Online and Offline Evaluation as a Service. In: Mothe et al. [301], pp. 497–517
235. Kim, S.J., Lee, J.H.: Subtopic Mining Based on Head-Modifier Relation and Co-occurrence of Intents Using Web Documents. In: Forner et al. [138], pp. 179–191
236. Kliegr, T., Kuchar, J.: Benchmark of Rule-Based Classifiers in the News Recommendation Task. In: Mothe et al. [301], pp. 130–141
237. Kluck, M.: The Domain-Specific Track in CLEF 2004: Overview of the Results and Remarks on the Assessment Process. In: Peters et al. [358], pp. 260–270
238. Kluck, M., Gey, F.C.: The Domain-Specific Task of CLEF – Specific Evaluation Strategies in Cross-Language Information Retrieval. In: Peters [345], pp. 48–56
239. Kluck, M., Stempfhuber, M.: Domain-Specific Track CLEF 2005: Overview of Results and Approaches, Remarks on the Assessment Analysis. In: Peters et al. [365], pp. 212–221
240. Kocher, M., Savoy, J.: Author Clustering with an Adaptive Threshold. In: Jones et al. [209], pp. 186–198
241. Koitka, S., Friedrich, C.M.: Optimized Convolutional Neural Network Ensembles for Medical Subfigure Classification. In: Jones et al. [209], pp. 57–68
242. Koolen, M., Bogers, T., Gäde, M., Hall, M., Hendrickx, I., Huurdeman, H.C., Kamps, J., Skov, M., Verberne, S., Walsh, D.: Overview of the CLEF 2016 Social Book Search Lab. In: Fuhr et al. [143], pp. 351–370
243. Koolen, M., Bogers, T., Gäde, M., Hall, M.A., Huurdeman, H.C., Kamps, J., Skov, M., Toms, E., Walsh, D.: Overview of the CLEF 2015 Social Book Search Lab. In: Mothe et al. [301], pp. 545–564
244. Koolen, M., Kazai, G., Kamps, J., Preminger, M., Doucet, A., Landoni, M.: Overview of the INEX 2012 Social Book Search Track. In: Forner et al. [137]
245. Koops, H.V., Van Balen, J., Wiering, F.: Automatic Segmentation and Deep Learning of Bird Sounds. In: Mothe et al. [301], pp. 261–267
246. Kordjamshidi, P., Rahgooy, T., Moens, M.F., Pustejovsky, J., Manzoor, U., Roberts, K.: CLEF 2017: Multimodal Spatial Role Labeling (mSpRL) Task Overview. In: Jones et al. [209], pp. 367–376
247. Kosmopoulos, A., Paliouras, G., Androutsopoulos, I.: The Effect of Dimensionality Reduction on Large Scale Hierarchical Classification. In: Kanoulas et al. [216], pp. 160–171
248. Kumar, N.K., Santosh, G.S.K., Varma, V.: A Language-Independent Approach to Identify the Named Entities in Under-Resourced Languages and Clustering Multilingual Documents. In: Forner et al. [136], pp. 74–82
249. Kurimo, M., Creutz, M., Varjokallio, M.: Morpho Challenge Evaluation Using a Linguistic Gold Standard. In: Peters et al. [366], pp. 864–872
250. Kurimo, M., Turunen, V.T., Varjokallio, M.: Overview of Morpho Challenge 2008. In: Peters et al. [359], pp. 951–966
251. Kurimo, M., Virpioja, S., Turunen, V.T., Blackwood, G.W., Byrne, W.: Overview and Results of Morpho Challenge 2009. In: Peters et al. [360], pp. 587–597
252. Kürsten, J., Eibl, M.: Comparing IR System Components Using Beanplots. In: Catarci et al. [75], pp. 136–137
253. Kvist, M., Velupillai, S.: SCAN: A Swedish Clinical Abbreviation Normalizer - Further Development and Adaptation to Radiology. In: Kanoulas et al. [216], pp. 62–73
254. de L. Pertile, S., Pereira Moreira, V.: A Test Collection to Evaluate Plagiarism by Missing or Incorrect References. In: Catarci et al. [75], pp. 141–143

255. de L. Pertile, S., Rosso, P., Pereira Moreira, V.: Counting Co-occurrences in Citations to Identify Plagiarised Text Fragments. In: Forner et al. [138], pp. 150–154
256. Lagopoulos, A., Anagnostou, A., Minas, A., Tsoumakas, G.: Learning-to-Rank and Relevance Feedback for Literature Appraisal in Empirical Medicine. In: Bellot et al. [42], pp. 52–63
257. Lai, M., Tambuscio, M., Patti, V., Ruffo, G., Rosso, P.: Extracting Graph Topological Information and Users' Opinion. In: Jones et al. [209], pp. 112–118
258. Landoni, M., Matteri, D., Murgia, E., Huibers, T., Soledad Pera, M.: Sonny, Cerca! Evaluating the Impact of Using a Vocal Assistant to Search at School. In: Crestani et al. [94]
259. Larsen, B.: The Scholarly Impact of CLEF 2010-2017. In: Ferro and Peters [130]
260. Larson, M., Newman, E., Jones, G.J.F.: Overview of VideoCLEF 2008: Automatic Generation of Topic-Based Feeds for Dual Language Audio-Visual Content. In: Peters et al. [359], pp. 906–917
261. Larson, M., Newman, E., Jones, G.J.F.: Overview of VideoCLEF 2009: New Perspectives on Speech-Based Multimedia Content Enrichment. In: Peters et al. [368], pp. 354–368
262. Lasneck, M.: Towards Automatic Large-Scale Identification of Birds in Audio Recordings. In: Mothe et al. [301], pp. 364–375
263. Leiva, L.A., Villegas, M., Paredes, R.: Relevant Clouds: Leveraging Relevance Feedback to Build Tag Clouds for Image Search. In: Forner et al. [138], pp. 143–149
264. Leong, C.W., Hassan, S., Ruiz, M.E., Rada, M.: Improving Query Expansion for Image Retrieval via Saliency and Picturability. In: Forner et al. [136], pp. 137–142
265. Lestari Paramita, M., Sanderson, M., Clough, P.: Diversity in Photo Retrieval: Overview of the ImageCLEFPhoto Task 2009. In: Peters et al. [368], pp. 45–59
266. Li, P., Jiang, X., Kambhamettu, C., Shatkay, H.: Segmenting Compound Biomedical Figures into Their Constituent Panels. In: Jones et al. [209], pp. 199–210
267. Li, W., Jones, G.J.F.: Enhancing Medical Information Retrieval by Exploiting a Content-Based Recommender Method. In: Mothe et al. [301], pp. 142–153
