

What Happened in CLEF... For Another While?

Nicola Ferro

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

Abstract. 2024 marks the 25th 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, a description of how it has evolved over the years, and its major achievements.

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)*¹ 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)*² [266], 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)*³ [543], 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)*⁴, whose aim is to encourage research in Indian languages by creating a platform similar to CLEF, providing

¹ <http://www.clef-initiative.eu/>

² <http://trec.nist.gov/>

³ <http://research.nii.ac.jp/ntcir/>

⁴ <http://fire.irsi.res.in/>

100% Voluntary Effort Based

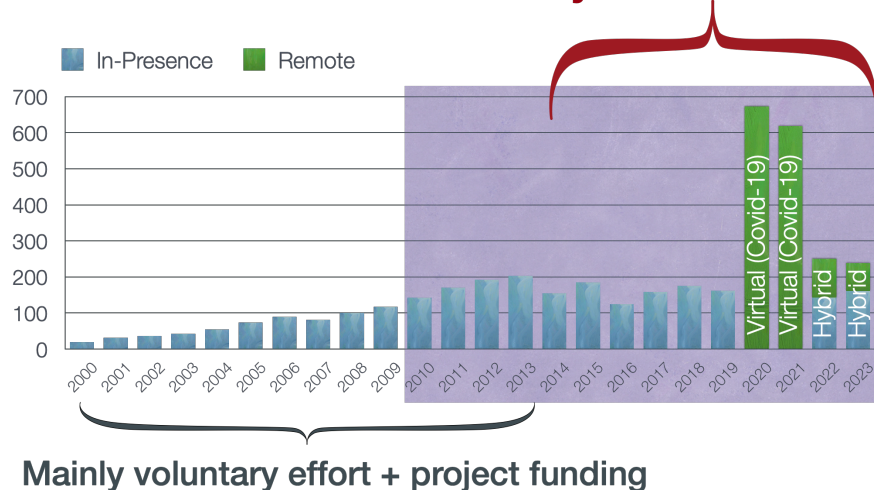


Fig. 1: Participation in CLEF over the years (CLEF “Classic” period un-shaded; CLEF Initiative period shaded).

data and a common forum for comparing models and techniques applied to these languages.

This year marks the 25th 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.

CLEF can be divided into a CLEF “Classic” (2000–2009) period, where CLEF began and was supported by the European funding, and a “CLEF Initiative” period (2010–2024), where CLEF was profoundly reorganized and became a self-sustained activity not backed by dedicated European funding, thanks to the contribution of the PROMISE network of excellence [94].

Figure 1 shows the attendance to the CLEF event over the years: we can observe a stable and consistent growth in participation, 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. We can also see that in 2020 and 2021 CLEF has gone completely virtual due to the Covid-19 pandemic and this caused a spike in attendance. CLEF 2022 and 2023 represent the first editions back in presence after Covid-19 which offered a hybrid modality of participation: we can observe as the in-presence attendance is back to the pre-Covid levels and how it is complemented by a substantial remote attendance as well.

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 impact of CLEF in the IR community and beyond; Section 7 summarizes the book which was prepared for celebrating the past 20th anniversary of CLEF; finally, Section 8 presents the CLEF Association, the no-profit legal entity committed to sustaining and running CLEF.

2 CLEF “Classic”: 2000–2009

The *Cross-Language Evaluation Forum* (CLEF) began as a cross-lingual track at TREC in 1997 [563], moving to an independent activity in 2000 [480], 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* [275]. 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 [481], 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;

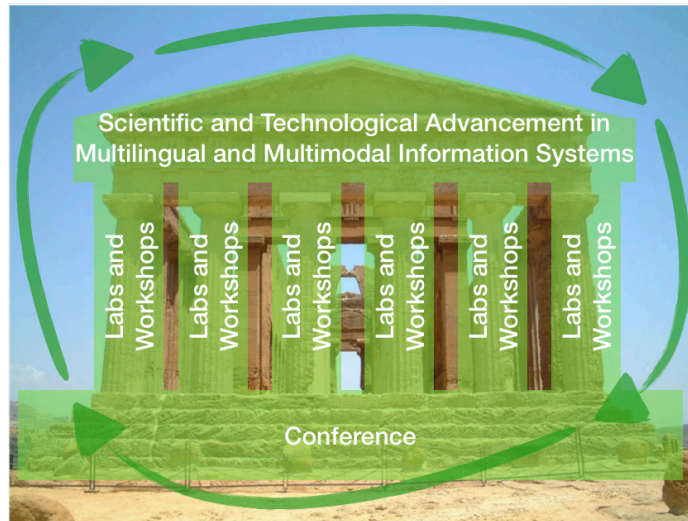


Fig. 2: Structure of the CLEF Initiative.

- 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 TPDL – 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 2, 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 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*⁵ are published in the Springer Lecture Notes in Computer Science (LNCS) [10, 36, 37, 55, 64, 104, 117, 138, 209, 211, 219, 240, 241, 305, 312, 420, 480, 483–489, 494, 495, 497] 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)⁶ series [18, 51, 87–89, 96, 106–111, 186–188, 210, 212, 435, 436, 490–493, 496, 502] 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 peer-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 2027.

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

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

⁶ <http://ceur-ws.org/>

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.

3.4 Beyond CLEF

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. In occasion of the past 20th anniversary of CLEF in 2019, 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. Since then, ECIR kept hosting a CLEF session and similar plans are already in place for CLEF 2025 and ECIR 2025.

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.

4 The Conference

Figure 3 gives an overview of the topics addressed by the CLEF conference over the years, together with the number of papers published for each topic. Figure 3 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, Ranking, 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”. Figure 3 shows how the *evaluation* and *multilinguality and multimodality* streams attract a good number of papers across all the editions of CLEF. On the

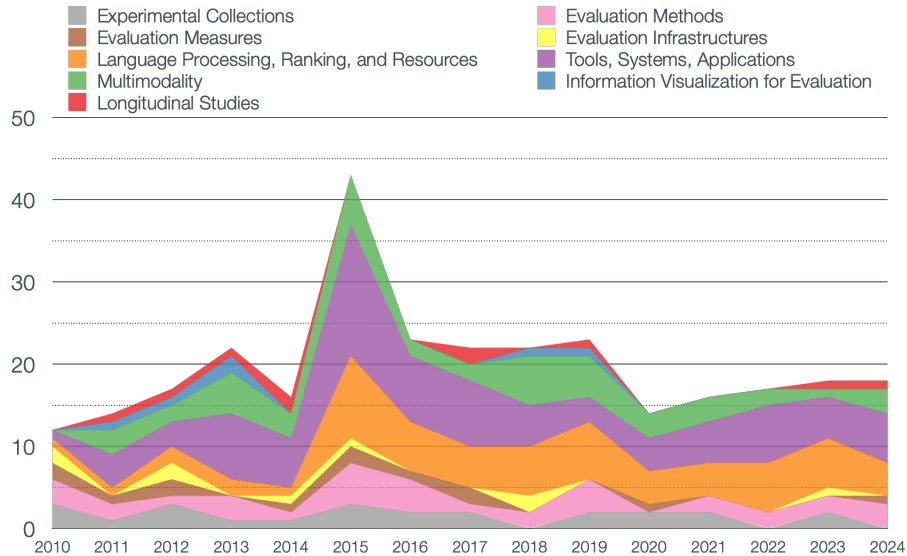


Fig. 3: Topics addressed by the CLEF conference over the years and number of papers published for each topic.

other hand, “Information Visualization for Evaluation” attracted a few papers for about ten years (2010–2019) but it has faded away in the last few years. Finally, “Longitudinal Studies” keep appearing at CLEF, even if in a discontinuous way but this is understandable when you consider the additional effort required to conduct a longitudinal study.

Since CLEF 2023 the conference has introduced a new *result-less review process*, inspired by the “Dagstuhl Seminar 23031 on Frontiers of Information Access Experimentation for Research and Education” [60,61], where the reviewing process is split in two parts. Firstly, the papers are reviewed by their methodological contribution, their research questions, and their experimental design – the submitted papers do not contain an experimentation part. Secondly, those papers which pass the first step are then reviewed by their experimentation, analyses, and insights. The purpose of this new review process is: (i) to avoid accepting papers just because performance improvements with respect to some baseline; (ii) to ensure they have a grounded methodology; and, (iii) to verify that the research questions are driven by the methodology and not, post-hoc, by the experimental results.

We briefly summarize below the topics of Figure 3, covered by the conference over the years, with pointers to the main references:

Experimental Collections explored different issues concerning experimental collections such as: the creation of collections for languages such as Persian, Arabic, Ahmaric; 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; reading comprehension; subjectivity in news articles; spoken Portuguese document corpus; datasets for automatic gene variant interpretation; [17, 34, 52, 66, 67, 167, 221, 223, 226, 232, 271, 278, 319, 362, 385, 387, 406, 417, 431, 507, 561, 609, 621, 636];

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 translation 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; longitudinal evaluation of IR systems and prediction of their performance; multidimensional relevance; noise reduction in relevance judgements: benchmarking methodologies for information extraction from documents; evaluation methods for knowledge acquisition; user simulation via generative AI; de-noising benchmarks [46, 49, 65, 105, 134, 170, 182, 206, 217, 245, 259, 277, 314, 318, 337, 341, 375, 413, 416, 457–459, 524, 534, 555, 558, 560, 570, 578, 620, 628, 630, 641, 645];

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; measures for merging multiple assessors; measures for reproducibility and replicability [28, 31, 47, 103, 193, 194, 222, 252, 323, 326, 393, 419, 464];

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; evaluation infrastructures for quantum computing for information access [8, 59, 158, 262, 263, 270, 380, 468];

Ranking, Language Processing, and Resources continued the CLEF interest in multilinguality by dealing with tools, algorithm, and resources for multiple languages such as: lemmatizers, decompounders and normalizers for underrepresented resources using statistical approaches; statistical stemmers; named entity extraction, linking and clustering in cross-lingual set-

tings; 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, fake news and mis-information; gender and bot identification argument retrieval; pseudo-relevance feedback; medical concept normalization; text analysis for authorship attribution; hate speech detection; query rewriting for health search; irony detection; entity representation using transformers; large language models [1, 35, 48, 77, 86, 99, 101, 116, 121, 132, 133, 135, 137, 139, 140, 150, 159, 166, 183, 192, 214, 216, 225, 230, 256, 264, 274, 313, 353, 356, 361, 365, 366, 379, 381, 386, 411, 412, 415, 443, 456, 471, 503, 525, 528, 529, 544, 545, 548, 550, 553, 557, 568, 571, 587, 596, 597, 603, 612, 626, 633, 646];

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; event detection; stock market prediction; skill extraction for job; tracking news stories in short messages search; studies on exposure of children to search technology; sensitive information classification; dataset recommendation; analysis of crime-related time series; entity linking; gambling disorders detection: prompt engineering; political bias in media; sexism identification; quality of Web search results for decision making; document sanitization; news recommendation [3, 4, 13–16, 19, 22, 42, 57, 58, 69, 70, 79, 80, 85, 115, 118, 131, 136, 142, 149, 151, 153, 162, 165, 172, 184, 185, 208, 215, 224, 227, 233, 234, 244, 251, 255, 260, 267, 288, 315, 327, 331, 335, 340, 345, 347, 351, 354, 363, 364, 372, 373, 376, 384, 404, 408–410, 414, 421, 444, 452, 514, 523, 530, 532, 533, 537, 547, 549, 554, 565, 569, 572, 573, 577, 579, 593, 595, 600, 624, 625, 631, 632, 634, 635, 637, 638, 641, 642, 644];

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; life-logging; voice question answering; image forgery; podcast retrieval; detecting eating disorders using text and images; image captioning; argument classification via images; sexism and hate speech identification via multiple media [2, 12, 23, 30, 38, 44, 68, 98, 100, 102, 114, 120, 122, 123, 125, 144, 145, 161, 226, 257, 258, 261, 268, 307, 316, 324, 325, 333, 334, 346, 355, 367, 371, 382, 392, 401, 442, 453, 454, 541, 564, 567, 581, 584, 585, 594, 643];

Information Visualization for Evaluation opened up a brand new area concerned with exploiting information visualization and visual analytics techniques 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 [33, 143, 152, 360, 546, 614];

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; reproducibility of technology assisted reviews [160, 163, 164, 195, 196, 205, 407, 445, 604, 607, 629].

5 Tracks and Labs

Figure 4 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 [5, 6, 90–93, 95, 154–156, 200];

Domain Specific Cross-Language IR (DS, 2000–2008) dealt with multilingual information retrieval on structured scientific data from the social sciences domain [91–93, 342–344, 498, 499, 586];

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 [246–250, 321, 449, 450];

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 [190, 191, 304, 451, 470, 627];

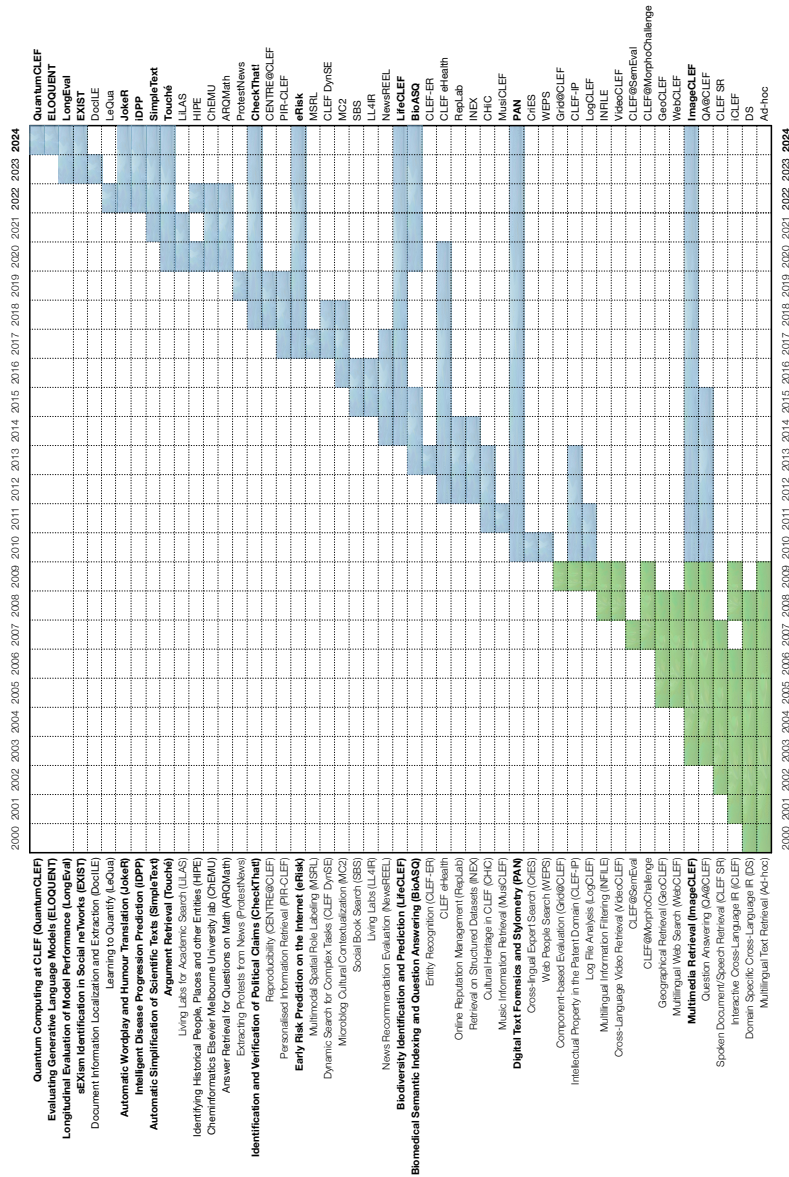


Fig. 4: Labs offered by CLEF over the years. CLEF “Classic” period in green; the CLEF Initiative period in blue; labs active in CLEF 2024 in bold.

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 [124, 213, 231, 394–396, 418, 472–476, 478, 479, 536, 552, 610, 611, 613];

Multimedia Retrieval (ImageCLEF, 2003–2024) studied the cross-language annotation and retrieval of images to support the advancement of the field of visual media analysis, indexing, classification, and retrieval [41, 112, 113, 126–129, 146–148, 235, 236, 253, 280–287, 308, 374, 405, 423, 424, 426–428, 430, 446–448, 513, 526, 527, 598, 601, 605, 606, 608, 615–617, 640];

Multilingual Web Search (WebCLEF, 2005–2008) addressed multilingual Web search, exploring different faces of navigational queries and known-item search [50, 289, 290, 574];

Geographical Retrieval (GeoCLEF, 2005–2008) evaluated cross-language *Geographic Information Retrieval (GIR)* against search tasks involving both spatial and multilingual aspects [228, 229, 398, 400];

CLEF@SemEval (2007) explored the impact of *Word Sense Disambiguation (WSD)* on multilingual information retrieval [7]; 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 [357–359];

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 [369, 370]; it went on to become the *MediaEval Benchmarking Initiative for Multimedia Evaluation*⁷ 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 [71, 72];

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 [157, 397, 399];

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 [505, 508–510, 535];

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 [198];

Web People Search (WEPS, 2010) focused on person name ambiguity and person attribute extraction on Web pages and on online reputation

⁷ <http://www.multimediaeval.org/>

management for organizations [26, 43]; 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 [580];

Digital Text Forensics and Stylometry (PAN, 2010–2024) studied plagiarism, authorship attribution, social software misuse, different types of profiling [29, 39, 45, 73–76, 141, 243, 279, 306, 515–518, 520–522, 539, 582, 583];

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 [455]; this activity has continued as part of MediaEval;

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 [220, 500, 501];

Retrieval on Structured Datasets (INEX, 2012–2014) was a stand-alone initiative pioneering structured and XML retrieval from 2002⁸; it joined forces with CLEF in 2012 to further promote the evaluation of focused retrieval by providing large test collections of structured documents [62, 63, 119, 350, 551, 602, 623];

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 [24, 25, 27];

CLEF eHealth (2012–2021) 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 [237, 238, 242, 328–330, 588, 589, 591, 592];

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 [531];

Biodiversity Identification and Prediction (LifeCLEF, 2014–2024) 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 [291–294, 296–302];

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 [273, 338, 339, 383]

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) [566];

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

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 [348,349].