268. Linhares Pontes, E., Huet, S., Torres-Moreno, J.M.: Microblog Contextualization: Advantages and Limitations of a Multi-sentence Compression Approach. In: Bellot et al. [42], pp. 181–190
269. Lipani, A., Piroi, F., Andersson, L., Hanbury, A.: An Information Retrieval Ontology for Information Retrieval Nanopublications. In: Kanoulas et al. [216], pp. 44–49
270. Litvinova, T., Seredin, P., Litvinova, O., Ryzhkova, E.: Estimating the Similarities Between Texts of Right-Handed and Left-Handed Males and Females. In: Jones et al. [209], pp. 119–124
271. Liu, F., Peng, Y., Rosen, M.P.: An Effective Deep Transfer Learning and Information Fusion Framework for Medical Visual Question Answering. In: Crestani et al. [94]
272. Lommatzsch, A., Kille, B., Hopfgartner, F., Larson, M., Brodt, T., Seiler, J., Özgöbek, Ö.: CLEF 2017 NewsREEL Overview: A Stream-Based Recommender Task for Evaluation and Education. In: Jones et al. [209], pp. 239–254
273. Lommatzsch, A., Werner, S.: Optimizing and Evaluating Stream-Based News Recommendation Algorithms. In: Mothe et al. [301], pp. 376–388
274. Loponen, A., Järvelin, K.: A Dictionary- and Corpus-Independent Statistical Lemmatizer for Information Retrieval in Low Resource Languages. In: Agosti et al. [9], pp. 3–14

275. Losada, D.E., Crestani, F.: A Test Collection for Research on Depression and Language Use. In: Fuhr et al. [143], pp. 28–39
276. Losada, D.E., Crestani, F., Parapar, J.: eRISK 2017: CLEF Lab on Early Risk Prediction on the Internet: Experimental Foundations. In: Jones et al. [209], pp. 346–360
277. Losada, D.E., Crestani, F., Parapar, J.: Overview of eRisk: Early Risk Prediction on the Internet. In: Bellot et al. [42], pp. 343–361
278. Losada, D.E., Crestani, F., Parapar, J.: Overview of eRisk 2019: Early Risk Prediction on the Internet. In: Crestani et al. [94]
279. Mackie, S., McCreddie, R., Macdonald, C., Ounis, I.: Comparing Algorithms for Microblog Summarisation. In: Kanoulas et al. [216], pp. 153–159
280. Magdy, W., Jones, G.J.F.: Examining the Robustness of Evaluation Metrics for Patent Retrieval with Incomplete Relevance Judgements. In: Agosti et al. [9], pp. 82–93
281. Magnini, B., Giampiccolo, D., Forner, P., Ayache, C., Jijkoun, V., Osenova, P., Peñas, A., Rocha, P., Sacaleanu, B., Sutcliffe, R.F.E.: Overview of the CLEF 2006 Multilingual Question Answering Track. In: Peters et al. [357], pp. 223–256
282. Magnini, B., Romagnoli, S., Vallin, A., Herrera, J., Peñas, A., Peinado, V., Verdejo, M.F., de Rijke, M.: The Multiple Language Question Answering Track at CLEF 2003. In: Peters et al. [356], pp. 471–486
283. Magnini, B., Vallin, A., Ayache, C., Erbach, G., Peñas, A., de Rijke, M., Rocha, P., Simov, K.I., Sutcliffe, R.F.E.: Overview of the CLEF 2004 Multilingual Question Answering Track. In: Peters et al. [358], pp. 371–391
284. Mandl, T., Agosti, M., Di Nunzio, G.M., Yeh, A.S., Mani, I., Doran, C., Schulz, J.M.: LogCLEF 2009: The CLEF 2009 Multilingual Logfile Analysis Track Overview. In: Peters et al. [360], pp. 508–517
285. Mandl, T., Carvalho, P., Di Nunzio, G.M., Gey, F., Larson, R., Santos, D., Womser-Hacker, C.: GeoCLEF 2008: The CLEF 2008 Cross-Language Geographic Information Retrieval Track Overview. In: Peters et al. [359], pp. 808–821
286. Mandl, T., Di Nunzio, G.M., Schulz, J.M.: LogCLEF 2010: the CLEF 2010 Multilingual Logfile Analysis Track Overview. In: Braschler et al. [61]
287. Mandl, T., Gey, F., Di Nunzio, G.M., Ferro, N., Larson, R., Sanderson, M., Santos, D., Womser-Hacker, C., Xie, X.: GeoCLEF 2007: the CLEF 2007 Cross-Language Geographic Information Retrieval Track Overview. In: Peters et al. [366], pp. 745–772
288. Manotumruksa, J., Macdonald, C., Ounis, I.: Predicting Contextually Appropriate Venues in Location-Based Social Networks. In: Fuhr et al. [143], pp. 96–109
289. Martínez-Gómez, J., García-Varea, I., Caputo, B.: Overview of the ImageCLEF 2012 Robot Vision Task. In: Forner et al. [137]
290. Mayfield, J., Lawrie, D., McNamee, P., Oard, D.W.: Building a Cross-Language Entity Linking Collection in Twenty-One Languages. In: Forner et al. [136], pp. 3–13
291. McCreddie, R., Macdonald, C., Ounis, I., Brassey, J.: A Study of Personalised Medical Literature Search. In: Kanoulas et al. [216], pp. 74–85
292. McMinn, A.J., Jose, J.M.: Real-Time Entity-Based Event Detection for Twitter. In: Mothe et al. [301], pp. 65–77
293. Medvedeva, M., Haagsma, H., Nissim, M.: An Analysis of Cross-Genre and In-Genre Performance for Author Profiling in Social Media. In: Jones et al. [209], pp. 211–223

294. Miftahutdinov, Z., Tutubalina, E.: Deep Learning for ICD Coding: Looking for Medical Concepts in Clinical Documents in English and in French. In: Bellot et al. [42], pp. 203–215
295. Mirsarraf, M.R., Dehghani, N.: A Dependency-Inspired Semantic Evaluation of Machine Translation Systems. In: Forner et al. [138], pp. 71–74
296. Mitrovic, S., Müller, H.: Summarizing Citation Contexts of Scientific Publications. In: Mothe et al. [301], pp. 154–165
297. Molina, A., SanJuan, E., Torres-Moreno, J.M.: A Turing Test to Evaluate a Complex Summarization Task. In: Forner et al. [138], pp. 75–80
298. Molina, S., Mothe, J., Roques, D., Tanguy, L., Ullah, M.Z.: IRIT-QFR: IRIT Query Feature Resource. In: Jones et al. [209], pp. 69–81
299. Morante, R., Daelemans, W.: Overview of the QA4MRE Pilot Task: Annotating Modality and Negation for a Machine Reading Evaluation. In: Petras et al. [373]
300. Moreno, R., Huáng, W., Younus, A., O'Mahony, M.P., Hurley, N.J.: Evaluation of Hierarchical Clustering via Markov Decision Processes for Efficient Navigation and Search. In: Jones et al. [209], pp. 125–131
301. Mothe, J., Savoy, J., Kamps, J., Pinel-Sauvagnat, K., Jones, G.J.F., SanJuan, E., Cappellato, L., Ferro, N. (eds.): Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Sixth International Conference of the CLEF Association (CLEF 2015). Lecture Notes in Computer Science (LNCS) 9283, Springer, Heidelberg, Germany (2015)
302. Mulhem, P., Goeuriot, L., Dogra, N., Amer, N.O.: TimeLine Illustration Based on Microblogs: When Diversification Meets Metadata Re-ranking. In: Jones et al. [209], pp. 224–235
303. Müller, H., Clough, P., Deselaers, T., Caputo, B. (eds.): ImageCLEF – Experimental Evaluation in Visual Information Retrieval. Springer-Verlag, Heidelberg, Germany (2010)
304. Müller, H., Deselaers, T., Deserno, T.M., Clough, P., Kim, E., Hersh, W.R.: Overview of the ImageCLEFmed 2006 Medical Retrieval and Medical Annotation Tasks. In: Peters et al. [357], pp. 595–608
305. Müller, H., Deselaers, T., Deserno, T.M., Kalpathy-Cramer, J., Kim, E., Hersh, W.R.: Overview of the ImageCLEFmed 2007 Medical Retrieval and Medical Annotation Tasks. In: Peters et al. [366], pp. 472–491
306. Müller, H., Hanbury, A.: EaaS: Evaluation-as-a-Service and Experiences from the VISCERAL Project. In: Ferro and Peters [130]
307. Müller, H., Garcia Seco de Herrera, A., Kalpathy-Cramer, J., Demner-Fushman, D., Antani, S., Eggel, I.: Overview of the ImageCLEF 2012 Medical Image Retrieval and Classification Tasks. In: Forner et al. [137]
308. Müller, H., Kalpathy-Cramer, J., Eggel, I., Bedrick, S., Radhouani, S., Bakke, B., Khan Jr., C.E., Hersh, W.R.: Overview of the CLEF 2009 Medical Image Retrieval Track. In: Peters et al. [368], pp. 72–84
309. Müller, H., Kalpathy-Cramer, J., Eggel, I., Bedrick, S., Reisetter, J., Khan Jr., C.E., Hersh, W.R.: Overview of the CLEF 2010 Medical Image Retrieval Track. In: Braschler et al. [61]
310. Müller, H., Kalpathy-Cramer, J., Garcia Seco de Herrera, A.: Experiences from the ImageCLEF Medical Retrieval and Annotation Tasks. In: Ferro and Peters [130]
311. Müller, H., Kalpathy-Cramer, J., Kahn, C.E., Hatt, W., Bedrick, S., Hersh, W.: Overview of the ImageCLEFmed 2008 Medical image Retrieval Task. In: Peters et al. [359], pp. 512–522

312. Murauer, B., Specht, G.: Generating Cross-Domain Text Corpora from Social Media Comments. In: Crestani et al. [94]
313. Nakov, P., Barrón-Cedeño, A., Elsayed, T., Suwaileh, R., Marquez, L., Zaghoulani, W., Atanasova, P., Kyuchukov, S., Da San Martino, G.: Overview of the CLEF-2018 CheckThat! Lab on Automatic Identification and Verification of Political Claims. In: Bellot et al. [42], pp. 372–387
314. Nardi, A., Peters, C., Ferro, N. (eds.): CLEF 2007 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-1173/> (2007)
315. Nardi, A., Peters, C., Vicedo, J.L., Ferro, N. (eds.): CLEF 2006 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-1172/> (2006)
316. Nordlie, R., Pharo, N.: Seven Years of INEX Interactive Retrieval Experiments - Lessons and Challenges. In: Catarci et al. [75], pp. 13–23
317. Nowak, S., Dunker, P.: Overview of the CLEF 2009 Large-Scale Visual Concept Detection and Annotation Task. In: Peters et al. [368], pp. 94–109
318. Nowak, S., Huiskes, M.J.: New Strategies for Image Annotation: Overview of the Photo Annotation Task at ImageCLEF 2010. In: Braschler et al. [61]
319. Nowak, S., Nagel, K., Liebetrau, J.: The CLEF 2011 Photo Annotation and Concept-based Retrieval Tasks. In: Petras et al. [373]
320. Oard, D.W., Gonzalo, J.: The CLEF 2001 Interactive Track. In: Peters et al. [354], pp. 308–319
321. Oard, D.W., Gonzalo, J.: The CLEF 2003 Interactive Track. In: Peters et al. [356], pp. 425–434
322. Oard, D.W., Wang, J., Jones, G.J.F., White, R.W., Pecina, P., Soergel, D., Huang, X., Shafran, I.: Overview of the CLEF-2006 Cross-Language Speech Retrieval Track. In: Peters et al. [357], pp. 744–758
323. Oh, H.S., Jung, Y., Kim, K.Y.: A Multiple-Stage Approach to Re-ranking Medical Documents. In: Mothe et al. [301], pp. 166–177
324. Olvera-Lobo, M.D., Gutiérrez-Artacho, J.: Multilingual Question-Answering System in Biomedical Domain on the Web: An Evaluation. In: Forner et al. [136], pp. 83–88
325. Orio, N., Liem, C.C.S., Peeters, G., Schedl, M.: MusiClef: Multimodal Music Tagging Task. In: Catarci et al. [75], pp. 36–41
326. Orio, N., Rizo, D.: Overview of MusiCLEF 2011. In: Petras et al. [373]
327. Ortega-Mendoza, M., Franco-Arcega, A., López-Monroy, A.P., Montes-y Gómez, M.: I, Me, Mine: The Role of Personal Phrases in Author Profiling. In: Fuhr et al. [143], pp. 110–122
328. Otterbacher, J.: Addressing Social Bias in Information Retrieval. In: Bellot et al. [42], pp. 121–127
329. Pääkkönen, T., Järvelin, K., Kekäläinen, J., Keskustalo, H., Baskaya, F., Maxwell, D., Azzopardi, L.: Exploring Behavioral Dimensions in Session Effectiveness. In: Mothe et al. [301], pp. 178–189
330. Palotti, J., Zuccon, G., Bernhardt, J., Hanbury, A., Goeuriot, L.: Assessors Agreement: A Case Study Across Assessor Type, Payment Levels, Query Variations and Relevance Dimensions. In: Fuhr et al. [143], pp. 40–53
331. Parks, M., Karlgren, J., Stymne, S.: Plausibility Testing for Lexical Resources. In: Jones et al. [209], pp. 132–137
332. Pasi, G., Jones, G.J.F., Curtis, K., Marrara, S., Sanvitto, C., Ganguly, D., Sen, P.: Evaluation of Personalised Information Retrieval at CLEF 2018 (PIR-CLEF). In: Bellot et al. [42], pp. 335–342