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 [175,239].

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 [310,311].

Multimodal Spatial Role Labeling (MSRL, 2017) explored the extraction of spatial information from two information resources that is image and text, which is importa 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 [352].

Early Risk Prediction on the Internet (eRisk, 2017–2024) explored the evaluation methodology, effectiveness metrics and practical applications (particularly those related to health and safety) of early risk detection on the Internet [388–391,460–463].

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 [465–467].

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 [197,199].

Identification and Verification of Political Claims (CheckThat!, 2018–2024) 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 [53,54,56,171,432–434].

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 [276].

Biomedical Semantic Indexing and Question Answering (BioASQ, 2020–2024) aimed to push the research frontier towards systems that use the diverse and voluminous information available online to respond directly to the information needs of biomedical scientists [437–441].

- Answer Retrieval for Questions on Math (ARQMath, 2020–2022))** aimed to advance math-aware search and the semantic analysis of mathematical notation and texts [402, 403, 639].
- Cheminformatics Elsevier Melbourne University lab (ChEMU, 2020–2022))** aimed to provide a unique opportunity for the development of information extraction tools over chemical patents [269, 377, 378].
- Identifying Historical People, Places and other Entities (HIPE, 2020 and 2022)** aimed to promote named entity recognition and linking in historical documents, with the objective of assessing and advancing the development of robust, adaptable, and transferable named entity processing systems [168, 169].
- Living Labs for Academic Search (LiLAS, 2020–2021)** aimed to advance online evaluation of academic search systems by improving the search for academic resources like literature (ranging from short bibliographic records to full-text papers), research data, and the interlinking between these resources [559, 562].
- Argument Retrieval (Touché, 2020–2024)** aimed to establish a collaborative platform for researchers in the field of argument retrieval and to provide tools for developing and evaluating argument retrieval approaches [81–84, 336].
- Automatic Simplification of Scientific Texts (SimpleText, 2021–2024)** aimed to make science more open and accessible via automatic generation of simplified summaries of scientific documents [173, 178–180].
- Intelligent Disease Progression Prediction (iDPP, 2022–2024)** aimed to design and develop an evaluation infrastructure for AI algorithms able to: (1) better describe disease mechanisms of Amyotrophic Lateral Sclerosis (ALS) and Multiple Sclerosis (MS); (2) stratify patients according to their phenotype assessed all over the disease evolution; and (3) predict disease progression in a probabilistic, time dependent fashion [78, 189, 254].
- Automatic Wordplay and Humour Translation (JokeR, 2022–2024)** aimed to bring together translators and computer scientists to work on an evaluation framework for creative language, including data and metric development, and to foster work on automatic methods for wordplay translation [174, 176, 177].
- Learning to Quantify (LeQua, 2022)** aimed to allow the comparative evaluation of methods for “learning to quantify” in textual datasets; i.e. methods for training predictors of the relative frequencies of the classes of interest in sets of unlabelled textual documents [181].
- Document Information Localization and Extraction (DocILE, 2023)** aimed to benchmark Key Information Localization and Extraction (KILE) and Line Item Recognition (LIR) from business documents like invoices [576].
- sEXism Identification in Social neTworks (EXIST, 2023–2024)** aimed to capture and categorize sexism, from explicit misogyny to other subtle behaviors, in social networks [511, 512].

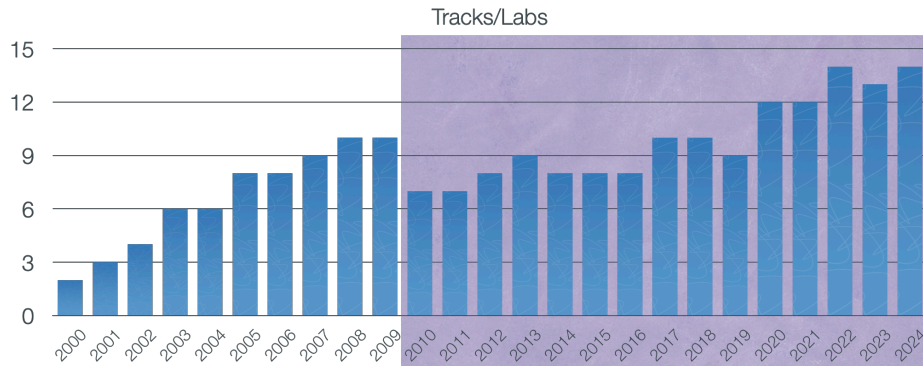


Fig. 5: Number of labs offered by CLEF over the years (CLEF “Classic” period un-shaded; the CLEF Initiative period shaded).

Longitudinal Evaluation of Model Performance (LongEval, 2023-2024)

aimed to evaluate the temporal persistence of information retrieval systems and text classifiers [20, 21].

Evaluating Generative Language Models (ELOQUENT, 2024)

aimed to evaluate the quality of generative language models under various angles like topical competence, veracity and hallucination, and robustness [320].

Quantum Computing at CLEF (QuantumCLEF, 2024)

aimed to advance the design and development of *Quantum Computing (QC)* algorithms and, in particular, for *Quantum Annealing (QA)* algorithms, for *IR* and *Recommender Systems (RS)* by providing access to real quantum computers [469].

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 cycle, typically three years, but also to introduce new activities every year. Also note that we put a cap on a maximum of 14 labs per edition, in order to avoid dispersion into too many activities. Finally, in the recent years, the lab selection process has become more and more competitive since CLEF is receiving around 20-25 lab proposals per year versus the 14 labs that can be accepted.

Figure 6 shows the number of papers published in the Working Notes papers, which represent the main output of the lab activities. We can observe how the number of papers published has a steady and consistent growth over the years, reaching quite a substantial number of papers in the last 2-3 years as a consequence of the increased participation in CLEF and the number of labs offered.

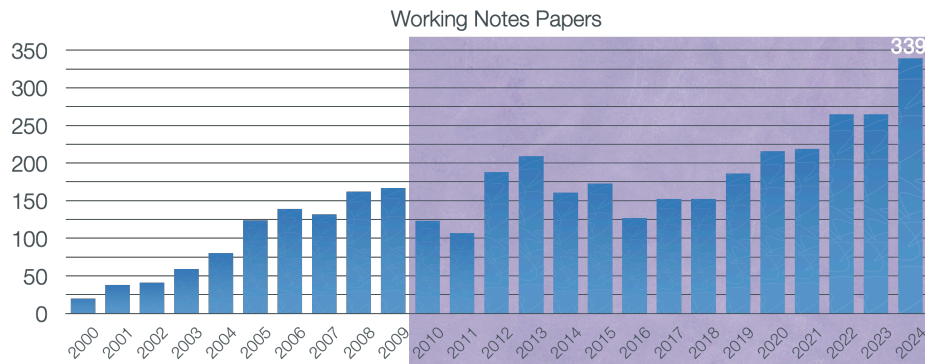


Fig. 6: Papers published in the CLEF Working Notes over the years (CLEF “Classic” period un-shaded; CLEF Initiative period shaded).

6 Impact of CLEF

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 [540]. 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 [599] and some steps in this direction have been performed for CLEF as well [32, 604, 607]. 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 [40, 332] and CLEF [205, 207].

Such rigorous studies are beyond the scope of the present report, here we concentrate on identifying rough indicators with respect to the maturity and liveness 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.

Figure 7 shows the trends of the h5-index (the largest number h such that at least h articles in that publication were cited at least h times each, only those of its articles that were published in the last five complete calendar years) and h5-median (the median number of citations for the articles that make up the h5-index), from 2016 to 2024 taken from Google Scholar Metrics⁹. We can observe a steady increase trend for both indicators, suggesting a positive scholarly impact for the research outcomes of the CLEF community. In Google Scholar Metrics 2024, CLEF achieved h5-index = 47, h5-median = 65; for a comparison: SIGIR h5-index = 103, h5-median = 149; CIKM h5-index = 91, h5-median = 133; ECIR h5-index = 42, h5-median = 60; TREC h5-index = 17, h5-median = 30.

Figure 8 reports the top-20 venues according to Google Scholar Metrics 2024 and CLEF is among them, as it happens since some years. In particular, Figure 8a reports the top-20 venues for the “Database and Information Systems” category¹⁰ while Figure 8b reports the top-20 venues according to a query¹¹ more targeted to IR and RS

```
"AIRS" OR "WWW" OR "information retrieval" OR
"Information and Knowledge Management" OR "SIGIR" OR
"information science and technology" OR "web search" OR
"TOIS" OR "information processing & Management" OR
"Transactions on Knowledge and Data Engineering" OR
"TWEB" OR "digital libraries" OR "cross language" OR
"recommender"
```

prepared by Mark Sanderson with further refinements suggested by Damiano Spina and Martin Tomko.

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 [265, 266, 543, 618] and CLEF [97, 422, 482], with a special publication which was prepared for the 20th anniversary of CLEF, as discussed in the next section.

⁹ Despite the expansion of the acronym for CLEF changed from *Cross-Language Evaluation Forum* to Conference and Labs of the Evaluation Forum in 2010, Google Scholar Metrics still indexes CLEF as “Cross-Language Evaluation Forum”. Maybe, this is also due to the fact that also Springer still calls CLEF “Cross-Language Evaluation Forum”; see <https://link.springer.com/conference/clef>.

¹⁰ https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=eng_databasesinformationsystems

¹¹ https://scholar.google.com/citations?hl=en&view_op=search_venues&vq=%22AIRS%22+OR+%22WWW%22+OR+%22information+retrieval%22+OR+%22Information+and+Knowledge+Management%22+OR+%22SIGIR%22+OR+%22information+science+and+technology%22+OR+%22web+search%22+OR+%22TOIS%22+OR+%22information+processing+%26+Management%22+OR+%22Transactions+on+Knowledge+and+Data+Engineering%22+OR+%22TWEB%22+OR+%22cross+language%22+OR+%22recommender%22+OR+%22digital+libraries%22

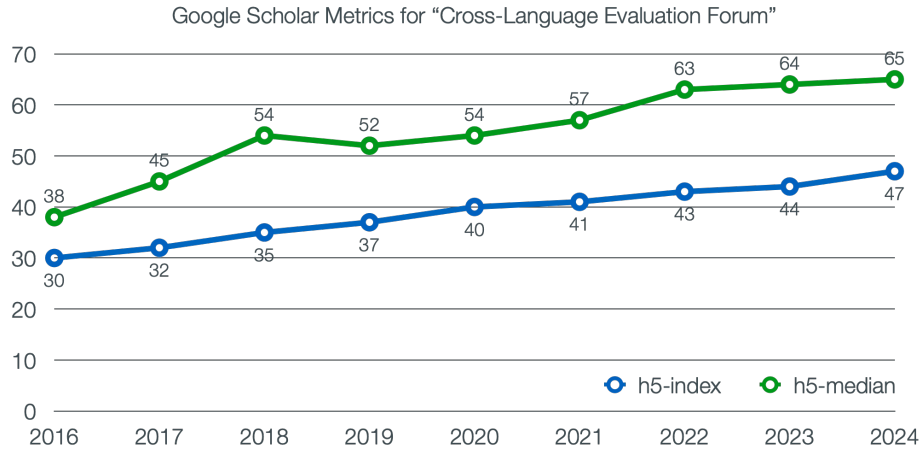
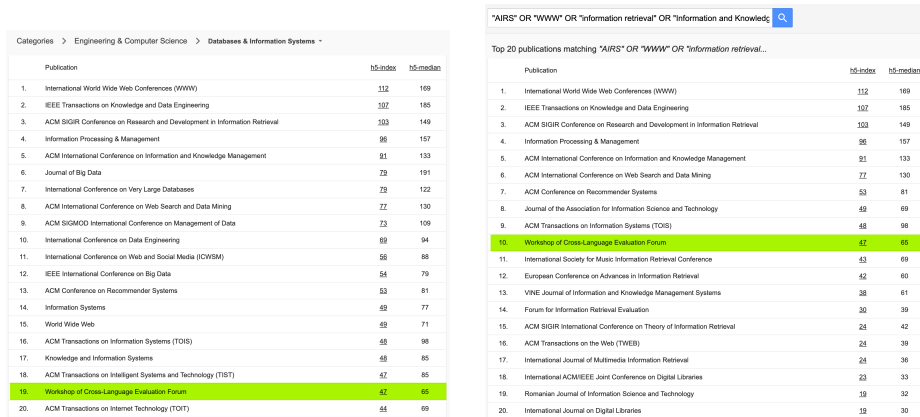


Fig. 7: Google Scholar Metrics for “Cross-Language Evaluation Forum”: h5-index (blue) and h5-median (green).



(a) Top-20 venues in the category “Database and Information Systems”

(b) Top-20 venues for a query targeted to IR and RS venues.

Fig. 8: Top venues according to Google Scholar Metrics 2024.

7 The CLEF Book

In occasion of the past 20th anniversary of CLEF, we prepared a book [202] 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 [619], describing how this has been interpreted in CLEF and in other internationally recognized evaluation initiatives [542]. In addition, the introductory chapter illustrates the activity and results of CLEF over the years in some detail [201]. Part II “Evaluation Infrastructure” presents research architectures and infrastructures that have been developed to manage experimental data [9] and to provide evaluation services in CLEF and elsewhere [425, 519].

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 [556] and the impact of languages on information access [322]; it then addresses multimodality from the perspective of both images [130, 429, 504, 622] and sound and vision [303]. Part IV “Retrieval in New Domains” deals with the medical domain [590], the intellectual property and patent domain [506], the biodiversity domain [295], and the structured data and semantic search domains [309]. Part V “Beyond Retrieval” covers information access tasks other than pure retrieval, namely question answering [477], digital text forensics [538], online reputation management [11], and continuous evaluation and living labs [272].

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 [368] and we discuss open issues and areas for future development, such as reproducibility and validity [218] and *Visual Analytics (VA)* for experimental evaluation [204]. In particular, the concluding chapter discusses the relevance of IR benchmarking in an industrial setting [317].

8 The CLEF Association

The CLEF Association¹² is an independent no-profit legal entity, established in October 2013 as a result of activity of the PROMISE¹³ Network of Excellence which backed CLEF from 2010 to 2013.

¹² <http://www.clef-initiative.eu/association>

¹³ <http://www.promise-noe.eu/>

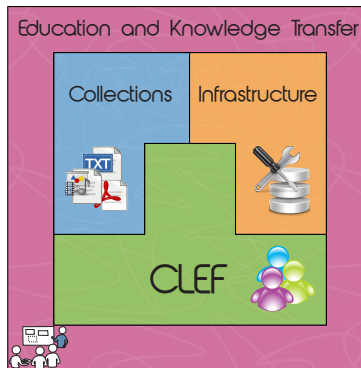


Fig. 9: Pillar activities of the CLEF Association.

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 9 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;
- *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 [575] 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.