333. Pasi, G., Jones, G.J.F., Goeuriot, L., Kelly, L., Marrara, S., Sanvitto, C.: Overview of the CLEF 2019 Personalised Information Retrieval Lab (PIR-CLEF 2019). In: Crestani et al. [94]
334. Pasi, G., Jones, G.J.F., Marrara, S., Sanvitto, C., Ganguly, D., Sen, P.: Overview of the CLEF 2017 Personalised Information Retrieval Pilot Lab (PIR-CLEF 2017). In: Jones et al. [209], pp. 338–345
335. Pecina, P., Hoffmannová, P., Jones, G.J.F., Zhang, Y., Oard, D.W.: Overview of the CLEF-2007 Cross-Language Speech Retrieval Track. In: Peters et al. [366], pp. 674–686
336. Pellegrin, L., Vanegas, J.A., Arevalo, J., Beltrán, V., Escalante, H.J., Montes-y Gómez, M., González, F.A.: A Two-Step Retrieval Method for Image Captioning. In: Fuhr et al. [143], pp. 150–161
337. Peñas, A., Forner, P., Rodrigo, A., Sutcliffe, R.F.E., Forascu, C., Mota, C.: Overview of ResPubliQA 2010: Question Answering Evaluation over European Legislation. In: Braschler et al. [61]
338. Peñas, A., Forner, P., Sutcliffe, R.F.E., Rodrigo, A., Forascu, C., Alegria, I., Giampiccolo, D., Moreau, N., Osenova, P.: Overview of ResPubliQA 2009: Question Answering Evaluation over European Legislation. In: Peters et al. [360], pp. 174–196
339. Peñas, A., Hovy, E.H., Forner, P., Rodrigo, A., Sutcliffe, R.F.E., Forascu, C., Sporleder, C.: Overview of QA4MRE at CLEF 2011: Question Answering for Machine Reading Evaluation. In: Petras et al. [373]
340. Peñas, A., Hovy, E.H., Forner, P., Rodrigo, A., Sutcliffe, R.F.E., Morante, R.: QA4MRE 2011-2013: Overview of Question Answering for Machine Reading Evaluation. In: Forner et al. [138], pp. 303–320
341. Peñas, A., Hovy, E.H., Forner, P., Rodrigo, A., Sutcliffe, R.F.E., Sporleder, C., Forascu, C., Benajiba, Y., Osenova, P.: Overview of QA4MRE at CLEF 2012: Question Answering for Machine Reading Evaluation. In: Forner et al. [137]
342. Peñas, A., Rodrigo, A., Magnini, B., Forner, P., Hovy, E.H., Sutcliffe, R., Giampiccolo, D.: Results and Lessons of the Question Answering Track at CLEF. In: Ferro and Peters [130]
343. Peñas, A., Unger, C., Ngonga Ngomo, A.C.a.: Overview of CLEF Question Answering Track 2014. In: Kanoulas et al. [216], pp. 300–306
344. Peñas, A., Unger, C., Paliouras, P., Kakadiaris, I.A.: Overview of the CLEF Question Answering Track 2015. In: Mothe et al. [301], pp. 539–544
345. Peters, C. (ed.): Cross-Language Information Retrieval and Evaluation: Workshop of Cross-Language Evaluation Forum (CLEF 2000). Lecture Notes in Computer Science (LNCS) 2069, Springer, Heidelberg, Germany (2001)
346. Peters, C.: Introduction. In: Cross-Language Information Retrieval and Evaluation: Workshop of Cross-Language Evaluation Forum (CLEF 2000) [345], pp. 1–6
347. Peters, C.: What Happened in CLEF 2004? In: Peters et al. [358], pp. 1–9
348. Peters, C.: What Happened in CLEF 2005? In: Peters et al. [365], pp. 1–10
349. Peters, C.: What Happened in CLEF 2006? In: Peters et al. [357], pp. 1–10
350. Peters, C.: What Happened in CLEF 2007? In: Peters et al. [366], pp. 1–12
351. Peters, C.: What Happened in CLEF 2008? In: Peters et al. [359], pp. 1–14
352. Peters, C.: What Happened in CLEF 2009? In: Peters et al. [360], pp. 1–12
353. Peters, C., Braschler, M., Clough, P.: Multilingual Information Retrieval. Springer-Verlag, Heidelberg, Germany (2011)

354. Peters, C., Braschler, M., Gonzalo, J., Kluck, M. (eds.): Evaluation of Cross-Language Information Retrieval Systems: Second Workshop of the Cross-Language Evaluation Forum (CLEF 2001) Revised Papers. Lecture Notes in Computer Science (LNCS) 2406, Springer, Heidelberg, Germany (2002)
355. Peters, C., Braschler, M., Gonzalo, J., Kluck, M. (eds.): Advances in Cross-Language Information Retrieval: Third Workshop of the Cross-Language Evaluation Forum (CLEF 2002) Revised Papers. Lecture Notes in Computer Science (LNCS) 2785, Springer, Heidelberg, Germany (2003)
356. Peters, C., Braschler, M., Gonzalo, J., Kluck, M. (eds.): Comparative Evaluation of Multilingual Information Access Systems: Fourth Workshop of the Cross-Language Evaluation Forum (CLEF 2003) Revised Selected Papers. Lecture Notes in Computer Science (LNCS) 3237, Springer, Heidelberg, Germany (2004)
357. Peters, C., Clough, P., Gey, F.C., Karlgren, J., Magnini, B., Oard, D.W., de Rijke, M., Stempfhuber, M. (eds.): Evaluation of Multilingual and Multi-modal Information Retrieval : Seventh Workshop of the Cross-Language Evaluation Forum (CLEF 2006). Revised Selected Papers. Lecture Notes in Computer Science (LNCS) 4730, Springer, Heidelberg, Germany (2007)
358. Peters, C., Clough, P., Gonzalo, J., Jones, G.J.F., Kluck, M., Magnini, B. (eds.): Multilingual Information Access for Text, Speech and Images: Fifth Workshop of the Cross-Language Evaluation Forum (CLEF 2004) Revised Selected Papers. Lecture Notes in Computer Science (LNCS) 3491, Springer, Heidelberg, Germany (2005)
359. Peters, C., Deselaers, T., Ferro, N., Gonzalo, J., Jones, G.J.F., Kurimo, M., Mandl, T., Peñas, A. (eds.): Evaluating Systems for Multilingual and Multi-modal Information Access: Ninth Workshop of the Cross-Language Evaluation Forum (CLEF 2008). Revised Selected Papers. Lecture Notes in Computer Science (LNCS) 5706, Springer, Heidelberg, Germany (2009)
360. Peters, C., Di Nunzio, G.M., Kurimo, M., Mandl, T., Mostefa, D., Peñas, A., Roda, G. (eds.): Multilingual Information Access Evaluation Vol. I Text Retrieval Experiments – Tenth Workshop of the Cross-Language Evaluation Forum (CLEF 2009). Revised Selected Papers. Lecture Notes in Computer Science (LNCS) 6241, Springer, Heidelberg, Germany (2010)
361. Peters, C., Ferro, N. (eds.): CLEF 2000 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-1166/> (2000)
362. Peters, C., Ferro, N. (eds.): CLEF 2001 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-1167/> (2001)
363. Peters, C., Ferro, N. (eds.): CLEF 2002 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-1168/> (2002)
364. Peters, C., Ferro, N. (eds.): CLEF 2003 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-1169/> (2003)
365. Peters, C., Gey, F.C., Gonzalo, J., Jones, G.J.F., Kluck, M., Magnini, B., Müller, H., de Rijke, M. (eds.): Accessing Multilingual Information Repositories: Sixth Workshop of the Cross-Language Evaluation Forum (CLEF 2005). Revised Selected Papers. Lecture Notes in Computer Science (LNCS) 4022, Springer, Heidelberg, Germany (2006)

366. Peters, C., Jijkoun, V., Mandl, T., Müller, H., Oard, D.W., Peñas, A., Petras, V., Santos, D. (eds.): Advances in Multilingual and Multimodal Information Retrieval: Eighth Workshop of the Cross-Language Evaluation Forum (CLEF 2007). Revised Selected Papers. Lecture Notes in Computer Science (LNCS) 5152, Springer, Heidelberg, Germany (2008)
367. Peters, C., Quochi, V., Ferro, N. (eds.): CLEF 2005 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-1171/> (2005)
368. Peters, C., Tsikrika, T., Müller, H., Kalpathy-Cramer, J., Jones, G.J.F., Gonzalo, J., Caputo, B. (eds.): Multilingual Information Access Evaluation Vol. II Multimedia Experiments – Tenth Workshop of the Cross-Language Evaluation Forum (CLEF 2009). Revised Selected Papers. Lecture Notes in Computer Science (LNCS), Springer, Heidelberg, Germany (2010)
369. Petras, V., Baerisch, S.: The Domain-Specific Track at CLEF 2008. In: Peters et al. [359], pp. 186–198
370. Petras, V., Baerisch, S., Stempfhuber, M.: The Domain-Specific Track at CLEF 2007. In: Peters et al. [366], pp. 160–173
371. Petras, V., Bogers, T., Hall, M., Savoy, J., Malak, P., Pawlowski, A., Ferro, N., Masiero, I.: Cultural Heritage in CLEF (CHiC) 2013. In: Forner et al. [138], pp. 192–211
372. Petras, V., Ferro, N., Gäde, M., Isaac, A., Kleineberg, M., Masiero, I., Nicchio, M., Stiller, J.: Cultural Heritage in CLEF (CHiC) Overview 2012. In: Forner et al. [137]
373. Petras, V., Forner, P., Clough, P., Ferro, N. (eds.): CLEF 2011 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-1177/> (2011)
374. Piras, L., Caputo, B., Dang-Nguyen, D.T., Riegler, M., Halvorsen, P.: Image Retrieval Evaluation in Specific Domains. In: Ferro and Peters [130]
375. Piroi, F.: CLEF-IP 2010: Retrieval Experiments in the Intellectual Property Domain. In: Braschler et al. [61]
376. Piroi, F., Hanbury, A.: Multilingual Patent Text Retrieval Evaluation: CLEF-IP. In: Ferro and Peters [130]
377. Piroi, F., Lupu, M., Hanbury, A.: Effects of Language and Topic Size in Patent IR: An Empirical Study. In: Catarci et al. [75], pp. 54–66
378. Piroi, F., Lupu, M., Hanbury, A.: Overview of CLEF-IP 2013 Lab - Information Retrieval in the Patent Domain. In: Forner et al. [138], pp. 232–249
379. Piroi, F., Lupu, M., Hanbury, A., Sexton, A.P., Magdy, W., Filippov, I.V.: CLEF-IP 2012: Retrieval Experiments in the Intellectual Property Domain. In: Forner et al. [137]
380. Piroi, F., Lupu, M., Hanbury, A., Zenz, V.: CLEF-IP 2011: Retrieval in the Intellectual Property Domain. In: Petras et al. [373]
381. Popescu, A., Tsikrika, T., Kludas, J.: Overview of the Wikipedia Retrieval Task at ImageCLEF 2010. In: Braschler et al. [61]
382. Potha, N., Stamatatos, E.: An Improved Impostors Method for Authorship Verification. In: Jones et al. [209], pp. 138–144
383. Potthast, M., Barrón-Cedeño, A., Eiselt, A., Stein, B., Rosso, P.: Overview of the 2nd International Competition on Plagiarism Detection. In: Braschler et al. [61]
384. Potthast, M., Eiselt, A., Barrón-Cedeño, A., Stein, B., Rosso, P.: Overview of the 3rd International Competition on Plagiarism Detection. In: Petras et al. [373]

385. Potthast, M., Gollub, T., Hagen, M., Kiesel, J., Michel, M., Oberländer, A., Tippmann, M., Barrón-Cedeño, A., Gupta, P., Rosso, P., Stein, B.: Overview of the 4th International Competition on Plagiarism Detection. In: Forner et al. [137]
386. Potthast, M., Gollub, T., Rangel Pardo, F., Rosso, P., Stamatatos, E., Stein, B.: Improving the Reproducibility of PAN's Shared Tasks: Plagiarism Detection, Author Identification, and Author Profiling. In: Kanoulas et al. [216], pp. 268–299
387. Potthast, M., Gollub, T., Wiegmann, M., Stein, B.: TIRA Integrated Research Architecture. In: Ferro and Peters [130]
388. Potthast, M., Holfeld, T.: Overview of the 2nd International Competition on Wikipedia Vandalism Detection. In: Petras et al. [373]
389. Potthast, M., Rangel Pardo, F.M., Tschuggnall, M., Stamatatos, E., Rosso, P., Stein, B.: Overview of PAN'17 - Author Identification, Author Profiling, and Author Obfuscation. In: Jones et al. [209], pp. 275–290
390. Potthast, M., Stein, B., Holfeld, T.: Overview of the 1st International Competition on Wikipedia Vandalism Detection. In: Braschler et al. [61]
391. Pritsos, D.A., Stamatatos, E.: The Impact of Noise in Web Genre Identification. In: Mothe et al. [301], pp. 268–273