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., Shakeri, A.: Meta Text Aligner: Text Alignment Based on Predicted Plagiarism Relation. In: Mothe et al. [420], pp. 193–199
2. Achilles, L., Mandl, T., Womser-Hacker, C.: “Meanspo Please, I Want to Lose Weight”: A Characterization Study of Meanspiration Content on Tumblr Based on Images and Texts. In: Barrón-Cedeño et al. [55], pp. 3–17
3. Adams, J., Bedrick, S.: Automatic Indexing of Journal Abstracts with Latent Semantic Analysis. In: Mothe et al. [420], pp. 200–208
4. 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. [219], pp. 162–173
5. Agirre, E., Di Nunzio, G.M., Ferro, N., Mandl, T., Peters, C.: CLEF 2008: Ad Hoc Track Overview. In: Peters et al. [488], pp. 15–37
6. Agirre, E., Di Nunzio, G.M., Mandl, T., Otegi, A.: CLEF 2009 Ad Hoc Track Overview: Robust-WSD Task. In: Peters et al. [489], pp. 36–49
7. 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. [495], pp. 908–917
8. 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. [117], pp. 88–99
9. 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 [203], pp. 105–122
10. 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)
11. Carrillo-de Albornoz, J., Gonzalo, J., Amigó, E.: RepLab: an Evaluation Campaign for Online Monitoring Systems. In: Ferro and Peters [203]

12. Alexander, A., Mars, M., Tingey, J.C., Yu, H., Backhouse, C., Reddy, S., Karlgren, J.: Audio Features, Precomputed for Podcast Retrieval and Information Access Experiments. In: Candan et al. [104], pp. 3–14
13. Alfalahi, A., Eriksson, G., Sneider, E.: Shadow Answers as an Intermediary in Email Answer Retrieval. In: Mothe et al. [420], pp. 209–214
14. Alhamzeh, A., Bouhaouel, M., Egyed-Zsigmond, E., Mitrović, J., Brunie, L., Kosch, H.: Query Expansion, Argument Mining and Document Scoring for an Efficient Question Answering System. In: Barrón-Cedeño et al. [55], pp. 162–174
15. Alhamzeh, A., Mukhopadhyaya, S., Hafid, S., Bremard, A., Egyed-Zsigmond, E., H., K., Brunie, L.: A Hybrid Approach for Stock Market Prediction Using Financial News and Stocktwits. In: Candan et al. [104], pp. 15–26
16. Alharbi, A., Stevenson, M.: Improving Ranking for Systematic Reviews Using Query Adaptation. In: Crestani et al. [138], pp. 141–148
17. Ali, E., Caputo, A., Lawless, S., Conlan, O.: Dataset Creation Framework for Personalized Type-Based Facet Ranking Tasks Evaluation. In: Candan et al. [104], pp. 27–39
18. Aliannejadi, M., Faggioli, G., Ferro, N., Vlachos, M. (eds.): CLEF 2023 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <https://ceur-ws.org/Vol-3497/>. (2023)
19. Aliannejadi, M., Huibers, T., Landoni, M., Murgia, E., Pera, M.S.: The Effect of Prolonged Exposure to Online Education on a Classroom Search Companion. In: Barrón-Cedeño et al. [55], pp. 62–78
20. Alkhalifa, R., Bilal, I., Borkakoty, H., Camacho-Collados, J., Deveaud, R., El-Ebshihy, A., Espinosa-Anke, L., Gonzalez-Saez, G., Galuščáková, P., Goeriot, L., Kochkina, E., Liakata, M., Loureiro, D., Mulhem, P., Piroi, F., Popel, M., Servan, C., Tayyar Madabushi, H., Zubiaga, A.: Overview of the CLEF-2023 LongEval Lab on Longitudinal Evaluation of Model Performance. In: Arampatzis et al. [36], pp. 440–458
21. Alkhalifa, R., Borkakoty, H., Deveaud, R., El-Ebshihy, A., Espinosa-Anke, L., Fink, T., Galuščáková, P., González Sáez, G.N., Goeriot, L., Iommi, D., Liakata, M., Tayyar Madabushi, H., Medina-Alias, P., Mulhem, P., Piroi, F., Popel, M., Zubiaga, A.: Overview of the CLEF 2024 LongEval Lab on Longitudinal Evaluation of Model Performance. In: Goeriot et al. [241]
22. Alkhalwaldeh, R.S., Jose, J.M.: Experimental Study on Semi-structured Peer-to-Peer Information Retrieval Network. In: Mothe et al. [420], pp. 3–14
23. Almeida, J.R., Fajarda, O., Oliveira, J.L.: File Forgery Detection Using a Weighted Rule-Based System. In: Arampatzis et al. [37], pp. 85–96
24. 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. [211], pp. 333–352
25. 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. [312], pp. 307–322
26. 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. [96]
27. Amigó, E., Corujo, A., Gonzalo, J., Meij, E., de Rijke, M.: Overview of RepLab 2012: Evaluating Online Reputation Management Systems. In: Forner et al. [210]
28. Amigó, E., Gonzalo, J., Verdejo, M.F.: A Comparison of Evaluation Metrics for Document Filtering. In: Forner et al. [209], pp. 38–49

29. Anderka, M., Stein, B.: Overview of the 1th International Competition on Quality Flaw Prediction in Wikipedia. In: Forner et al. [210]
30. Andrearczyk, V., Müller, H.: Deep Multimodal Classification of Image Types in Biomedical Journal Figures. In: Bellot et al. [64], pp. 3–14
31. 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. [117], pp. 112–123
32. 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)
33. Angelini, M., Ferro, N., Santucci, G., Silvello, G.: Improving Ranking Evaluation Employing Visual Analytics. In: Forner et al. [211], pp. 29–40
34. Antici, F., Bolognini, K., Inajetovic, M.A., Ivasiuk, B., Galassi, A., Ruggeri, F.: SubjectivITA: An Italian Corpus for Subjectivity Detection in Newspapers. In: Candan et al. [104], pp. 40–52
35. Antunes, H., Lopes, C.: Analyzing the adequacy of readability indicators to a non-English language. In: Crestani et al. [138], pp. 149–155
36. Arampatzis, A., Kanoulas, E., Tsikrika, T., Vrochidis, S., Giachanou, A., Li, D., Aliannejadi, M., Vlachos, M., Faggioli, G., Ferro, N. (eds.): Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Fourteenth International Conference of the CLEF Association (CLEF 2023). *Lecture Notes in Computer Science (LNCS)* 14163, Springer, Heidelberg, Germany (2023)
37. Arampatzis, A., Kanoulas, E., Tsikrika, T., Vrochidis, S., Joho, H., Lioma, C., Eickhoff, K., Névóel, A., Cappellato, L., Ferro, N. (eds.): Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Eleventh International Conference of the CLEF Association (CLEF 2020). *Lecture Notes in Computer Science (LNCS)* 12260, Springer, Heidelberg, Germany (2020)
38. Arcos, I., Rosso, P.: Sexism Identification on TikTok: A Multimodal AI Approach with Text, Audio, and Video. In: Goeriot et al. [240]
39. Argamon, S., Juola, P.: Overview of the International Authorship Identification Competition at PAN-2011. In: Petras et al. [502]
40. 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)
41. Arni, T., Clough, P., Sanderson, M., Grubinger, M.: Overview of the ImageCLEF-photo 2008 Photographic Retrieval Task. In: Peters et al. [488], pp. 500–511
42. Arora, P., Foster, J., Jones, G.J.F.: Query Expansion for Sentence Retrieval Using Pseudo Relevance Feedback and Word Embedding. In: Jones et al. [305], pp. 97–103
43. 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. [96]
44. Avila, J., Rodrigo, A., Centeno, R.: Best of Touché 2023 task 4: Testing Data Augmentation and Label Propagation for Multilingual multi-target Stance Detection. In: Goeriot et al. [240]

45. Ayele, A.A., Babakov, N., Bevendorff, J., Bonet Casals, X., Chulvi, B., Dementieva, D., Elnagar, A., Freitag, D., Fröbe, M., Korenčić, D., Mayerl, M., Moskovskiy, D., Mukherjee, A., Panchenko, A., Potthast, M., Rangel Pardo, F., Rizwan, N., Rosso, P., Schneider, F., Smirnova, A., Stamatatos, E., Stakovskii, E., Stein, B., Taulé, M., Ustalov, D., Wang, X., Wiegmann, N., Muhie Yimam, S., Zangerle, E.: Overview of PAN 2024: Multi-Author Writing Style Analysis, Multilingual Text Detoxification, Oppositional Thinking Analysis, and Generative AI Authorship Verification. In: Goeuriot et al. [241]
46. Azarbyad, H., Marx, M.: How many labels? Determining the Number of Labels in Multi-Label Text Classification. In: Crestani et al. [138], pp. 156–163
47. Azarbyad, H., Saan, F., Dehghani, D., Marx, M., Kamps, J.: Are Topically Diverse Documents Also Interesting? In: Mothe et al. [420], pp. 215–221
48. Azarbyad, H., Shakery, A., Faili, H.: Exploiting Multiple Translation Resources for English-Persian Cross Language Information Retrieval. In: Forner et al. [211], pp. 93–99
49. 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. [209], pp. 26–37
50. Balog, K., Azzopardi, L., Kamps, J., de Rijke, M.: Overview of WebCLEF 2006. In: Peters et al. [486], pp. 803–819
51. 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)
52. Baradaran Hashemi, H., Shakery, A., Feili, H.: Creating a Persian-English Comparable Corpus. In: Agosti et al. [10], pp. 27–39
53. Barrón-Cedeño, A., Alam, F., Galassi, A., Da San Martino, G., Nakov, P., Elsayed, T., Azizov, D., Caselli, T., Cheema, G.S., Haouari, F., Hasanain, M., Kutlu, M., Li, C., Ruggeri, F., Struß, J.M., Zaghouani, W.: Overview of the CLEF-2023 CheckThat! Lab on Checkworthiness, Subjectivity, Political Bias, Factuality, and Authority of News Articles and Their Source. In: Arampatzis et al. [36], pp. 251–275
54. Barrón-Cedeño, A., Alam, F., Struß, J.M., Nakov, P., Chakraborty, T., Elsayed, T., Przybyła, P., Caselli, T., Da San Martino, G., Haouari, F., Hasanain, M., Li, C., Piskorski, J., Ruggeri, F., Song, X., Suwaileh, R.: Overview of the CLEF-2024 CheckThat! Lab: Check-Worthiness, Subjectivity, Persuasion, Roles, Authorities, and Adversarial Robustness. In: Goeuriot et al. [241]
55. Barrón-Cedeño, A., Da San Martino, G., Degli Esposti, M., Sebastiani, F., Macdonald, C., Pasi, G., Hanbury, A., Potthast, M., Faggioli, G., Ferro, N. (eds.): Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Thirteenth International Conference of the CLEF Association (CLEF 2022). Lecture Notes in Computer Science (LNCS) 13390, Springer, Heidelberg, Germany (2022)
56. Barrón-Cedeño, A., Elsayed, T., Nakov, P., Da San Martino, G., Hasanain, M., Suwaileh, R., Haouari, F., Babulkov, N., Hamdan, B., Nikolov, A., Shaar, S., Sheikh Ali, Z.: Overview of CheckThat! 2020: Automatic Identification and Verification of Claims in Social Media. In: Arampatzis et al. [37], pp. 215–236
57. Basile, A., Caselli, T.: Protest Event Detection: When Task-Specific Models Outperform an Event-Driven Method. In: Arampatzis et al. [37], pp. 97–111
58. 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. [64], pp. 143–156

59. Baudis, P., Sedivý, J.: Modeling of the Question Answering Task in the YodaQA System. In: Mothe et al. [420], pp. 222–228
60. Bauer, C., Carterette, B., Ferro, N., Fuhr, N., Beel, J., Breuer, T., Clarke, C.L.A., Crescenzi, A., Demartini, G., Di Nunzio, G.M., Dietz, L., Faggioli, G., Ferwerda, B., Fröbe, M., Hagen, M., Hanbury, A., Hauff, C., Jannach, D., Kando, N., Kanoulas, E., Knijnenburg, B.P., Kruschwitz, U., Li, M., Maistro, M., Michiels, L., Papenmeier, A., Potthast, M., Rosso, P., Said, A., Schaer, P., Seifert, C., Spina, D., Stein, B., Tintarev, N., Urbano, J., Wachsmuth, H., Willemsen, M.C., Zobel, J.: Report on the Dagstuhl Seminar on Frontiers of Information Access Experimentation for Research and Education. SIGIR Forum **57**(1), 7:1–7:28 (June 2023)
61. Bauer, C., Carterette, B.A., Ferro, N., Fuhr, N., Faggioli, G. (eds.): Report from Dagstuhl Seminar 23031: Frontiers of Information Access Experimentation for Research and Education. Dagstuhl Reports, Volume 13, Number 1, Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Germany (2023)
62. 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. [312], pp. 212–228
63. 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. [211], pp. 269–281
64. 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)
65. Beloborodov, A., Braslavski, P., Driker, M.: Towards Automatic Evaluation of Health-Related CQA Data. In: Kanoulas et al. [312], pp. 7–18
66. Bensalem, I., Rosso, P., Chikhi, S.: A New Corpus for the Evaluation of Arabic Intrinsic Plagiarism Detection. In: Forner et al. [211], pp. 53–58
67. Berendsen, R., Tsagkias, M., de Rijke, M., Meij, E.: Generating Pseudo Test Collections for Learning to Rank Scientific Articles. In: Catarci et al. [117], pp. 42–53
68. 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. [211], pp. 120–123
69. 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. [211], pp. 167–178
70. Bernard, G., Suire, C., Faucher, C., Doucet, A., Rosso, P.: Tracking News Stories in Short Messages in the Era of Infodemic. In: Barrón-Cedeño et al. [55], pp. 18–32
71. Besançon, R., Chaudiron, S., Mostefa, D., Hamon, O., I., T., Choukri, K.: Overview of CLEF 2008 INFILE Pilot Track. In: Peters et al. [488], pp. 939–946
72. 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. [489], pp. 342–353

73. Bevendorff, J., Borrego-Obrador, I., China-Ríos, M., Franco-Salvador, M., Fröbe, M., Heini, A., Kredens, K., Mayerl, M., Pezik, P., Potthast, M., Rangel, F., Rosso, P., Stamatatos, E., Stein, B., Wiegmann, M., Wolska, M., Zangerle, E.: Overview of PAN 2023: Authorship Verification, Multi-Author Writing Style Analysis, Profiling Cryptocurrency Influencers, and Trigger Detection. In: Arampatzis et al. [36], pp. 459–481
74. Bevendorff, J., Chulvi, B., De La Peña Sarracén, G.L., Kestemont, M., Manjavacas, E., Markov, I., Mayerl, M., Potthast, M., Rangel, F., Rosso, P., Stamatatos, E., Stein, B., Wiegmann, M., Wolska, M., Zangerle, E.: Overview of PAN 2021: Authorship Verification, Profiling Hate Speech Spreaders on Twitter, and Style Change Detection. In: Candan et al. [104], pp. 419–431
75. Bevendorff, J., Chulvi, B., Fersini, E., Heini, A., Kestemont, M., Kredens, K., Mayerl, M., Ortega-Bueno, R., Pezik, P., Potthast, M., Rangel, F., Rosso, P., Stamatatos, E., Stein, B., Wiegmann, M., Wolska, M., Zangerle, E.: Overview of PAN 2022: Authorship Verification, Profiling Irony and Stereotype Spreaders, and Style Change Detection. In: Barrón-Cedeño et al. [55], pp. 382–394
76. Bevendorff, J., Ghanem, B., Giachanou, A., Kestemont, M., Manjavacas, E., Markov, I., Mayerl, M., Potthast, M., Rangel Pardo, F.M., Rosso, P., Specht, G., Stamatatos, E., Stein, B., Wiegmann, M., Zangerle, E.: Overview of PAN 2020: Authorship Verification, Celebrity Profiling, Profiling Fake News Spreaders on Twitter, and Style Change Detection. In: Arampatzis et al. [37], pp. 372–383
77. Bhaskar, P., Bandyopadhyay, S.: Language Independent Query Focused Snippet Generation. In: Catarci et al. [117], pp. 138–140
78. Birolo, G., Bosoni, P., Faggioli, G., Aidos, H., Bergamaschi, R., Cavalla, P., Chiò, A., Dagliati, A., de Carvalho, M., Di Nunzio, G.M., Fariselli, P., Garcia Dominguez, J.M., Gromicho, M., Guazzo, A., Longato, E., Madeira, S., Manera, U., Marchesin, S., Menotti, L., Silvello, G., Tavazzi, E., Tavazzi, E., Trescato, I., Vettoretti, M., Di Camillo, B., Ferro, N.: Intelligent Disease Progression Prediction: Overview of iDPP@CLEF 2024. In: Goeuriot et al. [241]
79. Blinov, V., Mishchenko, K., Bolotova, V., Braslavski, P.: A Pinch of Humor for Short-Text Conversation: An Information Retrieval Approach. In: Jones et al. [305], pp. 3–15
80. Boenninghoff, B., Kolossa, D., Nickel, R.M.: Self-calibrating Neural-Probabilistic Model for Authorship Verification Under Covariate Shift. In: Candan et al. [104], pp. 145–158
81. Bondarenko, A., Fröbe, M., Beloucif, M., Gienapp, L., Ajjour, Y., Panchenko, A., Biemann, C., Stein, B., Wachsmuth, H., Potthast, M., Hagen, M.: Overview of Touché 2020: Argument Retrieval - Extended Abstract. In: Arampatzis et al. [37], pp. 384–395
82. Bondarenko, A., Fröbe, M., Kiesel, J., Schlatt, F., Barriere, V., Ravenet, B., Hemamou, L., Luck, S., Reimer, J.H., Stein, B., Potthast, M., Hagen, M.: Overview of Touché 2023: Argument and Causal Retrieval. In: Arampatzis et al. [36], pp. 507–530
83. Bondarenko, A., Fröbe, M., Kiesel, J., Syed, S., Gurcke, T., Beloucif, M., Panchenko, A., Biemann, C., Stein, B., Wachsmuth, H., Potthast, M., Hagen, M.: Overview of Touché 2022: Argument Retrieval. In: Barrón-Cedeño et al. [55], pp. 311–336
84. Bondarenko, A., Gienapp, L., Fröbe, M., Beloucif, M., Ajjour, Y., Panchenko, A., Biemann, C., Stein, B., Wachsmuth, J., Potthast, M., Hagen, M.: Overview of Touché 2021: Argument Retrieval. In: Candan et al. [104], pp. 450–467