392. Pronobis, A., Fornoni, M., Christensen, H.I., Caputo, B.: The Robot Vision Track at ImageCLEF 2010. In: Braschler et al. [61]
393. Pronobis, A., Xing, L., Caputo, B.: Overview of the CLEF 2009 Robot Vision Track. In: Peters et al. [368], pp. 110–119
394. Raghavi, K.C., Chinnakotla, M.K., Black, A.W., Shrivastava, M.: WebShodh: A Code Mixed Factoid Question Answering System for Web. In: Jones et al. [209], pp. 104–111
395. Ragheb, W., Azé, J., Bringay, S., Servajean, M.: Language Modeling in Temporal Mood Variation Models for Early Risk Detection on The Internet. In: Crestani et al. [94]
396. Rangel Pardo, F., Rosso, P.: On the Multilingual and Genre Robustness of EmoGraphs for Author Profiling in Social Media. In: Mothe et al. [301], pp. 274–280
397. Rebholz-Schuhmann, D., Clematide, S., Rinaldi, F., Kafkas, S., van Mulligen, E.M., Bui, Q.C., Hellrich, J., Lewin, J., Milward, D., Poprat, M., Jimeno-Yepes, A., Hahn, U., Kors, J.A.: Entity Recognition in Parallel Multi-lingual Biomedical Corpora: The CLEF-ER Laboratory Overview. In: Forner et al. [138], pp. 353–367
398. Rekabsaz, N., Lupu, M.: A Real-World Framework for Translator as Expert Retrieval. In: Kanoulas et al. [216], pp. 141–152
399. de Rijke, M., Balog, K., Bogers, T., van den Bosch, A.: On the Evaluation of Entity Profiles. In: Agosti et al. [9], pp. 94–99
400. Roda, G., Tait, J., Piroi, F., Zenz, V.: CLEF-IP 2009: Retrieval Experiments in the Intellectual Property Domain. In: Peters et al. [360], pp. 385–409
401. Rodrigo, A., Peñas, A., Verdejo, M.F.: Overview of the Answer Validation Exercise 2008. In: Peters et al. [359], pp. 296–313
402. Roller, R., Stevenson, M.: Self-supervised Relation Extraction Using UMLS. In: Kanoulas et al. [216], pp. 116–127
403. Rosso, P., Potthast, M., Stein, B., Stamatatos, E., Rangel Pardo, F.M., Daelemans, W.: Evolution of the PAN Lab on Digital Text Forensics. In: Ferro and Peters [130]
404. Rosso, P., Rangel, F., Potthast, M., Stamatatos, E., Tschuggnall, M., Stein, B.: Overview of PAN'16. In: Fuhr et al. [143], pp. 332–350

405. Rowe, B.R., Wood, D.W., Link, A.L., Simoni, D.A.: Economic Impact Assessment of NIST's Text REtrieval Conference (TREC) Program. RTI Project Number 0211875, RTI International, USA. <http://trec.nist.gov/pubs/2010.economic.impact.pdf> (July 2010)
406. Sabetghadam, S., Bierig, R., Rauber, A.: A Hybrid Approach for Multi-faceted IR in Multimodal Domain. In: Kanoulas et al. [216], pp. 86–97
407. Sakai, T.: How to Run an Evaluation Task. In: Ferro and Peters [130]
408. Saleh, S., Pecina, P.: Reranking Hypotheses of Machine-Translated Queries for Cross-Lingual Information Retrieval. In: Fuhr et al. [143], pp. 54–68
409. Samuel, J.: Analyzing and Visualizing Translation Patterns of Wikidata Properties. In: Bellot et al. [42], pp. 128–134
410. Sánchez-Junquera, J., Villaseñor-Pineda, L., Montes-y Gómez, M., Rosso, P.: Character N-Grams for Detecting Deceptive Controversial Opinions. In: Bellot et al. [42], pp. 135–140
411. Sanchez-Perez, M.A., Gelbukh, A.F., Sidorov, G.: Adaptive Algorithm for Plagiarism Detection: The Best-Performing Approach at PAN 2014 Text Alignment Competition. In: Mothe et al. [301], pp. 402–413
412. Sanchez-Perez, M.A., Markov, I., Gómez-Adorno, H., Sidorov, G.: Comparison of Character n-grams and Lexical Features on Author, Gender, and Language Variety Identification on the Same Spanish News Corpus. In: Jones et al. [209], pp. 145–151
413. SanJuan, E., Moriceau, V., Tannier, X., Bellot, P., Mothe, J.: Overview of the INEX 2012 Tweet Contextualization Track. In: Forner et al. [137]
414. Santos, D., Cabral, L.M.: GikiCLEF: Expectations and Lessons Learned. In: Peters et al. [360], pp. 212–222
415. Ramos dos Santos, W., Paraboni, I.: Personality facets recognition from text. In: Crestani et al. [94]
416. Savenkov, D., Braslavski, P., Lebedev, M.: Search Snippet Evaluation at Yandex: Lessons Learned and Future Directions. In: Forner et al. [136], pp. 14–25
417. Savoy, J., Braschler, M.: Lessons Learnt from Experiments on the Ad-Hoc Multilingual Test Collections at CLEF. In: Ferro and Peters [130]
418. Schaer, P.: Better than Their Reputation? On the Reliability of Relevance Assessments with Students. In: Catarci et al. [75], pp. 124–135
419. Schaer, P., Mayr, P., Sünkler, S., Lewandowski, D.: How Relevant is the Long Tail?: A Relevance Assessment Study on Million Short. In: Fuhr et al. [143], pp. 227–233
420. Schaer, P., Neumann, M.: Enriching Existing Test Collections with OXPath. In: Jones et al. [209], pp. 152–158
421. Schauble, P., Sheridan, P.: Cross-Language Information Retrieval (CLIR) Track Overview. In: Voorhees, E.M., Harman, D.K. (eds.) *The Sixth Text REtrieval Conference (TREC-6)*. pp. 31–44. National Institute of Standards and Technology (NIST), Special Publication 500-240, Washington, USA (1997)
422. Schubotz, M., Krämer, L., Meuschke, N., Hamborg, F., Gipp, B.: Evaluating and Improving the Extraction of Mathematical Identifier Definitions. In: Jones et al. [209], pp. 82–94
423. Schuth, A., Balog, K., Kelly, L.: Overview of the Living Labs for Information Retrieval Evaluation (LL4IR) CLEF Lab 2015. In: Mothe et al. [301], pp. 484–496
424. Schuth, A., Marx, M.: Evaluation Methods for Rankings of Facetvalues for Faceted Search. In: Forner et al. [136], pp. 131–136

425. Shen, W., Nie, J.Y.: Is Concept Mapping Useful for Biomedical Information Retrieval. In: Mothe et al. [301], pp. 281–286