85. Borchert, F., Llorca, I., Schapranow, M.P.: Cross-Lingual Candidate Retrieval and Re-ranking for Biomedical Entity Linking. In: Arampatzis et al. [36], pp. 135–147
86. Bordea, G., Thiessard, F., Hamon, T., Mougin, F.: Automatic Query Selection for Acquisition and Discovery of Food-Drug Interactions. In: Bellot et al. [64], pp. 115–120
87. 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)
88. 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)
89. 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)
90. Braschler, M.: CLEF 2000 – Overview of Results. In: Peters [480], pp. 89–101
91. Braschler, M.: CLEF 2001 – Overview of Results. In: Peters et al. [483], pp. 9–26
92. Braschler, M.: CLEF 2002 – Overview of Results. In: Peters et al. [484], pp. 9–27
93. Braschler, M.: CLEF 2003 – Overview of Results. In: Peters et al. [485], pp. 44–63
94. Braschler, M., Choukri, K., Ferro, N., Hanbury, A., Karlgren, J., Müller, H., Petras, V., Pianta, E., de Rijke, M., Santucci, G.: A PROMISE for Experimental Evaluation. In: Agosti et al. [10], pp. 140–144
95. Braschler, M., Di Nunzio, G.M., Ferro, N., Peters, C.: CLEF 2004: Ad Hoc Track Overview and Results Analysis. In: Peters et al. [487], pp. 10–26
96. 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)
97. Braschler, M., Peters, C.: Cross-Language Evaluation Forum: Objectives, Results, Achievements. *Information Retrieval* **7**(1–2), 7–31 (2004)
98. Braslavski, P., Verberne, S., Talipov, R.: Show Me How to Tie a Tie: Evaluation of Cross-Lingual Video Retrieval. In: Fuhr et al. [219], pp. 3–15
99. Bucur, A.M.: Leveraging LLM-Generated Data for Detecting Depression Symptoms on Social Media. In: Goeuriot et al. [240]
100. Budíková, P., Batko, M., Botorek, J., Zezula, P.: Search-Based Image Annotation: Extracting Semantics from Similar Images. In: Mothe et al. [420], pp. 327–339
101. Buraya, K., Farseev, A., Filchenkov, A.: Multi-view Personality Profiling Based on Longitudinal Data. In: Bellot et al. [64], pp. 15–27
102. Buscaldi, D., Rosso, P., Chulvi, B., Wang, T.: Classification of Social Media Hateful Screenshots Inciting Violence and Discrimination. In: Goeuriot et al. [240]
103. Cabanac, G., Hubert, G., Boughanem, M., Chrisment, C.: Tie-Breaking Bias: Effect of an Uncontrolled Parameter on Information Retrieval Evaluation. In: Agosti et al. [10], pp. 112–123
104. Candan, K.S., Ionescu, B., Goeuriot, L., Larsen, B., Müller, H., Joly, A., Maistro, M., Piroi, F., Faggioli, G., Ferro, N. (eds.): Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Twelfth International Conference of the CLEF Association (CLEF 2021). Lecture Notes in Computer Science (LNCS) 12880, Springer, Heidelberg, Germany (2021)
105. Capari, A., Azarbondy, H., Tsatsaronis, G., Afzal, Z., Dunham, J., Kamps, J.: Knowledge Acquisition Passage Retrieval: Corpus, Ranking Models, and Evaluation Resources. In: Goeuriot et al. [240]

106. Cappellato, L., Eickhoff, C., Ferro, N., Névéol, A. (eds.): CLEF 2020 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-2696/> (2020)
107. 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)
108. 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)
109. 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)
110. 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)
111. 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)
112. 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. [312], pp. 192–211
113. 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. [211], pp. 250–268
114. Cardoso, R., Marinho, Z., Mendes, A., S., M.: Priberam at MESINESP Multilabel Classification of Medical Texts Task. In: Candan et al. [104], pp. 159–172
115. Cassani, L., Livraga, G., Viviani, M.: Assessing Document Sanitization for Controlled Information Release and Retrieval in Data Marketplaces. In: Goeuriot et al. [240]
116. Cassidy, T., Ji, H., Deng, H., Zheng, J., Han, J.: Analysis and Refinement of Cross-Lingual Entity Linking. In: Catarci et al. [117], pp. 1–12
117. 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)
118. Chaa, M., Nouali, O., Bellot, P.: Combining Tags and Reviews to Improve Social Book Search Performance. In: Bellot et al. [64], pp. 64–75
119. Chappell, T., Geva, S.: Overview of the INEX 2012 Relevance Feedback Track. In: Forner et al. [210]
120. Chidlovskii, B., Csurka, G., Clinchant, S.: Evaluating Stacked Marginalised Denoising Autoencoders Within Domain Adaptation Methods. In: Mothe et al. [420], pp. 15–27
121. Chikka, V.R., Mariyasagayam, N., Niwa, Y., Karlapalem, K.: Information Extraction from Clinical Documents: Towards Disease/Disorder Template Filling. In: Mothe et al. [420], pp. 389–401
122. 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. [138], pp. 164–170

123. Chulif, S., Loong Chang, Y.: Herbarium-Field Triplet Network for Cross-Domain Plant Identification. In: Candan et al. [104], pp. 173–188
124. 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. [211], pp. 321–332
125. Clough, P., Goodale, P.: Selecting Success Criteria: Experiences with an Academic Library Catalogue. In: Forner et al. [211], pp. 59–70
126. 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. [486], pp. 223–256
127. 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. [494], pp. 535–557
128. Clough, P., Müller, H., Sanderson, M.: The CLEF 2004 Cross-Language Image Retrieval Track. In: Peters et al. [487], pp. 597–613
129. Clough, P., Sanderson, M.: The CLEF 2003 Cross Language Image Retrieval Track. In: Peters et al. [485], pp. 581–593
130. Clough, P., Tsirikika, T.: Multi-lingual Retrieval of Pictures in ImageCLEF. In: Ferro and Peters [203]
131. Clough, P., Willett, P., Lim, J.: Unfair Means: Use Cases Beyond Plagiarism. In: Mothe et al. [420], pp. 229–234
132. 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. [138], pp. 49–61
133. Collovini, S., de Bairros P. Filho, M., Vieira, R.: Analysing the Role of Representation Choices in Portuguese Relation Extraction. In: Mothe et al. [420], pp. 105–116
134. Conlan, O., Fraser, K., Kelly, L., Yousuf, B.: A User Modeling Shared Challenge Proposal. In: Crestani et al. [138], pp. 171–177
135. Corbara, S., Chulvi, B., Rosso, P., Moreo, A.: Rhythmic and Psycholinguistic Features for Authorship Tasks in the Spanish Parliament: Evaluation and Analysis. In: Barrón-Cedeño et al. [55], pp. 79–92
136. Corezola Pereira, R., Pereira Moreira, V., Galante, R.: A New Approach for Cross-Language Plagiarism Analysis. In: Agosti et al. [10], pp. 15–26
137. 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. [420], pp. 340–351
138. Crestani, F., Braschler, M., Savoy, J., Rauber, A., Müller, H., Losada, D.E., Heinatz Bürki, 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)
139. Custódio, J., Paraboni, I.: An Ensemble Approach to Cross-Domain Authorship Attribution. In: Crestani et al. [138], pp. 201–212
140. Dadashkarimi, J., Esfahani, H.N., Faili, H., Shakery, A.: SS4MCT: A Statistical Stemmer for Morphologically Complex Texts. In: Fuhr et al. [219], pp. 201–207
141. 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. [138], pp. 402–416

142. Dehghani, M., Azarbondy, H., Kamps, J., Marx, M.: Two-Way Parsimonious Classification Models for Evolving Hierarchies. In: Fuhr et al. [219], pp. 69–82
143. Déjean, S., Mothe, J., Ullah, M.Z.: Studying the Variability of System Setting Effectiveness by Data Analytics and Visualization. In: Crestani et al. [138], pp. 62–74
144. Del Moro, M., Tudosie, S.C., Vannoni, F., Galassi, A., Ruggeri, F.: Inception Models for Fashion Image Captioning: An Extensive Study on Multiple Datasets. In: Arampatzis et al. [36], pp. 3–14
145. Deneu, B., Servajean, M., Botella, C., Joly, A.: Evaluation of Deep Species Distribution Models using Environment and Co-occurrences. In: Crestani et al. [138], pp. 213–225
146. Deselaers, T., Deserno, T.M.: Medical Image Annotation in ImageCLEF 2008. In: Peters et al. [488], pp. 523–530
147. Deselaers, T., Hanbury, A.: The Visual Concept Detection Task in ImageCLEF 2008. In: Peters et al. [488], pp. 531–538
148. 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. [495], pp. 445–471
149. Devezas, J., Nunes, S.: Index-Based Semantic Tagging for Efficient Query Interpretation. In: Fuhr et al. [219], pp. 208–213
150. Dhanani, F., Rafi, M., Atif Tahir, M.: Humour Translation with Transformers. In: Arampatzis et al. [36], pp. 148–160
151. Dhrangadhariya, A., Aguilar, G., Solorio, T., Hilfiker, R., Müller, H.: End-to-End Fine-Grained Neural Entity Recognition of Patients, Interventions, Outcomes. In: Candan et al. [104], pp. 65–77
152. 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. [209], pp. 119–130
153. Di Nunzio, G.M.: A Study on a Stopping Strategy for Systematic Reviews Based on a Distributed Effort Approach. In: Arampatzis et al. [37], pp. 112–123
154. Di Nunzio, G.M., Ferro, N., Jones, G.J.F., Peters, C.: CLEF 2005: Ad Hoc Track Overview. In: Peters et al. [494], pp. 11–36
155. Di Nunzio, G.M., Ferro, N., Mandl, T., Peters, C.: CLEF 2006: Ad Hoc Track Overview. In: Peters et al. [486], pp. 21–34
156. Di Nunzio, G.M., Ferro, N., Mandl, T., Peters, C.: CLEF 2007: Ad Hoc Track Overview. In: Peters et al. [495], pp. 13–32
157. 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. [502]
158. Di Nunzio, G.M., Vezzani, F.: Using R Markdown for Replicable Experiments in Evidence Based Medicine. In: Bellot et al. [64], pp. 28–39
159. Di Nunzio, G.M., Vezzani, F.: Did I Miss Anything? A Study on Ranking Fusion and Manual Query Rewriting in Consumer Health Search. In: Barrón-Cedeño et al. [55], pp. 217–229
160. Di Nunzio, G.M., Vezzani, F.: The Best is Yet to Come: A Reproducible Analysis of CLEF eHealth TAR Experiments. In: Arampatzis et al. [36], pp. 15–20
161. 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. [64], pp. 157–168

162. Dietz, F.: The Curious Case of Session Identification. In: Arampatzis et al. [37], pp. 69–74
163. Dietz, F., Petras, V.: A Component-Level Analysis of an Academic Search Test Collection. - Part I: System and Collection Configurations. In: Jones et al. [305], pp. 16–28
164. Dietz, F., Petras, V.: A Component-Level Analysis of an Academic Search Test Collection. - Part II: Query Analysis. In: Jones et al. [305], pp. 29–42
165. Domann, J., Lommatzsch, A.: A Highly Available Real-Time News Recommender Based on Apache Spark. In: Jones et al. [305], pp. 161–172
166. Dsilva, R.R.: From Sentence Embeddings to Large Language Models to Detect and Understand Wordplay. In: Goeuriot et al. [240]
167. Efimov, P., Chertok, A., Boytsov, L., Braslavski, P.: SberQuAD – Russian Reading Comprehension Dataset: Description and Analysis. In: Arampatzis et al. [37], pp. 3–15
168. Ehrmann, M., Romanello, M., Flückiger, A., Clematide, S.: Overview of CLEF HIPE 2020: Named Entity Recognition and Linking on Historical Newspapers. In: Arampatzis et al. [37], pp. 288–310
169. Ehrmann, M., Romanello, M., Najem-Meyer, S., Doucet, A., Clematide, S.: Overview of HIPE-2022: Named Entity Recognition and Linking in Multilingual Historical Documents. In: Barrón-Cedeño et al. [55], pp. 423–446
170. El-Ebshihy, A., Fink, P., Gonzalez-Saez, G., Galuščáková, P., Piroi, F., Iommi, D., Goeuriot, G., Mulhem, P.: Predicting Retrieval Performance Changes in Evolving Evaluation Environments. In: Arampatzis et al. [36], pp. 21–33
171. 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. [138], pp. 301–321
172. Ermakova, K.: A Method for Short Message Contextualization: Experiments at CLEF/INEX. In: Mothe et al. [420], pp. 352–363
173. Ermakova, L., Bellot, P., Braslavski, P., Kamps, J., Mothe, J., Nurbakova, D., Ovchinnikova, I., SanJuan, E.: Overview of SimpleText 2021 - CLEF Workshop on Text Simplification for Scientific Information Access. In: Candan et al. [104], pp. 432–449
174. Ermakova, L., Bosser, A.G., Miller, T., Palma Preciado, V.M., Sidorov, G., Jatowt, A.: Overview of JOKER @ CLEF-2024: Automatic Humour Analysis. In: Goeuriot et al. [241]
175. Ermakova, L., Goeuriot, L., Mothe, J., Mulhem, P., Nie, J.Y., SanJuan, E.: CLEF 2017 Microblog Cultural Contextualization Lab Overview. In: Jones et al. [305], pp. 304–314
176. Ermakova, L., Miller, T., Bosser, A.G., Palma Preciado, V.M., Sidorov, G., Jatowt, A.: Overview of JOKER – CLEF-2023 Track on Automatic Wordplay Analysis. In: Arampatzis et al. [36], pp. 397–415
177. Ermakova, L., Miller, T., Regattin, F., Bosser, A.G., Borg, C., Mathurin, E., Le Corre, G., Araújo, S., Hannachi, R., Boccou, J., Digue, A., Damoy, A., Jean-jean, B.: Overview of JOKER@CLEF 2022: Automatic Wordplay and Humour Translation Workshop. In: Barrón-Cedeño et al. [55], pp. 447–469
178. Ermakova, L., SanJuan, E., Huët, S., Azarbonyad, H., Augereau, O., Kamps, J.: Overview of the CLEF 2023 SimpleText Lab: Automatic Simplification of Scientific Texts. In: Arampatzis et al. [36], pp. 482–506

179. Ermakova, L., Sanjuan, E., Huet, S., Azaronyad, H., Di Nunzio, G.M., Vezani, F., D'Souza, J., Kamps, J.: Overview of the CLEF 2024 SimpleText Track: Improving Access to Scientific Texts for Everyone. In: Goeuriot et al. [241]
180. Ermakova, L., SanJuan, E., Kamps, J., Huet, S., Ovchinnikova, I., Nurbakova, D., Araújo, S., Hannachi, R., Mathurin, E., Bellot, P.: Overview of the CLEF 2022 SimpleText Lab: Automatic Simplification of Scientific Texts. In: Barrón-Cedeño et al. [55], pp. 470–494
181. Esuli, A., Moreo, A., Sebastiani, F., Sperduti, G.: A Concise Overview of LeQua@CLEF 2022: Learning to Quantify. In: Barrón-Cedeño et al. [55], pp. 362–381
182. Esuli, F., Sebastiani, F.: Evaluating Information Extraction. In: Agosti et al. [10], pp. 100–111
183. Eyuboglu, A.B., Altun, B., Arslan, M.B., Sonmezer, E., Kutlu, M.: Fight Against Misinformation on Social Media: Detecting Attention-Worthy and Harmful Tweets and Verifiable and Check-Worthy Claims. In: Arampatzis et al. [36], pp. 161–173
184. Ezzeldin, A.M., Kholief, M.H., El-Sonbaty, Y.: ALQASIM: Arabic Language Question Answer Selection in Machines. In: Forner et al. [211], pp. 100–103
185. Fabregat, H., Duque, A., Araujo, L., Martinez-Romo, J.: A Re-labeling Approach Based on Approximate Nearest Neighbors for Identifying Gambling Disorders in Social Media. In: Arampatzis et al. [36], pp. 174–185
186. Faggioli, G., Ferro, N., Galuščáková, P., Garcia Seco de Herrera, A. (eds.): CLEF 2024 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073. (2024)
187. Faggioli, G., Ferro, N., Hanbury, A., Potthast, M. (eds.): CLEF 2022 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073. <http://ceur-ws.org/Vol-3180/> (2022)
188. Faggioli, G., Ferro, N., Joly, A., Maistro, M., Piroi, F. (eds.): CLEF 2021 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-2936/> (2021)
189. Faggioli, G., Guazzo, A., Marchesin, S., Menotti, L., Trescato, I., Aidos, H., Bergamaschi, R., Birolo, G., Cavalla, P., Chiò, A., Dagliati, A., de Carvalho, M., Di Nunzio, G.M., Fariselli, P., García Dominguez, J.M., Gromicho, M., Longato, E., Madeira, S.C., Manera, U., Silvello, G., Tavazzi, E., Tavazzi, E., Vettoretti, M., Di Camillo, B., Ferro, N.: Intelligent Disease Progression Prediction: Overview of iDPP@CLEF 2023. In: Arampatzis et al. [36], pp. 343–369
190. Federico, M., Bertoldi, N., Levow, G.A., Jones, G.J.F.: CLEF 2004 Cross-Language Spoken Document Retrieval Track. In: Peters et al. [487], pp. 816–820
191. Federico, M., Jones, G.J.F.: The CLEF 2003 Cross-Language Spoken Document Retrieval Track. In: Peters et al. [485], p. 646
192. Fernández-Pichel, M., Losada, D.E., Pichel, J.C., Elswailer, D.: Comparing Traditional and Neural Approaches for Detecting Health-Related Misinformation. In: Candan et al. [104], pp. 78–90
193. Ferrante, M., Ferro, N., Maistro, M.: Rethinking How to Extend Average Precision to Graded Relevance. In: Kanoulas et al. [312], pp. 19–30
194. Ferrante, M., Ferro, N., Piazzon, L.: s-AWARE: supervised measure-based methods for crowd-assessors combination. In: Arampatzis et al. [37], pp. 16–27
195. Ferro, N.: What Happened in CLEF... For a While? In: Crestani et al. [138], pp. 3–45
196. Ferro, N.: What Happened in CLEF... For Another While? In: Goeuriot et al. [240]

197. 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)
198. Ferro, N., Harman, D.: CLEF 2009: Grid@CLEF Pilot Track Overview. In: Peters et al. [489], pp. 552–565
199. Ferro, N., Maistro, M., Sakai, T., Soboroff, I.: CENTRE@CLEF2018: Overview of the Replicability Task. In: Cappellato et al. [111]
200. Ferro, N., Peters, C.: CLEF 2009 Ad Hoc Track Overview: TEL & Persian Tasks. In: Peters et al. [489], pp. 13–35
201. 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 [203], pp. 3–44
202. 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)
203. 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)
204. Ferro, N., Santucci, G.: Visual Analytics and IR Experimental Evaluation. In: Ferro and Peters [203], pp. 565–582
205. Ferro, N., Silvello, G.: CLEF 15th Birthday: What Can We Learn From Ad Hoc Retrieval? In: Kanoulas et al. [312], pp. 31–43
206. Ferro, N., Silvello, G.: The CLEF Monolingual Grid of Points. In: Fuhr et al. [219], pp. 16–27
207. 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)
208. Fontanella, S., Rodríguez-Sánchez, A.J., Piater, J., Szedmak, S.: Kronecker Decomposition for Image Classification. In: Fuhr et al. [219], pp. 137–149
209. 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)
210. 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)
211. 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)
212. 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)
213. 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. [488], pp. 262–295

214. 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. [420], pp. 28–40
215. Frayling, E., Macdonald, C., McDonald, G., Ounis, I.: Using Entities in Knowledge Graph Hierarchies to Classify Sensitive Information. In: Barrón-Cedeño et al. [55], pp. 125–132
216. Frenda, S. and Patti, V., Rosso, P.: When Sarcasm Hurts: Irony-Aware Models for Abusive Language Detection. In: Arampatzis et al. [36], pp. 34–47
217. Fröbe, M., Akiki, C., Potthast, M., Hagen, M.: Noise-Reduction for Automatically Transferred Relevance Judgments. In: Barrón-Cedeño et al. [55], pp. 48–61
218. Fuhr, N.: Reproducibility and Validity in CLEF. In: Ferro and Peters [203]
219. 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)
220. 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. [502]
221. Gäde, M., Stiller, J., Petras, V.: Which Log for Which Information? Gathering Multilingual Data from Different Log File Types. In: Agosti et al. [10], pp. 70–81
222. Galuscáková, P., Pecina, P., Hajic, J.: Penalty Functions for Evaluation Measures of Unsegmented Speech Retrieval. In: Catarci et al. [117], pp. 100–111
223. Ganguly, D., Jones, G.J.F.: A Gamified Approach to Relevance Judgement. In: Fuhr et al. [219], pp. 214–220
224. Ganguly, D., Leveling, J., Jones, G.J.F.: Simulation of Within-Session Query Variations Using a Text Segmentation Approach. In: Forner et al. [209], pp. 89–94
225. Ganguly, D., Leveling, J., Jones, G.J.F.: A Case Study in Decomposing for Bengali Information Retrieval. In: Forner et al. [211], pp. 108–119
226. Garmash, E., Tanaka, E., Clifton, A., Correia, J., Jat, S., Zhu, W., Jones, R., Karlgren, J.: Cem Mil Podcasts: A Spoken Portuguese Document Corpus for Multi-modal, Multi-lingual and Multi-dialect Information Access Research. In: Arampatzis et al. [36], pp. 48–59
227. Gebremeskel, G.G., de Vries, A.P.: Random Performance Differences Between Online Recommender System Algorithms. In: Fuhr et al. [219], pp. 187–200
228. 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. [486], pp. 852–876
229. Gey, F.C., Larson, R.R., Sanderson, M., Joho, H., Clough, P., Petras, V.: GeoCLEF: The CLEF 2005 Cross-Language Geographic Information Retrieval Track Overview. In: Peters et al. [494], pp. 908–919
230. Ghosh, S., Singhanian, P., Singh, S., Rudra, K., Ghosh, S.: Stance Detection in Web and Social Media : A Comparative Study. In: Crestani et al. [138], pp. 75–87
231. Giampiccolo, D., Forner, P., Herrera, J., Peñas, A., Ayache, C., Forascu, C., Jijikoun, V., Osenova, P., Rocha, P., Sacaleanu, B., Sutcliffe, R.F.E.: Overview of the CLEF 2007 Multilingual Question Answering Track. In: Peters et al. [495], pp. 200–236
232. Gînsca, A.L., Popescu, A., Lupu, M., Iftene, A., Kanellos, I.: Evaluating User Image Tagging Credibility. In: Mothe et al. [420], pp. 41–52

233. Glinos, D.G.: Discovering Similar Passages within Large Text Documents. In: Kanoulas et al. [312], pp. 98–109
234. Gobeill, J., Gaudinat, A., Ruch, P.: Instance-Based Learning for Tweet Monitoring and Categorization. In: Mothe et al. [420], pp. 235–240
235. 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. [502]
236. 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. [210]
237. Goeuriot, L., Kelly, L., Suominen, H., Hanlen, L., Névéol, A., Grouin, C., Palotti, J.R.M., Zuccon, G.: Overview of the CLEF eHealth Evaluation Lab 2015. In: Mothe et al. [420], pp. 429–443
238. Goeuriot, L., Kelly, L., Suominen, H., Névéol, A., Robert, A., Kanoulas, E., Spijker, R., Palotti, J.R.M., Zuccon, G.: CLEF 2017 eHealth Evaluation Lab Overview. In: Jones et al. [305], pp. 291–303
239. Goeuriot, L., Mothe, J., Mulhem, P., Murtagh, F., SanJuan, E.: Overview of the CLEF 2016 Cultural Micro-blog Contextualization Workshop. In: Fuhr et al. [219], pp. 371–378
240. Goeuriot, L., Mulhem, P., Quénot, G., Schwab, D., Soulier, L., Di Nunzio, G.M., Galuščáková, P., García Seco de Herrera, A., Faggioli, G., Ferro, N. (eds.): Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Fifteenth International Conference of the CLEF Association (CLEF 2024) – Part 1. Lecture Notes in Computer Science (LNCS) 14958, Springer, Heidelberg, Germany (2024)
241. Goeuriot, L., Mulhem, P., Quénot, G., Schwab, D., Soulier, L., Di Nunzio, G.M., Galuščáková, P., García Seco de Herrera, A., Faggioli, G., Ferro, N. (eds.): Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Fifteenth International Conference of the CLEF Association (CLEF 2024) – Part 2. Lecture Notes in Computer Science (LNCS) 14959, Springer, Heidelberg, Germany (2024)
242. Goeuriot, L., Suominen, H., Kelly, L., Miranda-Escalada, A., Krallinger, M., Liu, Z., Pasi, G., Gonzalez Saez, G., Viviani, M., Xu, C.: Overview of the CLEF eHealth Evaluation Lab 2020. In: Arampatzis et al. [37], pp. 255–271
243. 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. [211], pp. 282–302
244. 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. [64], pp. 216–223
245. González-Sáez, G.N., Mulhem, P., Goeuriot, L.: Towards the Evaluation of Information Retrieval Systems on Evolving Datasets with Pivot Systems. In: Candan et al. [104], pp. 91–102
246. Gonzalo, J., Clough, P., Karlgren, J.: Overview of iCLEF 2008: Search Log Analysis for Multilingual Image Retrieval. In: Peters et al. [488], pp. 227–235
247. Gonzalo, J., Clough, P., Vallin, A.: Overview of the CLEF 2005 Interactive Track. In: Peters et al. [494], pp. 251–262
248. Gonzalo, J., Oard, D.W.: The CLEF 2002 Interactive Track. In: Peters et al. [484], pp. 372–382

249. Gonzalo, J., Oard, D.W.: iCLEF 2004 Track Overview: Pilot Experiments in Interactive Cross-Language Question Answering. In: Peters et al. [487], pp. 310–322
250. Gonzalo, J., Peinado, V., Clough, P., Karlgren, J.: Overview of iCLEF 2009: Exploring Search Behaviour in a Multilingual Folksonomy Environment. In: Peters et al. [497], pp. 13–20
251. Goodwin, T., Harabagiu, S.M.: The Impact of Belief Values on the Identification of Patient Cohorts. In: Forner et al. [211], pp. 155–166
252. 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. [420], pp. 78–90
253. Grubinger, M., Clough, P., Hanbury, A., Müller, H.: Overview of the ImageCLEF-photo 2007 Photographic Retrieval Task. In: Peters et al. [495], pp. 433–444
254. Guazzo, A., Trescato, I., Longato, E., Hazizaj, E., Dosso, D., Faggioli, G., Di Nunzio, G.M., Silvello, G., Vettoretti, M., Tavazzi, E., Roversi, C., Fariselli, P., Madeira, S.C., de Carvalho, M., Gromicho, M., Chiò, A., Manera, U., Dagliati, A., Birolo, G., Aidos, H., Di Camillo, B., Ferro, N.: Intelligent Disease Progression Prediction: Overview of iDPP@CLEF 2022. In: Barrón-Cedeño et al. [55], pp. 395–422
255. Gupta, P., Barrón-Cedeño, A., Rosso, P.: Cross-Language High Similarity Search Using a Conceptual Thesaurus. In: Catarci et al. [117], pp. 67–75
256. Gupta, S., Reda Bouadjenek, M., Robles-Kelly, A.: An Analysis of Logic Rule Dissemination in Sentiment Classifiers. In: Barrón-Cedeño et al. [55], pp. 118–124
257. Hagen, M., Glimm, C.: Supporting More-Like-This Information Needs: Finding Similar Web Content in Different Scenarios. In: Kanoulas et al. [312], pp. 50–61
258. Hall, M., Toms, E.: Building a Common Framework for IIR Evaluation. In: Forner et al. [211], pp. 17–28
259. Halvani, O., Graner, L.: Rethinking the Evaluation Methodology of Authorship Verification Methods. In: Bellot et al. [64], pp. 40–51
260. Hammarström, H.: Automatic Annotation of Bibliographical References for Descriptive Language Materials. In: Forner et al. [209], pp. 62–73
261. 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. [64], pp. 169–180
262. Hanbury, A., Müller, H.: Automated Component-Level Evaluation: Present and Future. In: Agosti et al. [10], pp. 124–135
263. 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. [117], pp. 24–29
264. Hansen, C., Hansen, C., Grue Simonsen, J., Lioma, C.: Fact Check-Worthiness Detection with Contrastive Ranking. In: Arampatzis et al. [37], pp. 124–130
265. Harman, D.K.: Information Retrieval Evaluation. Morgan & Claypool Publishers, USA (2011)
266. Harman, D.K., Voorhees, E.M. (eds.): TREC. Experiment and Evaluation in Information Retrieval. MIT Press, Cambridge (MA), USA (2005)
267. Harris, C.G., Xu, T.: The Importance of Visual Context Clues in Multimedia Translation. In: Forner et al. [209], pp. 107–118
268. 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. [64], pp. 224–230

269. He, J., Nguyen, D.Q., Akhondi, S.A., Druckenbrodt, C., Thorne, C., Hoessel, R., Afzal, Z., Zhai, Z., Fang, B., Yoshikawa, H., Albahem, A., Cavedon, L., Cohn, T., Baldwin, T., Verspoor, K.: Overview of ChEMU 2020: Named Entity Recognition and Event Extraction of Chemical Reactions from Patents. In: Arampatzis et al. [37], pp. 237–254
270. Hiemstra, D., Hauff, C.: MapReduce for Information Retrieval Evaluation: “Let’s Quickly Test This on 12 TB of Data”. In: Agosti et al. [10], pp. 64–69
271. 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. [138], pp. 88–100
272. 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 [203]
273. Hopfgartner, F., Kille, B., Lommatzsch, A., Plumbaum, T., Brodt, T., Heintz, T.: Benchmarking News Recommendations in a Living Lab. In: Kanoulas et al. [312], pp. 250–267
274. Huertas-Tato, J., Martín, A., Camacho, D.: Using Authorship Embeddings to Understand Writing Style in Social Media. In: Arampatzis et al. [36], pp. 60–71
275. 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)
276. 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. [138], pp. 425–432
277. Huurnink, B., Hofmann, K., de Rijke, M., Bron, M.: Validating Query Simulators: An Experiment Using Commercial Searches and Purchases. In: Agosti et al. [10], pp. 40–51
278. Imhof, M., Braschler, M.: Are Test Collections “Real”? Mirroring Real-World Complexity in IR Test Collections. In: Mothe et al. [420], pp. 241–247
279. Inches, G., Crestani, F.: Overview of the International Sexual Predator Identification Competition at PAN-2012. In: Forner et al. [210]
280. Ionescu, B., Müller, H., Drăgulinescu, A.M., Rückert, J., Ben Abacha, A., García Seco De Herrera, A., Bloch, L., Brüngel, R., Idrissi-Yaghir, A., Schäfer, H., Schmidt, C.S., Pakull, T.M.G., Damm, H., Bracke, B., Friedrich, C.M., Andrei, A.G., Prokopchuk, Y., Karapenka, D., Radzhabov, A., Kovalev, V., Macaire, C., Schwab, D., Lecouteux, B., Esperança-Rodier, E., Yim, W.W., Fu, Y., Sun, Z., Yetisgen, M., Xia, F., Hicks, S.A., Riegler, M.A., Thambawita, V., Storås, A., Halvorsen, P., Heinrich, M., Kiesel, J., Potthas, M., Stein, B.: Overview of the ImageCLEF 2024: Multimedia retrieval in medical applications. In: Goeuriot et al. [241]
281. Ionescu, B., Müller, H., Drăgulinescu, A.M., Yim, W.W., Ben Abacha, A., Snider, N., Adams, G., Yetisgen, M., Rückert, J., García Seco de Herrera, A., Friedrich, C.M., Bloch, L., Brüngel, R., Idrissi-Yaghir, A., Schäfer, H., Hicks, S.A., Riegler, M.A., Thambawita, V., Storås, A.M., Halvorsen, P., Papachrysos, N., Schöler, J., Jha, D., Andrei, A.G., Coman, I., Kovalev, V., Radzhabov, A., Prokopchuk, Y., Ştefan, L.D., Constantin, M.G., Dogariu, M., Deshayes, J., Popescu, A.: Overview of the ImageCLEF 2023: Multimedia Retrieval in Medical, Social Media and Internet Applications. In: Arampatzis et al. [36], pp. 370–396

282. Ionescu, B., Müller, H., Péteri, R., Abacha, A.B., Sarrouiti, M., Demner-Fushman, D., Hasan, S.A., Kozlovski, S., Liauchuk, V., Cid, Y.D., Kovalev, V., Pelka, O., García Seco de Herrera, A., Jacutprakart, J., Friedrich, C.M., Berari, R., Tauteanu, A., Fichou, D., Brie, P., Dogariu, M., Ștefan, L.D., Constantin, M.G., Chamberlain, J., Campello, A., Clark, A., Oliver, T.A., Moustahfid, H., Popescu, A., Deshayes-Chossart, J.: Overview of the ImageCLEF 2021: Multimedia Retrieval in Medical, Nature, Internet and Social Media Applications. In: Candan et al. [104], pp. 345–370
283. Ionescu, B., Müller, H., Péteri, R., Ben Abacha, A., Datla, V.V., Hasan, S.A., Demner-Fushman, D., Kozlovski, S., Liauchuk, V., Dicente Cid, Y., Kovalev, V., Pelka, O., Friedrich, C.M., García Seco de Herrera, A., Ninh, V.T., Le, T.K., Zhou, L., Piras, L., Riegler, M., Halvorsen, P., Tran, M.T., Lux, M., Gurrin, C., Dang-Nguyen, D.T., Chamberlain, J., Clark, A., Campello, A., Fichou, D., Berari, R., Brie, P., Dogariu, M., Stefan, L.D., Constantin, M.G.: Overview of the ImageCLEF 2020: Multimedia Retrieval in Medical, Lifelogging, Nature, and Internet Applications. In: Arampatzis et al. [37], pp. 311–341
284. 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. [138], pp. 358–386
285. Ionescu, B., Müller, H., Péteri, R., Rückert, J., Ben Abacha, A., García Seco de Herrera, A., Friedrich, C.M., Bloch, L., Brüngel, R., Idrissi-Yaghir, A., Schäfer, H., Kozlovski, S., Dicente Cid, Y., Kovalev, V., Ștefan, L.D., Constantin, M.G., Dogariu, M., Popescu, A., Deshayes-Chossart, J., Schindler, H., Chamberlain, J., Campello, A., Clark, A.: Overview of the ImageCLEF 2022: Multimedia Retrieval in Medical, Social Media and Nature Applications. In: Barrón-Cedeño et al. [55], pp. 541–564
286. 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. [305], pp. 315–337
287. 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. [64], pp. 309–334
288. 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. [219], pp. 174–186
289. Jijkoun, V., de Rijke, M.: Overview of WebCLEF 2007. In: Peters et al. [495], pp. 725–731
290. Jijkoun, V., de Rijke, M.: Overview of WebCLEF 2008. In: Peters et al. [488], pp. 787–793
291. Joly, A., Botella, C., Picek, L., Kahl, S., Goëau, H., Deneu, B., Marcos, D., Estopinan, J., Leblanc, C., Larcher, T., Chamidullin, R., Šulc, M., Hruz, M., Servajean, M., Glotin, H., Planqué, R., Vellinga, W.P., Klinck, H., Denton, T.,