426. Shepeleva, N., Balog, K.: Towards an Understanding of Transactional Tasks. In: Fuhr et al. [143], pp. 234–240
427. Sherkat, E., Velcin, J., Milios, E.E.: Fast and Simple Deterministic Seeding of KMeans for Text Document Clustering. In: Bellot et al. [42], pp. 76–88
428. Shing, H.S., Barrow, J., Galuščáková, P., Oard, D.W., Resnik, P.: Unsupervised System Combination for Set-based Retrieval with Expectation Maximization. In: Crestani et al. [94]
429. Sierek, T., Hanbury, A.: Using Health Statistics to Improve Medical and Health Search. In: Mothe et al. [301], pp. 287–292
430. Sigurbjörnsson, B., Kamps, J., de Rijke, M.: Overview of WebCLEF 2005. In: Peters et al. [365], pp. 810–824
431. Silvello, G., Bordea, G., Ferro, N., Buitelaar, P., Bogers, T.: Semantic Representation and Enrichment of Information Retrieval Experimental Data. *International Journal on Digital Libraries (IJDL)* 18(2), 145–172 (June 2017)
432. Singh, G., Mantrach, A., Silvestri, F.: Improving Profiles of Weakly-Engaged Users: With Applications to Recommender Systems. In: Fuhr et al. [143], pp. 123–136
433. Sorg, P., Cimiano, P., Schultz, A., Sizov, S.: Overview of the Cross-lingual Expert Search (CriES) Pilot Challenge. In: Braschler et al. [61]
434. Spina, D., Amigó, E., Gonzalo, J.: Filter Keywords and Majority Class Strategies for Company Name Disambiguation in Twitter. In: Forner et al. [136], pp. 38–49
435. Stamatatos, E., Potthast, M., Rangel Pardo, F.M., Rosso, P., Stein, B.: Overview of the PAN/CLEF 2015 Evaluation Lab. In: Mothe et al. [301], pp. 518–538
436. Stamatatos, E., Rangel Pardo, F., Tschuggnall, M., Stein, B., Kestemont, M., Rosso, P., Potthast, M.: Overview of PAN 2018. In: Bellot et al. [42], pp. 267–285
437. Stathopoulos, S., Kalamboukis, T.: Medical Image Classification with Weighted Latent Semantic Tensors and Deep Convolutional Neural Networks. In: Bellot et al. [42], pp. 89–100
438. Stefanov, V., Sachs, A., Kritz, M., Samwald, M., Gschwandtner, M., Hanbury, A.: A Formative Evaluation of a Comprehensive Search System for Medical Professionals. In: Forner et al. [138], pp. 81–92
439. Stempfhuber, M., Baerisch, S.: The Domain-Specific Track at CLEF 2006: Overview of Approaches, Results and Assessment. In: Peters et al. [357], pp. 163–169
440. Suchomel, S., Brandeys, M.: Determining Window Size from Plagiarism Corpus for Stylometric Features. In: Mothe et al. [301], pp. 293–299
441. Suominen, H.: CLEFeHealth2012 - The CLEF 2012 Workshop on Cross-Language Evaluation of Methods, Applications, and Resources for eHealth Document Analysis. In: Forner et al. [137]
442. Suominen, H., Kelly, L., Goeuriot, L.: The Scholarly Impact and Strategic Intent of CLEF eHealth Labs from 2012–2017. In: Ferro and Peters [130]
443. Suominen, H., Kelly, L., Goeuriot, L., Névél, A., Ramadier, L., Robert, A., Kanoulas, E., Spijker, R., Azzopardi, L., Li, D., Jimmy, Palotti, J., Zuccon, G.: Overview of the CLEF eHealth Evaluation Lab 2018. In: Bellot et al. [42], pp. 286–301
444. Suominen, H., Kelly, L., Goeuriot, L., Neves, M., Kanoulas, E., Li, D., Azzopardi, L., Spijker, R.: Overview of the CLEF eHealth Evaluation Lab 2019. In: Crestani et al. [94]

445. Suominen, H., Salanterä, S., Velupillai, S., Webber Chapman, W., Savova, G.K., Elhadad, N., Pradhan, S., South, B.R., Mowery, D.L., Jones, G.J.F., Leveling, J., Kelly, L., Goeuriot, L., Martínez, D., Zuccon, G.: Overview of the ShARe/CLEF eHealth Evaluation Lab 2013. In: Forner et al. [138], pp. 212–231
446. Tannebaum, W., Mahdabi, P., Rauber, A.: Effect of Log-Based Query Term Expansion on Retrieval Effectiveness in Patent Searching. In: Mothe et al. [301], pp. 300–305
447. Tannebaum, W., Rauber, A.: Mining Query Logs of USPTO Patent Examiners. In: Forner et al. [138], pp. 136–142
448. Teixeira Lopes, C., Almeida Fernandes, T.: Health Suggestions: A Chrome Extension to Help Laypersons Search for Health Information. In: Fuhr et al. [143], pp. 241–246
449. Teixeira Lopes, C., Ribeiro, C.: Effects of Language and Terminology on the Usage of Health Query Suggestions. In: Fuhr et al. [143], pp. 83–95
450. Teixeira Lopes, C., Ribeiro, C.: Effects of Language and Terminology of Query Suggestions on the Precision of Health Searches. In: Bellot et al. [42], pp. 101–111
451. Thomee, B., Popescu, A.: Overview of the ImageCLEF 2012 Flickr Photo Annotation and Retrieval Task. In: Forner et al. [137]
452. Thornley, C.V., Johnson, A.C., Smeaton, A.F., Lee, H.: The Scholarly Impact of TRECVID (2003–2009). *Journal of the American Society for Information Science and Technology (JASIST)* 62(4), 613–627 (April 2011)
453. Tommasi, T., Caputo, B., Welter, P., Güld, M.O., Deserno, T.M.: Overview of the CLEF 2009 Medical Image Annotation Track. In: Peters et al. [368], pp. 85–93
454. Trappett, M., Geva, S., Trotman, A., Scholer, F., Sanderson, M.: Overview of the INEX 2012 Snippet Retrieval Track. In: Forner et al. [137]
455. Trotszek, M., Koitka, S., Friedrich, C.M.: Early Detection of Depression Based on Linguistic Metadata Augmented Classifiers Revisited. In: Bellot et al. [42], pp. 191–202
456. Tsikrika, T., Garcia Seco de Herrera, A., Müller, H.: Assessing the Scholarly Impact of ImageCLEF. In: Forner et al. [136], pp. 95–106
457. Tsikrika, T., Kludas, J.: Overview of the WikipediaMM Task at ImageCLEF 2008. In: Peters et al. [359], pp. 539–550
458. Tsikrika, T., Kludas, J.: Overview of the WikipediaMM Task at ImageCLEF 2009. In: Peters et al. [368], pp. 60–71