- Eggel, I., Bonnet, P., Müller, H.: Overview of LifeCLEF 2023: Evaluation of AI Models for the Identification and Prediction of Birds, Plants, Snakes and Fungi. In: Arampatzis et al. [36], pp. 416–439
292. 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. [64], pp. 247–266
 293. 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. [138], pp. 387–401
 294. 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. [305], pp. 255–274
 295. 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 [203]
 296. 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. [312], pp. 229–249
 297. 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. [420], pp. 462–483
 298. Joly, A., Goëau, H., Kahl, S., Deneu, B., Servajean, M., Cole, E., Picek, L., Ruiz De Castañeda, R.L., Bolon, I., Durso, A., Lorieul, T., Botella, C., Glotin, H., Champ, J., Eggel, I., Vellinga, W.P., Bonnet, P., Müller, H.: Overview of LifeCLEF 2020: A System-Oriented Evaluation of Automated Species Identification and Species Distribution Prediction. In: Arampatzis et al. [37], pp. 342–363
 299. Joly, A., Goëau, H., Kahl, S., Picek, L., Lorieul, T., Cole, E., Deneu, B., Servajean, M., Durso, A., Bolon, I., Glotin, H., Planqué, R., Ruiz de Castañeda, R.L., Vellinga, W.P., Klinck, H., Denton, T., Eggel, I., Bonnet, P., Müller, H.: Overview of LifeCLEF 2021: An Evaluation of Machine-Learning Based Species Identification and Species Distribution Prediction. In: Candan et al. [104], pp. 371–393
 300. Joly, A., Goëau, H., Kahl, S., Picek, L., Lorieul, T., Cole, E., Deneu, B., Servajean, M., Durso, A., Glotin, H., Planqué, R., Vellinga, W.P., Navine, A., Klinck, H., Denton, T., Eggel, I., Bonnet, P., Šulc, M., Hruz, M.: Overview of LifeCLEF 2022: An Evaluation of Machine-Learning Based Species Identification and Species Distribution Prediction. In: Barrón-Cedeño et al. [55], pp. 257–285
 301. 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. [219], pp. 286–310
 302. Joly, A., Picek, L., Kahl, S., Goëau, H., Espitalier, V., Botella, C., Marcos, D., Estopinan, J., Leblanc, C., Larcher, T., Šulc, M., Hruz, M., Servajean, M., Glotin, H., Planqué, R., Vellinga, W.P., Klinck, H., Denton, T., Eggel, I., Bonnet, P., Müller, H.: Overview of LifeCLEF 2024: Challenges on Species Distribution Prediction and Identification. In: Goeuriot et al. [241]
 303. Jones, G.J.F.: 'Bout Sound and Vision: CLEF beyond Text Retrieval Tasks. In: Ferro and Peters [203]

304. Jones, G.J.F., Federico, M.: CLEF 2002 Cross-Language Spoken Document Retrieval Pilot Track Report. In: Peters et al. [484], pp. 446–457
305. 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)
306. Juola, P.: An Overview of the Traditional Authorship Attribution Subtask. In: Forner et al. [210]
307. Jürgens, J., Hansen, P., Womser-Hacker, C.: Going beyond CLEF-IP: The 'Reality' for Patent Searchers. In: Catarci et al. [117], pp. 30–35
308. Kalpathy-Cramer, J., Müller, H., Bedrick, S., Eggel, I., Garcia Seco de Herrera, A., Tsirikla, T.: Overview of the CLEF 2011 Medical Image Classification and Retrieval Tasks. In: Petras et al. [502]
309. Kamps, J., Koolen, M., Geva, S., Schenkel, R., SanJuan, E., Bogers, T.: From XML Retrieval to Semantic Search and Beyond. In: Ferro and Peters [203]
310. Kanoulas, E., Azzopardi, L.: CLEF 2017 Dynamic Search Evaluation Lab Overview. In: Jones et al. [305], pp. 361–366
311. Kanoulas, E., Azzopardi, L., Hui Yang, G.: Overview of the CLEF Dynamic Search Evaluation Lab 2018. In: Bellot et al. [64], pp. 362–371
312. 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)
313. 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. [305], pp. 173–185
314. Karan, M., Snajder, J.: Evaluation of Manual Query Expansion Rules on a Domain Specific FAQ Collection. In: Mothe et al. [420], pp. 248–253
315. Karimi, M.: SessionPrint: Accelerating kNN via Locality-Sensitive Hashing for Session-based News Recommendation. In: Goeuriot et al. [240]
316. Karisani, P., Oroumchian, F., Rahgozar, M.: Tweet Expansion Method for Filtering Task in Twitter. In: Mothe et al. [420], pp. 55–64
317. Karlgren, J.: Adopting Systematic Evaluation Benchmarks in Operational Settings. In: Ferro and Peters [203]
318. Karlgren, J.: How Lexical Gold Standards Have Effects On The Usefulness Of Text Analysis Tools For Digital Scholarship. In: Crestani et al. [138], pp. 178–184
319. Karlgren, J., Callin, J., Collins-Thompson, K., Cuba Gyllensten, A., Ekgren, A., Jurgens, D., Korhonen, A., Olsson, F., Sahlgren, M., Schütze, H.: Evaluating Learning Language Representations. In: Mothe et al. [420], pp. 254–260
320. Karlgren, J., Dürlich, L., Gogoulou, E., Guillou, L., Nivre, J., Sahlgren, M., Talmán, A., Zahra, S.: Overview of ELOQUENT 2024 — shared tasks for evaluating generative language model quality. In: Goeuriot et al. [241]
321. Karlgren, J., Gonzalo, J., Clough, P.: iCLEF 2006 Overview: Searching the Flickr WWW Photo-Sharing Repository. In: Peters et al. [486], pp. 186–194
322. Karlgren, J., Hedlund, T., Järvelin, K., Keskustalo, H., Kettunen, K.: The Challenges of Language Variation in Information Access. In: Ferro and Peters [203]
323. Karlsson, V., Herman, P., Karlgren, J.: Evaluating Categorisation in Real Life – An Argument Against Simple but Impractical Metrics. In: Fuhr et al. [219], pp. 221–226

324. Kavallieratou, E., del Blanco, C.R., Cuevas, C., García, N.: Interactive Learning-based Retrieval Technique for Visual Lifelogging. In: Crestani et al. [138], pp. 226–237
325. Kazlouski, S.: Tuberculosis CT Image Analysis Using Image Features Extracted by 3D Autoencoder. In: Arampatzis et al. [37], pp. 131–140
326. Keller, J., Breuer, T., Schaer, P.: Replicability Measures for Longitudinal Information Retrieval Evaluation. In: Goeuriot et al. [240]
327. Keller, J., Paul, L., Munz, M.: Evaluating Research Dataset Recommendations in a Living Lab. In: Barrón-Cedeño et al. [55], pp. 135–148
328. Kelly, L., Goeuriot, L., Suominen, H., Névéol, A., Palotti, J., Zuccon, G.: Overview of the CLEF eHealth Evaluation Lab 2016. In: Fuhr et al. [219], pp. 255–266
329. 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. [312], pp. 172–191
330. Kelly, L., Suominen, H., Goeuriot, L., Neves, M., Kanoulas, E., Li, D., Azzopardi, L., Spijker, R., Zuccon, G., Scells, H., Palotti, J.: Overview of the CLEF eHealth Evaluation Lab 2019. In: Crestani et al. [138], pp. 322–339
331. 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. [117], pp. 76–81
332. 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)
333. Khwileh, A., Ganguly, D., Jones, G.J.F.: An Investigation of Cross-Language Information Retrieval for User-Generated Internet Video. In: Mothe et al. [420], pp. 117–129
334. 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. [305], pp. 43–56
335. Ki Ng, Y., Fraser, D.J., Kassaie, B., Tompa, F.W.: Dowsing for Math Answers. In: Candan et al. [104], pp. 201–212
336. Kiesel, J., Çöltekin, C., Heinrich, H., Fröbe, M., Alshomary, M., De Longueville, B., Erjavec, T., Handke, N., Kopp, M., Ljubešić, N., Meden, K., Mirzakhmedova, N., Morkevičius, V., Reitis-Münstermann, T., Scharfbillig, M., Stefanovitch, N., Wachsmuth, H., Potthast, M., Stein, B.: Overview of Touché 2024: Argumentation Systems. In: Goeuriot et al. [241]
337. Kiesel, J., Gohsen, M., Mirzakhmedova, N., Hagen, M., Stein, B.: Who Will Evaluate the Evaluators? Exploring the Gen-IR User Simulation Space. In: Goeuriot et al. [240]
338. 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 News-REEL’16: Multi-dimensional Evaluation of Real-Time Stream-Recommendation Algorithms. In: Fuhr et al. [219], pp. 311–331
339. 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. [420], pp. 497–517
340. 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. [211], pp. 179–191

341. Kliegr, T., Kuchar, J.: Benchmark of Rule-Based Classifiers in the News Recommendation Task. In: Mothe et al. [420], pp. 130–141
342. Kluck, M.: The Domain-Specific Track in CLEF 2004: Overview of the Results and Remarks on the Assessment Process. In: Peters et al. [487], pp. 260–270
343. Kluck, M., Gey, F.C.: The Domain-Specific Task of CLEF – Specific Evaluation Strategies in Cross-Language Information Retrieval. In: Peters [480], pp. 48–56
344. Kluck, M., Stempfhuber, M.: Domain-Specific Track CLEF 2005: Overview of Results and Approaches, Remarks on the Assessment Analysis. In: Peters et al. [494], pp. 212–221
345. Kocher, M., Savoy, J.: Author Clustering with an Adaptive Threshold. In: Jones et al. [305], pp. 186–198
346. Koitka, S., Friedrich, C.M.: Optimized Convolutional Neural Network Ensembles for Medical Subfigure Classification. In: Jones et al. [305], pp. 57–68
347. Konstantinou, A., Chatzakou, D., Theodosiadou, O., Tsirikla, T., Vrochidis, S., Kompatsiaris, I.: Trend Detection in Crime-Related Time Series with Change Point Detection Methods. In: Arampatzis et al. [36], pp. 72–84
348. 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. [219], pp. 351–370
349. 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. [420], pp. 545–564
350. 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. [210]
351. Koops, H.V., Van Balen, J., Wiering, F.: Automatic Segmentation and Deep Learning of Bird Sounds. In: Mothe et al. [420], pp. 261–267
352. 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. [305], pp. 367–376
353. Kosmajac, D., Keselj, V.: Twitter User Profiling: Bot and Gender Identification. In: Arampatzis et al. [37], pp. 141–153
354. Kosmopoulos, A., Paliouras, G., Androutsopoulos, I.: The Effect of Dimensionality Reduction on Large Scale Hierarchical Classification. In: Kanoulas et al. [312], pp. 160–171
355. Kougia, V., Pavlopoulos, J., Androutsopoulos, I.: Medical Image Tagging by Deep Learning and Retrieval. In: Arampatzis et al. [37], pp. 154–166
356. 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. [209], pp. 74–82
357. Kurimo, M., Creutz, M., Varjokallio, M.: Morpho Challenge Evaluation Using a Linguistic Gold Standard. In: Peters et al. [495], pp. 864–872
358. Kurimo, M., Turunen, V.T., Varjokallio, M.: Overview of Morpho Challenge 2008. In: Peters et al. [488], pp. 951–966
359. Kurimo, M., Virpioja, S., Turunen, V.T., Blackwood, G.W., Byrne, W.: Overview and Results of Morpho Challenge 2009. In: Peters et al. [489], pp. 587–597
360. Kürsten, J., Eibl, M.: Comparing IR System Components Using Beanplots. In: Catarci et al. [117], pp. 136–137
361. Kvist, M., Velupillai, S.: SCAN: A Swedish Clinical Abbreviation Normalizer - Further Development and Adaptation to Radiology. In: Kanoulas et al. [312], pp. 62–73

362. de L. Pertile, S., Pereira Moreira, V.: A Test Collection to Evaluate Plagiarism by Missing or Incorrect References. In: Catarci et al. [117], pp. 141–143
363. de L. Pertile, S., Rosso, P., Pereira Moreira, V.: Counting Co-occurrences in Citations to Identify Plagiarised Text Fragments. In: Forner et al. [211], pp. 150–154
364. Lagopoulos, A., Anagnostou, A., Minas, A., Tsoumakas, G.: Learning-to-Rank and Relevance Feedback for Literature Appraisal in Empirical Medicine. In: Bellot et al. [64], pp. 52–63
365. Lai, M., Stranisci, M.A., Bosco, C., Damiano, R., Patti, V.: Analysing Moral Beliefs for Detecting Hate Speech Spreaders on Twitter. In: Barrón-Cedeño et al. [55], pp. 149–161
366. Lai, M., Tambuscio, M., Patti, V., Ruffo, G., Rosso, P.: Extracting Graph Topological Information and Users’ Opinion. In: Jones et al. [305], pp. 112–118
367. 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. [138], pp. 101–113
368. Larsen, B.: The Scholarly Impact of CLEF 2010-2017. In: Ferro and Peters [203], pp. 547–554
369. 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. [488], pp. 906–917
370. Larson, M., Newman, E., Jones, G.J.F.: Overview of VideoCLEF 2009: New Perspectives on Speech-Based Multimedia Content Enrichment. In: Peters et al. [497], pp. 354–368
371. Lasseck, M.: Towards Automatic Large-Scale Identification of Birds in Audio Recordings. In: Mothe et al. [420], pp. 364–375
372. Leiva, L.A., Villegas, M., Paredes, R.: Relevant Clouds: Leveraging Relevance Feedback to Build Tag Clouds for Image Search. In: Forner et al. [211], pp. 143–149
373. Leong, C.W., Hassan, S., Ruiz, M.E., Rada, M.: Improving Query Expansion for Image Retrieval via Saliency and Picturability. In: Forner et al. [209], pp. 137–142
374. Lestari Paramita, M., Sanderson, M., Clough, P.: Diversity in Photo Retrieval: Overview of the ImageCLEFPhoto Task 2009. In: Peters et al. [497], pp. 45–59
375. Li, P., Jiang, X., Kambhamettu, C., Shatkay, H.: Segmenting Compound Biomedical Figures into Their Constituent Panels. In: Jones et al. [305], pp. 199–210
376. Li, W., Jones, G.J.F.: Enhancing Medical Information Retrieval by Exploiting a Content-Based Recommender Method. In: Mothe et al. [420], pp. 142–153
377. Li, Y., Fang, B., He, J., Yoshikawa, H., Akhondi, S.A., Druckenbrodt, C., Thorne, C., Afzal, Z., Zhai, Z., Baldwin, T., Verspoor, K.: Overview of ChEMU 2021: Reaction Reference Resolution and Anaphora Resolution in Chemical Patents. In: Candan et al. [104], pp. 292–307
378. Li, Y., Fang, B., He, J., Yoshikawa, H., Akhondi, S.A., Druckenbrodt, C., Thorne, C., Afzal, Z., Zhai, Z., Baldwin, T., Verspoor, K.: Overview of ChEMU 2022 Evaluation Campaign: Information Extraction in Chemical Patents. In: Barrón-Cedeño et al. [55], pp. 521–540
379. Linhares Pontes, E., Huet, S., Torres-Moreno, J.M.: Microblog Contextualization: Advantages and Limitations of a Multi-sentence Compression Approach. In: Bellot et al. [64], pp. 181–190
380. Lipani, A., Piroi, F., Andersson, L., Hanbury, A.: An Information Retrieval Ontology for Information Retrieval Nanopublications. In: Kanoulas et al. [312], pp. 44–49

381. 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. [305], pp. 119–124
382. 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. [138], pp. 238–247
383. 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. [305], pp. 239–254
384. Lommatzsch, A., Werner, S.: Optimizing and Evaluating Stream-Based News Recommendation Algorithms. In: Mothe et al. [420], pp. 376–388
385. Longhin, F., Guazzo, A., Longato, E., Ferro, N., Di Camillo, B.: DAVI: a Dataset for Automatic Variant Interpretation. In: Arampatzis et al. [36], pp. 85–96
386. Loponen, A., Järvelin, K.: A Dictionary- and Corpus-Independent Statistical Lemmatizer for Information Retrieval in Low Resource Languages. In: Agosti et al. [10], pp. 3–14
387. Losada, D.E., Crestani, F.: A Test Collection for Research on Depression and Language Use. In: Fuhr et al. [219], pp. 28–39
388. Losada, D.E., Crestani, F., Parapar, J.: eRISK 2017: CLEF Lab on Early Risk Prediction on the Internet: Experimental Foundations. In: Jones et al. [305], pp. 346–360
389. Losada, D.E., Crestani, F., Parapar, J.: Overview of eRisk: Early Risk Prediction on the Internet. In: Bellot et al. [64], pp. 343–361
390. Losada, D.E., Crestani, F., Parapar, J.: Overview of eRisk 2019: Early Risk Prediction on the Internet. In: Crestani et al. [138], pp. 340–357
391. Losada, D.E., Crestani, F., Parapar, J.: Overview of eRisk 2020: Early Risk Prediction on the Internet. In: Arampatzis et al. [37], pp. 272–287
392. Mackie, S., McCreadie, R., Macdonald, C., Ounis, I.: Comparing Algorithms for Microblog Summarisation. In: Kanoulas et al. [312], pp. 153–159
393. Magdy, W., Jones, G.J.F.: Examining the Robustness of Evaluation Metrics for Patent Retrieval with Incomplete Relevance Judgements. In: Agosti et al. [10], pp. 82–93
394. 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. [486], pp. 223–256
395. 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. [485], pp. 471–486
396. 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. [487], pp. 371–391
397. 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. [489], pp. 508–517
398. 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. [488], pp. 808–821
399. Mandl, T., Di Nunzio, G.M., Schulz, J.M.: LogCLEF 2010: the CLEF 2010 Multilingual Logfile Analysis Track Overview. In: Braschler et al. [96]