459. Tsikrika, T., Larsen, B., Müller, H., Endrullis, S., Rahm, E.: The Scholarly Impact of CLEF (2000–2009). In: Forner et al. [138], pp. 1–12
460. Tsikrika, T., Popescu, A., Kludas, J.: Overview of the Wikipedia Image Retrieval Task at ImageCLEF 2011. In: Petras et al. [373]
461. Turchi, M., Steinberger, J., Alexandrov Kabadjov, M., Steinberger, R.: Using Parallel Corpora for Multilingual (Multi-document) Summarisation Evaluation. In: Agosti et al. [9], pp. 52–63
462. Turmo, J., Comas, P., Rosset, S., Galibert, O., Moreau, N., Mostefa, D., Rosso, P., Buscaldi, D.: Overview of QAST 2009. In: Peters et al. [360], pp. 197–211
463. Turmo, J., Comas, P., Rosset, S., Lamel, L., Moreau, N., Mostefa, D.: Overview of QAST 2008. In: Peters et al. [359], pp. 296–313
464. Vallin, A., Magnini, B., Giampiccolo, D., Aunimo, L., Ayache, C., Osenova, P., Peñas, A., de Rijke, M., Sacaleanu, B., Santos, D., Sutcliffe, R.F.E.: Overview of the CLEF 2005 Multilingual Question Answering Track. In: Peters et al. [365], pp. 307–331

465. Valverde-Albacete, F., Carrillo de Albornoz, J., Peláez-Moreno, C.: A Proposal for New Evaluation Metrics and Result Visualization Technique for Sentiment Analysis Tasks. In: Forner et al. [138], pp. 41–42
466. Villegas, M., Müller, H., Gilbert, A., Piras, L., Wang, J., Mikolajczyk, K., Garcia Seco de Herrera, A., Bromuri, S., Amin, M.A., Kazi Mohammed, M., Acar, B., Üsküdarlı, S., Barzegar Marvasti, N., Aldana Montes, J.F., del Mar Roldán García, M.: General Overview of ImageCLEF at the CLEF 2015 Labs. In: Mothe et al. [301], pp. 444–461
467. Villegas, M., Müller, H., Garcia Seco de Herrera, A., Schaer, R., Bromuri, S., Gilbert, A., Piras, L., Wang, J., Yan, F., Ramisa, A., Dellandrea, E., Gaizauskas, R., Mikolajczyk, K., Puigcerver, K., Toselli, A.H., Sánchez, J.A., Vidal, E.: General Overview of ImageCLEF at the CLEF 2016 Labs. In: Fuhr et al. [143], pp. 267–285
468. Villegas, M., Paredes, R.: Overview of the ImageCLEF 2012 Scalable Web Image Annotation Task. In: Forner et al. [137]
469. Voorhees, E.M.: TREC: Continuing Information Retrieval's Tradition of Experimentation. *Communications of the ACM (CACM)* 50(11), 51–54 (November 2007)
470. Voorhees, E.M.: The Evolution of Cranfield. In: Ferro and Peters [130]
471. Wakeling, S., Clough, P.: Integrating Mixed-Methods for Evaluating Information Access Systems. In: Mothe et al. [301], pp. 306–311
472. Walker, A., Starkey, A., Pan, J.Z., Siddharthan, A.: Making Test Corpora for Question Answering More Representative. In: Kanoulas et al. [216], pp. 1–6
473. Wang, J., Gilbert, A., Thomee, B., Villegas, M.: Automatic Image Annotation at ImageCLEF. In: Ferro and Peters [130]
474. Wang, Q., Kamps, J., Ramírez Camps, G., Marx, M., Schuth, A., Theobald, M., Gurajada, S., Mishra, A.: Overview of the INEX 2012 Linked Data Track. In: Forner et al. [137]
475. Wang, X., Guo, Z., Zhang, Y., Li, J.: Medical Image Labeling and Semantic Understanding for Clinical Applications. In: Crestani et al. [94]
476. Wang, X., X., W., Zhang, Q.: A Web-Based CLIR System with Cross-Lingual Topical Pseudo Relevance Feedback. In: Forner et al. [138], pp. 104–107
477. Weitzel, L., Bernardini, F., Quaresma, P., Alves, C.A., Zacharski, W., de Figueiredo, L.G.: Brazilian Social Mood: The Political Dimension of Emotion. In: Fuhr et al. [143], pp. 247–255
478. White, R.W., Oard, D.W., Jones, G.J.F., Soergel, D., Huang, X.: Overview of the CLEF-2005 Cross-Language Speech Retrieval Track. In: Peters et al. [365], pp. 744–759
479. Wilhelm-Stein, T., Eibl, M.: A Quantitative Look at the CLEF Working Notes. In: Forner et al. [138], pp. 13–16
480. Wilhelm-Stein, T., Eibl, M.: Teaching the IR Process Using Real Experiments Supported by Game Mechanics. In: Mothe et al. [301], pp. 312–317
481. Wilhelm-Stein, T., Herms, R., Ritter, M., Eibl, M.: Improving Transcript-Based Video Retrieval Using Unsupervised Language Model Adaptation. In: Kanoulas et al. [216], pp. 110–115
482. Xu, K., Feng, Y., Huang, S., Zhao, D.: Question Answering via Phrasal Semantic Parsing. In: Mothe et al. [301], pp. 414–426
483. Yan, X., Gao, G., Su, X., Wei, H., Zhang, X., Lu, Q.: Hidden Markov Model for Term Weighting in Verbose Queries. In: Catarci et al. [75], pp. 82–87
484. Yang, H., Gonçalves, T.: A Compound Model for Consumer Health Search. In: Bellot et al. [42], pp. 231–236

485. Zamani, H., Esfahani, H.N., Babaie, P., Abnar, S., Dehghani, M., Shakery, A.: Authorship Identification Using Dynamic Selection of Features from Probabilistic Feature Set. In: Kanoulas et al. [216], pp. 128–140
486. Zellhöfer, D.: Overview of the Personal Photo Retrieval Pilot Task at ImageCLEF 2012. In: Forner et al. [137]
487. Zhang, L., Rettinger, A., Färber, M., Tadic, M.: A Comparative Evaluation of Cross-Lingual Text Annotation Techniques. In: Forner et al. [138], pp. 124–135
488. Ziak, H., Kern: Evaluation of Pseudo Relevance Feedback Techniques for Cross Vertical Aggregated Search. In: Mothe et al. [301], pp. 91–102
489. Zingla, M.A., Latiri, C., Slimani, Y.: Tweet Contextualization Using Association Rules Mining and DBpedia. In: Mothe et al. [301], pp. 318–323
490. Zlabinger, M., Rekabsaz, N., Zlabinger, S., Hanbury, A.: Efficient Answer-Annotation for Frequent Questions. In: Crestani et al. [94]
491. Zuo, C., Karakas, A., Banerjee, R.: To Check or not to Check: Syntax, Semantics, and Context in the Language of Check-worthy Claims. In: Crestani et al. [94]