400. 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. [495], pp. 745–772
401. Manotumruksa, J., Macdonald, C., Ounis, I.: Predicting Contextually Appropriate Venues in Location-Based Social Networks. In: Fuhr et al. [219], pp. 96–109
402. Mansouri, B., Novotný, V., Agarwal, A., Oard, D.W., Zanibbi, R.: Overview of ARQMath-3 (2022): Third CLEF Lab on Answer Retrieval for Questions on Math. In: Barrón-Cedeño et al. [55], pp. 286–310
403. Mansouri, B., Zanibbi, R., Oard, D.W., Agarwal, A.: Overview of ARQMath-2 (2021): Second CLEF Lab on Answer Retrieval for Questions on Math. In: Candan et al. [104], pp. 215–238
404. Martínez-Castaño, R., Htait, A., Azzopardi, L., Moshfeghi, Y.: BERT-Based Transformers for Early Detection of Mental Health Illnesses. In: Candan et al. [104], pp. 189–200
405. Martínez-Gómez, J., García-Varea, I., Caputo, B.: Overview of the ImageCLEF 2012 Robot Vision Task. In: Forner et al. [210]
406. Mayfield, J., Lawrie, D., McNamee, P., Oard, D.W.: Building a Cross-Language Entity Linking Collection in Twenty-One Languages. In: Forner et al. [209], pp. 3–13
407. McCreadie, R., Macdonald, C., Ounis, I., Brassey, J.: A Study of Personalised Medical Literature Search. In: Kanoulas et al. [312], pp. 74–85
408. McMin, A.J., Jose, J.M.: Real-Time Entity-Based Event Detection for Twitter. In: Mothe et al. [420], pp. 65–77
409. 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. [305], pp. 211–223
410. Merker, J.H., Merker, L., Bondarenko, A.: The Impact of Web Search Result Quality on Decision-Making. In: Goeuriot et al. [240]
411. Michail, A., Andermatt, P.S., Fankhauser, T.: SimpleText Best of Labs in CLEF-2023: Scientific Text Simplification Using Multi-Prompt Minimum Bayes Risk Decoding. In: Goeuriot et al. [240]
412. 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. [64], pp. 203–215
413. Mirsarraf, M.R., Dehghani, N.: A Dependency-Inspired Semantic Evaluation of Machine Translation Systems. In: Forner et al. [211], pp. 71–74
414. Mitrovic, S., Müller, H.: Summarizing Citation Contexts of Scientific Publications. In: Mothe et al. [420], pp. 154–165
415. Mohtaj, S. and Möller, S.: The Impact of Pre-processing on the Performance of Automated Fake News Detection. In: Barrón-Cedeño et al. [55], pp. 93–102
416. Molina, A., SanJuan, E., Torres-Moreno, J.M.: A Turing Test to Evaluate a Complex Summarization Task. In: Forner et al. [211], pp. 75–80
417. Molina, S., Mothe, J., Roques, D., Tanguy, L., Ullah, M.Z.: IRIT-QFR: IRIT Query Feature Resource. In: Jones et al. [305], pp. 69–81
418. Morante, R., Daelemans, W.: Overview of the QA4MRE Pilot Task: Annotating Modality and Negation for a Machine Reading Evaluation. In: Petras et al. [502]
419. Moreno, R., Huang, 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. [305], pp. 125–131

420. 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)
421. Mulhem, P., Goeuriot, L., Dogra, N., Amer, N.O.: TimeLine Illustration Based on Microblogs: When Diversification Meets Metadata Re-ranking. In: Jones et al. [305], pp. 224–235
422. Müller, H., Clough, P., Deselaers, T., Caputo, B. (eds.): ImageCLEF – Experimental Evaluation in Visual Information Retrieval. Springer-Verlag, Heidelberg, Germany (2010)
423. 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. [486], pp. 595–608
424. 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. [495], pp. 472–491
425. Müller, H., Hanbury, A.: EaaS: Evaluation-as-a-Service and Experiences from the VISCERAL Project. In: Ferro and Peters [203]
426. 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. [210]
427. 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. [497], pp. 72–84
428. 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. [96]
429. Müller, H., Kalpathy-Cramer, J., Garcia Seco de Herrera, A.: Experiences from the ImageCLEF Medical Retrieval and Annotation Tasks. In: Ferro and Peters [203]
430. 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. [488], pp. 512–522
431. Muraier, B., Specht, G.: Generating Cross-Domain Text Corpora from Social Media Comments. In: Crestani et al. [138], pp. 114–125
432. Nakov, P., Barrón-Cedeño, A., Da San Martino, G., Alam, F., Struß, J.M., Mandl, T., Míguez, R., Caselli, T., Kutlu, M., Zaghoulani, W., Li, C., Shaar, S., Kishore Shahi, G., Mubarak, H., Nikolov, A., Babulkov, N., Selim Kartal, Y., Wiegand, M., Siegel, M., Köhler, J.: Overview of the CLEF-2022 CheckThat! Lab on Fighting the COVID-19 Infodemic and Fake News Detection. In: Barrón-Cedeño et al. [55], pp. 495–520
433. 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. [64], pp. 372–387
434. Nakov, P., Da San Martino, G., Elsayed, T., Barrón-Cedeño, A., Míguez, R., Shaar, S., Alam, F., Haouari, F., Hasanain, M., Mansour, W., Hamdan, B., Ali, Z.S., Babulkov, N., Nikolov, A., Shahi, G.K., Struß, J.M., Mandl, T., Kutlu, M.,

- Kartal, Y.S.: Overview of the CLEF-2021 CheckThat! Lab on Detecting Check-Worthy Claims, Previously Fact-Checked Claims, and Fake News. In: Candan et al. [104], pp. 264–291
435. 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)
436. 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)
437. Nentidis, A., Katsimpras, G., Krithara, A., Lima López, S., Farré-Maduell, E., Gasco, L., Krallinger, M., Paliouras, G.: Overview of BioASQ 2023: The Eleventh BioASQ Challenge on Large-Scale Biomedical Semantic Indexing and Question Answering. In: Arampatzis et al. [36], pp. 227–250
438. Nentidis, A., Katsimpras, G., Krithara, A., Lima López, S., Farré-Maduell, E., Krallinger, M., Loukachevitch, N., Davydova, V., Tutubalina, E., Paliouras, G.: Overview of BioASQ 2024: The Twelfth BioASQ Challenge on Large-Scale Biomedical Semantic Indexing and Question Answering. In: Goeriot et al. [241]
439. Nentidis, A., Katsimpras, G., Vadorou, E., Krithara, A., Gasco, L., Krallinger, M., Paliouras, G.: Overview of BioASQ 2021: The Ninth BioASQ Challenge on Large-Scale Biomedical Semantic Indexing and Question Answering. In: Candan et al. [104], pp. 239–263
440. Nentidis, A., Katsimpras, G., Vadorou, E., Krithara, A., Miranda-Escalada, A., Gasco, L., Krallinger, M., Paliouras, G.: Overview of BioASQ 2022: The Tenth BioASQ Challenge on Large-Scale Biomedical Semantic Indexing and Question Answering. In: Barrón-Cedeño et al. [55], pp. 337–361
441. Nentidis, A., Krithara, A., Bougiatiotis, K., Krallinger, M., Rodríguez Penagos, C., Villegas, M., Paliouras, G.: Overview of BioASQ 2020: The Eighth BioASQ Challenge on Large-Scale Biomedical Semantic Indexing and Question Answering. In: Arampatzis et al. [37], pp. 194–214
442. Nicolson, A., Dowling, J., Koopman, B.: ImageCLEF 2021 Best of Labs: The Curious Case of Caption Generation for Medical Images. In: Barrón-Cedeño et al. [55], pp. 175–189
443. Ningtyas, A.M., El-Ebshihy, A., Budi Herwanto, G., Piroi, F., Hanbury, A.: Leveraging Wikipedia Knowledge for Distant Supervision in Medical Concept Normalization. In: Barrón-Cedeño et al. [55], pp. 33–47
444. Ningtyas, A.M., El-Ebshihy, A., Piroi, F., Hanbury, A.: Improving Laypeople Familiarity with Medical Terms by Informal Medical Entity Linking. In: Goeriot et al. [240]
445. Nordlie, R., Pharo, N.: Seven Years of INEX Interactive Retrieval Experiments - Lessons and Challenges. In: Catarci et al. [117], pp. 13–23
446. Nowak, S., Dunker, P.: Overview of the CLEF 2009 Large-Scale Visual Concept Detection and Annotation Task. In: Peters et al. [497], pp. 94–109
447. Nowak, S., Huiskes, M.J.: New Strategies for Image Annotation: Overview of the Photo Annotation Task at ImageCLEF 2010. In: Braschler et al. [96]
448. Nowak, S., Nagel, K., Liebetrau, J.: The CLEF 2011 Photo Annotation and Concept-based Retrieval Tasks. In: Petras et al. [502]
449. Oard, D.W., Gonzalo, J.: The CLEF 2001 Interactive Track. In: Peters et al. [483], pp. 308–319
450. Oard, D.W., Gonzalo, J.: The CLEF 2003 Interactive Track. In: Peters et al. [485], pp. 425–434

451. 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. [486], pp. 744–758
452. Oh, H.S., Jung, Y., Kim, K.Y.: A Multiple-Stage Approach to Re-ranking Medical Documents. In: Mothe et al. [420], pp. 166–177
453. Olvera-Lobo, M.D., Gutiérrez-Artacho, J.: Multilingual Question-Answering System in Biomedical Domain on the Web: An Evaluation. In: Forner et al. [209], pp. 83–88
454. Orio, N., Liem, C.C.S., Peeters, G., Schedl, M.: MusiClef: Multimodal Music Tagging Task. In: Catarci et al. [117], pp. 36–41
455. Orio, N., Rizo, D.: Overview of MusiCLEF 2011. In: Petras et al. [502]
456. 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. [219], pp. 110–122
457. Otterbacher, J.: Addressing Social Bias in Information Retrieval. In: Bellot et al. [64], pp. 121–127
458. Pääkkönen, T., Järvelin, K., Kekäläinen, J., Keskustalo, H., Baskaya, F., Maxwell, D., Azzopardi, L.: Exploring Behavioral Dimensions in Session Effectiveness. In: Mothe et al. [420], pp. 178–189
459. 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. [219], pp. 40–53
460. Parapar, J., Martín-Rodilla, P., Losada, D.E., Crestani, F.: Overview of eRisk 2021: Early Risk Prediction on the Internet. In: Candan et al. [104], pp. 324–344
461. Parapar, J., Martín-Rodilla, P., Losada, D.E., Crestani, F.: Overview of eRisk 2022: Early Risk Prediction on the Internet. In: Barrón-Cedeño et al. [55], pp. 233–256
462. Parapar, J., Martín-Rodilla, P., Losada, D.E., Crestani, F.: Overview of eRisk 2023: Early Risk Prediction on the Internet. In: Arampatzis et al. [36], pp. 294–315
463. Parapar, J., Martín-Rodilla, P., Losada, D.E., Crestani, F.: Overview of eRisk 2024: Early Risk Prediction on the Internet. In: Goeuriot et al. [241]
464. Parks, M., Karlgren, J., Stymne, S.: Plausibility Testing for Lexical Resources. In: Jones et al. [305], pp. 132–137
465. 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. [64], pp. 335–342
466. 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. [138], pp. 417–424
467. 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. [305], pp. 338–345
468. Pasin, A., Ferrari Dacrema, M., Cremonesi, P., Ferro, N.: qCLEF: a Proposal to Evaluate Quantum Annealing for Information Retrieval and Recommender Systems. In: Arampatzis et al. [36], pp. 97–108
469. Pasin, A., Ferrari Dacrema, M., Cremonesi, P., Ferro, N.: Overview of Quantum-CLEF 2024: The Quantum Computing Challenge for Information Retrieval and Recommender Systems at CLEF. In: Goeuriot et al. [241]

470. 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. [495], pp. 674–686
471. 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. [219], pp. 150–161
472. 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. [96]
473. 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. [489], pp. 174–196
474. 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. [502]
475. 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. [211], pp. 303–320
476. 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. [210]
477. 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 [203]
478. Peñas, A., Unger, C., Ngonga Ngomo, A.C.a.: Overview of CLEF Question Answering Track 2014. In: Kanoulas et al. [312], pp. 300–306
479. Peñas, A., Unger, C., Paliouras, P., Kakadiaris, I.A.: Overview of the CLEF Question Answering Track 2015. In: Mothe et al. [420], pp. 539–544
480. 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)
481. Peters, C.: Introduction. In: Cross-Language Information Retrieval and Evaluation: Workshop of Cross-Language Evaluation Forum (CLEF 2000) [480], pp. 1–6
482. Peters, C., Braschler, M., Clough, P.: Multilingual Information Retrieval. Springer-Verlag, Heidelberg, Germany (2011)
483. 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)
484. 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)
485. 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)
486. 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)
487. 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)
488. Peters, C., Deselaers, T., Ferro, N., Gonzalo, J., Jones, G.J.F., Kurimo, M., Mandl, T., Peñas, A. (eds.): Evaluating Systems for Multilingual and Multimodal 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)
489. 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)
490. 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)
491. 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)
492. 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)
493. 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)
494. 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)
495. 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)
496. 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)
497. 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)
498. Petras, V., Baerisch, S.: The Domain-Specific Track at CLEF 2008. In: Peters et al. [488], pp. 186–198

499. Petras, V., Baerisch, S., Stempfhuber, M.: The Domain-Specific Track at CLEF 2007. In: Peters et al. [495], pp. 160–173
500. 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. [211], pp. 192–211
501. 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. [210]
502. 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)
503. Petras, V., Lüscho, A., Ramthun, R., Stiller, J., España-Bonet, C., Henning, S.: Query or Document Translation for Academic Search – What’s the Real Difference? In: Arampatzis et al. [37], pp. 28–42
504. Piras, L., Caputo, B., Dang-Nguyen, D.T., Riegler, M., Halvorsen, P.: Image Retrieval Evaluation in Specific Domains. In: Ferro and Peters [203]
505. Piroi, F.: CLEF-IP 2010: Retrieval Experiments in the Intellectual Property Domain. In: Braschler et al. [96]
506. Piroi, F., Hanbury, A.: Multilingual Patent Text Retrieval Evaluation: CLEF-IP. In: Ferro and Peters [203]
507. Piroi, F., Lupu, M., Hanbury, A.: Effects of Language and Topic Size in Patent IR: An Empirical Study. In: Catarci et al. [117], pp. 54–66
508. Piroi, F., Lupu, M., Hanbury, A.: Overview of CLEF-IP 2013 Lab - Information Retrieval in the Patent Domain. In: Forner et al. [211], pp. 232–249
509. 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. [210]
510. Piroi, F., Lupu, M., Hanbury, A., Zenz, V.: CLEF-IP 2011: Retrieval in the Intellectual Property Domain. In: Petras et al. [502]
511. Plaza, L., Carrillo-de Albornoz, J., Morante, R., Amigó, E., Gonzalo, J., Spina, D., Rosso, P.: Overview of EXIST 2023 – Learning with Disagreement for Sexism Identification and Characterization. In: Arampatzis et al. [36], pp. 316–342
512. Plaza, L., Carrillo-de Albornoz, J., Ruiz, V., Maeso, B., Chulvi, B., Rosso, P., Amigó, E., Gonzalo, J., Morante, R., Spina, D.: Overview of EXIST 2024 – Learning with Disagreement for Sexism Identification and Characterization in Tweets and Memes. In: Goeriot et al. [241]
513. Popescu, A., Tsikrika, T., Kludas, J.: Overview of the Wikipedia Retrieval Task at ImageCLEF 2010. In: Braschler et al. [96]
514. Potha, N., Stamatatos, E.: An Improved Impostors Method for Authorship Verification. In: Jones et al. [305], pp. 138–144
515. Potthast, M., Barrón-Cedeño, A., Eiselt, A., Stein, B., Rosso, P.: Overview of the 2nd International Competition on Plagiarism Detection. In: Braschler et al. [96]
516. Potthast, M., Eiselt, A., Barrón-Cedeño, A., Stein, B., Rosso, P.: Overview of the 3rd International Competition on Plagiarism Detection. In: Petras et al. [502]
517. 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. [210]
518. 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. [312], pp. 268–299

519. Potthast, M., Gollub, T., Wiegmann, M., Stein, b.: TIRA Integrated Research Architecture. In: Ferro and Peters [203], pp. 123–160
520. Potthast, M., Holfeld, T.: Overview of the 2nd International Competition on Wikipedia Vandalism Detection. In: Petras et al. [502]
521. 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. [305], pp. 275–290
522. Potthast, M., Stein, B., Holfeld, T.: Overview of the 1st International Competition on Wikipedia Vandalism Detection. In: Braschler et al. [96]
523. Pradel, C., Sileo, D., Rodrigo, A., Peñas, A., Agirre, E.: Question Answering When Knowledge Bases are Incomplete. In: Arampatzis et al. [37], pp. 43–54
524. Prasetyo Putri, D.G., Viviani, M., Pasi, G.: A Multi-Task Learning Model for Multidimensional Relevance Assessment. In: Candan et al. [104], pp. 103–115
525. Pritsos, D.A., Stamatatos, E.: The Impact of Noise in Web Genre Identification. In: Mothe et al. [420], pp. 268–273
526. Pronobis, A., Fornoni, M., Christensen, H.I., Caputo, B.: The Robot Vision Track at ImageCLEF 2010. In: Braschler et al. [96]
527. Pronobis, A., Xing, L., Caputo, B.: Overview of the CLEF 2009 Robot Vision Track. In: Peters et al. [497], pp. 110–119
528. 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. [305], pp. 104–111
529. 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. [138], pp. 248–259
530. Rangel Pardo, F., Rosso, P.: On the Multilingual and Genre Robustness of EmoGraphs for Author Profiling in Social Media. In: Mothe et al. [420], pp. 274–280
531. 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. [211], pp. 353–367
532. Rekabsaz, N., Lupu, M.: A Real-World Framework for Translator as Expert Retrieval. In: Kanoulas et al. [312], pp. 141–152
533. Reusch, A., Thiele, M., Lehner, W.: Transformer-Encoder-Based Mathematical Information Retrieval. In: Barrón-Cedeño et al. [55], pp. 175–189
534. de Rijke, M., Balog, K., Bogers, T., van den Bosch, A.: On the Evaluation of Entity Profiles. In: Agosti et al. [10], pp. 94–99
535. Roda, G., Tait, J., Piroi, F., Zenz, V.: CLEF-IP 2009: Retrieval Experiments in the Intellectual Property Domain. In: Peters et al. [489], pp. 385–409
536. Rodrigo, A., Peñas, A., Verdejo, M.F.: Overview of the Answer Validation Exercise 2008. In: Peters et al. [488], pp. 296–313
537. Roller, R., Stevenson, M.: Self-supervised Relation Extraction Using UMLS. In: Kanoulas et al. [312], pp. 116–127
538. 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 [203]
539. Rosso, P., Rangel, F., Potthast, M., Stamatatos, E., Tschuggnall, M., Stein, B.: Overview of PAN'16. In: Fuhr et al. [219], pp. 332–350

540. 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)
541. Sabetghadam, S., Bierig, R., Rauber, A.: A Hybrid Approach for Multi-faceted IR in Multimodal Domain. In: Kanoulas et al. [312], pp. 86–97
542. Sakai, T.: How to Run an Evaluation Task. In: Ferro and Peters [203], pp. 71–102
543. Sakai, T., Oard, D.W., Kando, N. (eds.): Evaluating Information Retrieval and Access Tasks – NTCIR's Legacy of Research Impact, The Information Retrieval Series, vol. 43. Springer International Publishing, Germany (2021)
544. Sakhovskiy, A., Semenova, N., Kadurin, A., Tutubalina, E.: Graph-Enriched Biomedical Entity Representation Transformer. In: Arampatzis et al. [36], pp. 109–120
545. Saleh, S., Pecina, P.: Reranking Hypotheses of Machine-Translated Queries for Cross-Lingual Information Retrieval. In: Fuhr et al. [219], pp. 54–68
546. Samuel, J.: Analyzing and Visualizing Translation Patterns of Wikidata Properties. In: Bellot et al. [64], pp. 128–134
547. Sánchez-Cortés, D., Burdisso, S., Villatoro-Tello, E., Motlicek, P.: Mapping the Media Landscape: Predicting Factual Reporting and Political Bias Through Web Interactions. In: Goeuriot et al. [240]
548. Sánchez-Junquera, J., Villaseñor-Pineda, L., Montes-y Gómez, M., Rosso, P.: Character N-Grams for Detecting Deceptive Controversial Opinions. In: Bellot et al. [64], pp. 135–140
549. 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. [420], pp. 402–413
550. 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. [305], pp. 145–151
551. SanJuan, E., Moriceau, V., Tannier, X., Bellot, P., Mothe, J.: Overview of the INEX 2012 Tweet Contextualization Track. In: Forner et al. [210]
552. Santos, D., Cabral, L.M.: GikiCLEF: Expectations and Lessons Learned. In: Peters et al. [489], pp. 212–222
553. Ramos dos Santos, W., Paraboni, I.: Personality facets recognition from text. In: Crestani et al. [138], pp. 185–190
554. Sarvazyan, A.M., González, J.A., Rosso, P., Franco-Salvador, M.: Supervised Machine-Generated Text Detectors: Family and Scale Matters. In: Arampatzis et al. [36], pp. 121–132
555. Savenkov, D., Braslavski, P., Lebedev, M.: Search Snippet Evaluation at Yandex: Lessons Learned and Future Directions. In: Forner et al. [209], pp. 14–25
556. Savoy, J., Braschler, M.: Lessons Learnt from Experiments on the Ad-Hoc Multilingual Test Collections at CLEF. In: Ferro and Peters [203]
557. Sawinski, M., Wecel, K., Ksieźniak, E.: Under-sampling strategies for better transformer-based classifications models. In: Goeuriot et al. [240]
558. Schaer, P.: Better than Their Reputation? On the Reliability of Relevance Assessments with Students. In: Catarci et al. [117], pp. 124–135
559. Schaer, P., Breuer, T., Castro, L.J., Wolff, B., Schaible, J., Tavakolpoursaleh, N.: Overview of LiLAS 2021 – Living Labs for Academic Search. In: Candan et al. [104], pp. 394–418

560. 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. [219], pp. 227–233
561. Schaer, P., Neumann, M.: Enriching Existing Test Collections with OXPath. In: Jones et al. [305], pp. 152–158
562. Schaer, P., Schaible, J., Jael García Castro, L.: Overview of LiLAS 2020 - Living Labs for Academic Search. In: Arampatzis et al. [37], pp. 364–371
563. 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)
564. Schreieder, T., Braker, J.: Touché 2022 Best of Labs: Neural Image Retrieval for Argumentation. In: Arampatzis et al. [36], pp. 186–197
565. Schubotz, M., Krämer, L., Meuschke, N., Hamborg, F., Gipp, B.: Evaluating and Improving the Extraction of Mathematical Identifier Definitions. In: Jones et al. [305], pp. 82–94
566. Schuth, A., Balog, K., Kelly, L.: Overview of the Living Labs for Information Retrieval Evaluation (LL4IR) CLEF Lab 2015. In: Mothe et al. [420], pp. 484–496
567. Schuth, A., Marx, M.: Evaluation Methods for Rankings of Facetvalues for Faceted Search. In: Forner et al. [209], pp. 131–136
568. Shahshahani, M.S., Kamps, J.: Argument Retrieval from Web. In: Arampatzis et al. [37], pp. 75–81
569. Shen, W., Nie, J.Y.: Is Concept Mapping Useful for Biomedical Information Retrieval. In: Mothe et al. [420], pp. 281–286
570. Shepeleva, N., Balog, K.: Towards an Understanding of Transactional Tasks. In: Fuhr et al. [219], pp. 234–240
571. Sherkat, E., Velcin, J., Milios, E.E.: Fast and Simple Deterministic Seeding of KMeans for Text Document Clustering. In: Bellot et al. [64], pp. 76–88
572. 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. [138], pp. 191–197
573. Sierek, T., Hanbury, A.: Using Health Statistics to Improve Medical and Health Search. In: Mothe et al. [420], pp. 287–292
574. Sigurbjörnsson, B., Kamps, J., de Rijke, M.: Overview of WebCLEF 2005. In: Peters et al. [494], pp. 810–824
575. 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)
576. Šimsa, S., Uříčář, M., Šulc, M., Patel, Y., Hamdi, A., Kocián, M., Skalický, M., Matas, J., Doucet, A., Coustaty, M., Karatzas, D.: Overview of DocILE 2023: Document Information Localization and Extraction. In: Arampatzis et al. [36], pp. 276–293
577. Singh, G., Mantrach, A., Silvestri, F.: Improving Profiles of Weakly-Engaged Users: With Applications to Recommender Systems. In: Fuhr et al. [219], pp. 123–136
578. Skalický, M., Šimsa, S., Uříčář, M., Šulc, M.: Business Document Information Extraction: Towards Practical Benchmarks. In: Barrón-Cedeño et al. [55], pp. 105–117
579. Smith, E., Weiler, A., Braschler, M.: Skill Extraction for Domain-Specific Text Retrieval in a Job-Matching Platform. In: Candan et al. [104], pp. 116–128

580. Sorg, P., Cimiano, P., Schultz, A., Sizov, S.: Overview of the Cross-lingual Expert Search (CriES) Pilot Challenge. In: Braschler et al. [96]
581. Spina, D., Amigó, E., Gonzalo, J.: Filter Keywords and Majority Class Strategies for Company Name Disambiguation in Twitter. In: Forner et al. [209], pp. 38–49
582. Stamatatos, E., Potthast, M., Rangel Pardo, F.M., Rosso, P., Stein, B.: Overview of the PAN/CLEF 2015 Evaluation Lab. In: Mothe et al. [420], pp. 518–538
583. Stamatatos, E., Rangel Pardo, F., Tschuggnall, M., Stein, B., Kestemont, M., Rosso, P., Potthast, M.: Overview of PAN 2018. In: Bellot et al. [64], pp. 267–285
584. Stathopoulos, S., Kalamboukis, T.: Medical Image Classification with Weighted Latent Semantic Tensors and Deep Convolutional Neural Networks. In: Bellot et al. [64], pp. 89–100
585. 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. [211], pp. 81–92
586. Stempfhuber, M., Baerisch, S.: The Domain-Specific Track at CLEF 2006: Overview of Approaches, Results and Assessment. In: Peters et al. [486], pp. 163–169
587. Suchomel, S., Brandejs, M.: Determining Window Size from Plagiarism Corpus for Stylographic Features. In: Mothe et al. [420], pp. 293–299
588. Suominen, H.: CLEFeHealth2012 - The CLEF 2012 Workshop on Cross-Language Evaluation of Methods, Applications, and Resources for eHealth Document Analysis. In: Forner et al. [210]
589. Suominen, H., Goeuriot, L., Kelly, L., Alonso Alemany, L., Bassani, E., Brew-Sam, N., Cotik, V., Filippo, D., González-Sáez, G., Luque, F., Mulhem, P., Pasi, G., Roller, R., Seneviratne, S., Upadhyay, R., Vivaldi, J., Viviani, M., Xu, C.: Overview of the CLEF eHealth Evaluation Lab 2021. In: Candan et al. [104], pp. 308–323
590. Suominen, H., Kelly, L., Goeuriot, L.: The Scholarly Impact and Strategic Intent of CLEF eHealth Labs from 2012–2017. In: Ferro and Peters [203]
591. Suominen, H., Kelly, L., Goeuriot, L., Névéol, 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. [64], pp. 286–301
592. 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. [211], pp. 212–231
593. Tannebaum, W., Mahdabi, P., Rauber, A.: Effect of Log-Based Query Term Expansion on Retrieval Effectiveness in Patent Searching. In: Mothe et al. [420], pp. 300–305
594. Tannebaum, W., Rauber, A.: Mining Query Logs of USPTO Patent Examiners. In: Forner et al. [211], pp. 136–142
595. Teixeira Lopes, C., Almeida Fernandes, T.: Health Suggestions: A Chrome Extension to Help Laypersons Search for Health Information. In: Fuhr et al. [219], pp. 241–246
596. Teixeira Lopes, C., Ribeiro, C.: Effects of Language and Terminology on the Usage of Health Query Suggestions. In: Fuhr et al. [219], pp. 83–95
597. Teixeira Lopes, C., Ribeiro, C.: Effects of Language and Terminology of Query Suggestions on the Precision of Health Searches. In: Bellot et al. [64], pp. 101–111
598. Thomee, B., Popescu, A.: Overview of the ImageCLEF 2012 Flickr Photo Annotation and Retrieval Task. In: Forner et al. [210]

599. 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)
600. Tian, L., Huang, N., Zhang, X.: Large Language Model Cascades and Persona-based In-context Learning for Multilingual Sexism Detection. In: Goeuriot et al. [240]
601. 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. [497], pp. 85–93
602. Trappett, M., Geva, S., Trotman, A., Scholer, F., Sanderson, M.: Overview of the INEX 2012 Snippet Retrieval Track. In: Forner et al. [210]
603. Trotzek, M., Koitka, S., Friedrich, C.M.: Early Detection of Depression Based on Linguistic Metadata Augmented Classifiers Revisited. In: Bellot et al. [64], pp. 191–202
604. Tsikrika, T., Garcia Seco de Herrera, A., Müller, H.: Assessing the Scholarly Impact of ImageCLEF. In: Forner et al. [209], pp. 95–106
605. Tsikrika, T., Kludas, J.: Overview of the WikipediaMM Task at ImageCLEF 2008. In: Peters et al. [488], pp. 539–550
606. Tsikrika, T., Kludas, J.: Overview of the WikipediaMM Task at ImageCLEF 2009. In: Peters et al. [497], pp. 60–71
607. Tsikrika, T., Larsen, B., Müller, H., Endrullis, S., Rahm, E.: The Scholarly Impact of CLEF (2000–2009). In: Forner et al. [211], pp. 1–12
608. Tsikrika, T., Popescu, A., Kludas, J.: Overview of the Wikipedia Image Retrieval Task at ImageCLEF 2011. In: Petras et al. [502]
609. Turchi, M., Steinberger, J., Alexandrov Kabadjov, M., Steinberger, R.: Using Parallel Corpora for Multilingual (Multi-document) Summarisation Evaluation. In: Agosti et al. [10], pp. 52–63
610. Turmo, J., Comas, P., Rosset, S., Galibert, O., Moreau, N., Mostefa, D., Rosso, P., Buscaldi, D.: Overview of QAST 2009. In: Peters et al. [489], pp. 197–211
611. Turmo, J., Comas, P., Rosset, S., Lamel, L., Moreau, N., Mostefa, D.: Overview of QAST 2008. In: Peters et al. [488], pp. 296–313
612. Turner, M., Ive, J., Velupillai, S.: Linguistic Uncertainty in Clinical NLP: A Taxonomy, Dataset and Approach. In: Candan et al. [104], pp. 129–141
613. 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. [494], pp. 307–331
614. 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. [211], pp. 41–42
615. 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. [420], pp. 444–461
616. 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. [219], pp. 267–285
617. Villegas, M., Paredes, R.: Overview of the ImageCLEF 2012 Scalable Web Image Annotation Task. In: Forner et al. [210]

618. Voorhees, E.M.: TREC: Continuing Information Retrieval's Tradition of Experimentation. *Communications of the ACM (CACM)* **50**(11), 51–54 (November 2007)
619. Voorhees, E.M.: The Evolution of Cranfield. In: Ferro and Peters [203], pp. 45–69
620. Wakeling, S., Clough, P.: Integrating Mixed-Methods for Evaluating Information Access Systems. In: Mothe et al. [420], pp. 306–311
621. Walker, A., Starkey, A., Pan, J.Z., Siddharthan, A.: Making Test Corpora for Question Answering More Representative. In: Kanoulas et al. [312], pp. 1–6
622. Wang, J., Gilbert, A., Thomee, B., Villegas, M.: Automatic Image Annotation at ImageCLEF. In: Ferro and Peters [203]
623. 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. [210]
624. Wang, X., Guo, Z., Zhang, Y., Li, J.: Medical Image Labeling and Semantic Understanding for Clinical Applications. In: Crestani et al. [138], pp. 260–270
625. Wang, X., X., W., Zhang, Q.: A Web-Based CLIR System with Cross-Lingual Topical Pseudo Relevance Feedback. In: Forner et al. [211], pp. 104–107
626. 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. [219], pp. 247–255
627. 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. [494], pp. 744–759
628. Wiegmann, M., Stein, B., Potthast, M.: De-Noising Document Classification Benchmarks via Prompt-based Rank Pruning: A Case Study. In: Goeuriot et al. [240]
629. Wilhelm-Stein, T., Eibl, M.: A Quantitative Look at the CLEF Working Notes. In: Forner et al. [211], pp. 13–16
630. Wilhelm-Stein, T., Eibl, M.: Teaching the IR Process Using Real Experiments Supported by Game Mechanics. In: Mothe et al. [420], pp. 312–317
631. Wilhelm-Stein, T., Herms, R., Ritter, M., Eibl, M.: Improving Transcript-Based Video Retrieval Using Unsupervised Language Model Adaptation. In: Kanoulas et al. [312], pp. 110–115
632. Wu, S.H., Huang, H.Y.: SimpleText Best of Labs in CLEF-2022: Simplify Text Generation with Prompt Engineering. In: Arampatzis et al. [36], pp. 198–208
633. Xu, K., Feng, Y., Huang, S., Zhao, D.: Question Answering via Phrasal Semantic Parsing. In: Mothe et al. [420], pp. 414–426
634. 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. [117], pp. 82–87
635. Yang, H., Gonçalves, T.: A Compound Model for Consumer Health Search. In: Bellot et al. [64], pp. 231–236
636. Yeshambel, T., Mothe, J., Assabie, Y.: 2AIRTC: The Amharic Adhoc Information Retrieval Test Collection. In: Arampatzis et al. [37], pp. 55–66
637. Yoon, W., Yoo, J., Seo, S., Sung, M., Jeong, M., Kim, G., Kang, J.: Data-Centric and Model-Centric Approaches for Biomedical Question Answering. In: Barrón-Cedeño et al. [55], pp. 204–216
638. 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. [312], pp. 128–140

639. Zanibbi, R., Oard, D.W., Agarwal, A., Mansouri, B.: Overview of ARQMath 2020: CLEF Lab on Answer Retrieval for Questions on Math. In: Arampatzis et al. [37], pp. 169–193
640. Zellhöfer, D.: Overview of the Personal Photo Retrieval Pilot Task at ImageCLEF 2012. In: Forner et al. [210]
641. Zhang, L., Rettinger, A., Färber, M., Tadic, M.: A Comparative Evaluation of Cross-Lingual Text Annotation Techniques. In: Forner et al. [211], pp. 124–135
642. Zhong, W., Xie, Y., Lin, J.: Answer Retrieval for Math Questions Using Structural and Dense Retrieval. In: Arampatzis et al. [36], pp. 209–223
643. Ziak, H., Kern: Evaluation of Pseudo Relevance Feedback Techniques for Cross Vertical Aggregated Search. In: Mothe et al. [420], pp. 91–102
644. Zingla, M.A., Latiri, C., Slimani, Y.: Tweet Contextualization Using Association Rules Mining and DBpedia. In: Mothe et al. [420], pp. 318–323
645. Zlabinger, M., Rekabsaz, N., Zlabinger, S., Hanbury, A.: Efficient Answer-Annotation for Frequent Questions. In: Crestani et al. [138], pp. 126–137
646. 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. [138], pp. 271–283