CLEF 15th Birthday: Past, Present, and Future

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

Abstract

2014 marks the 15th birthday for CLEF, an evaluation campaign activity which has applied the Cranfield evaluation paradigm to the testing of multilingual and multimodal information access systems in Europe. This paper provides a summary of the motivations which led to the establishment of CLEF, and a description of how it has evolved over the years, the major achievements, and what we see as the next challenges.

1 Introduction

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

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

The *Conference and Labs of the Evaluation Forum (CLEF)*[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], 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], 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 data and a common forum for comparing models and techniques applied to these languages.

[1] http://www.clef-initiative.eu/
[2] http://trec.nist.gov/
[3] http://research.nii.ac.jp/ntcir/
[4] http://www.isical.ac.in/~clia/

This year marks the 15th 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.

The paper is organized as follows: Section 2 describes the beginning and the first period of CLEF, the so-called "CLEF Classic" period; Section 3 introduces the second (and current) period of CLEF, known as the "CLEF Initiative" period; Sections 4 and 5 give an idea of the spread and extension of CLEF activities by providing a short account of the topics addressed in the conference, tracks and labs over the years together with pointers to papers providing more details; Section 6 attempts to provide an assessment of the status of CLEF in the IR community; finally, Section 7 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 [240], moving to an independent activity in 2000 [193].

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 [119]. 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, 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.

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)* and *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. 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

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

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

In the CLEF Initiative an increased focus is on the *multimodal* aspect, intended not only as the ability to deal with information coming in multiple media but also in different modalities, e.g. the Web, social media, news streams, specific domains and so on. These different modalities should, ideally, be addressed in an integrated way; rather than building vertical search systems for each domain/modality the interaction between the different modalities, languages, and user tasks needs to be exploited to provide comprehensive and aggregated search systems.

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

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

This is reflected in the new tasks offered by CLEF, as described in the next two sections.

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 Digital Libray Conference, but has become an independent event, held over 3.5-4 days and made up of two interrelated activities: the *Conference* and the *Labs*. 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 and fostering a deeper analysis and

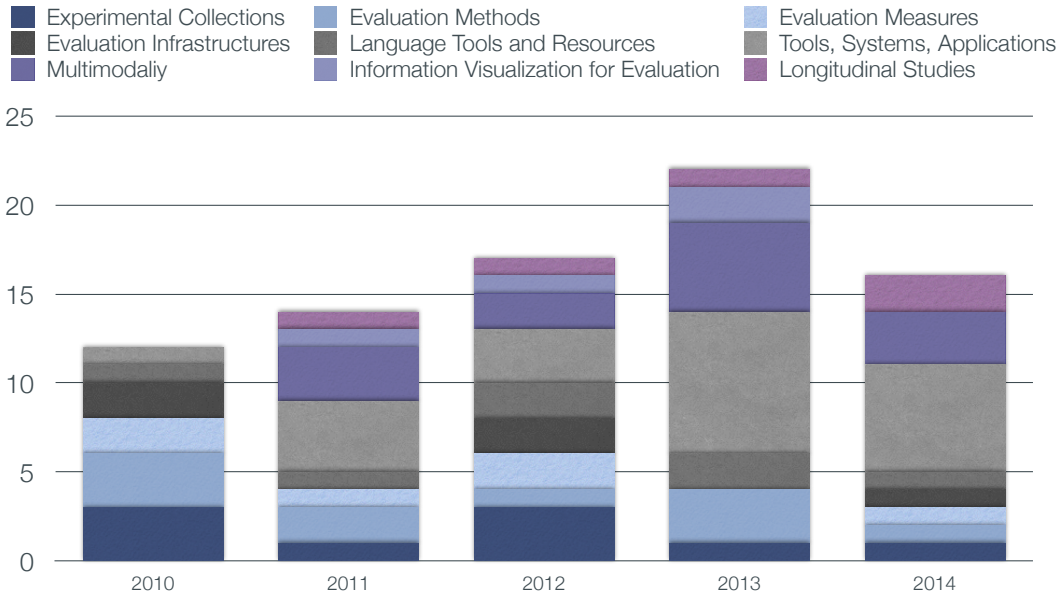

Figure 1: Topics addressed by the CLEF conference over the years and number of submissions for each topic.

understanding of experimental results. 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. The *Conference* and the *Labs* are expected to interact, bringing new interests and new expertise into CLEF.

In order to favour participation and the introduction of new perspectives, CLEF now has an open-bid process which allows research groups and institutions to bid to host the annual CLEF event and to propose themes. The bidding process follows a two-year cycle, i.e. in December 2014 bids to host CLEF 2017 will be sollicited.

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

4 The Conference

Figure 1 gives an overview of the topics addressed by the CLEF conference over the years, together with the number of submissions for each topics, as briefly summarized below with pointers to the main references:

Experimental Collections explored different issues concerning experimental collections such as: the creation of collections for Persian and Arabic languages; resource-effective creation of pseudo-test collections for specialised tasks; log-based experimental collections; collections for specific domains, e.g. question answering and plagiarism detection [27, 31, 32, 62, 90, 162, 216, 259, 266];

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 [25, 30, 64, 74, 120, 164, 165, 238, 239, 275];

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; and extension of measures to graded relevance [15, 17, 48, 78, 91, 153];

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 [4, 112, 113, 117, 150];

Language Tools 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; named entity extraction, linking and clustering in cross-lingual settings; exploitation of multiple translation resources; and language-independent generation of document snippets [24, 37, 52, 93, 139, 144, 151]

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; exploiting relevance feedback for building tag-clouds in image search; query expansion for image retrieval; and transcript-based video retrieval [34, 61, 63, 75, 92, 97, 106, 108, 111, 116, 132, 133, 138, 147, 148, 230, 233, 268, 271–273, 275];

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 and disambiguation; multimodal music tagging; multi-faceted IR in multimodal domains; ranking in faceted search [33, 56, 109, 110, 127, 152, 183, 184, 235, 241, 244, 245, 249];

Information Visualization for Evaluation opened up a brand new area concerned with exploiting information visualization and visual analytics techniques not only for presenting the

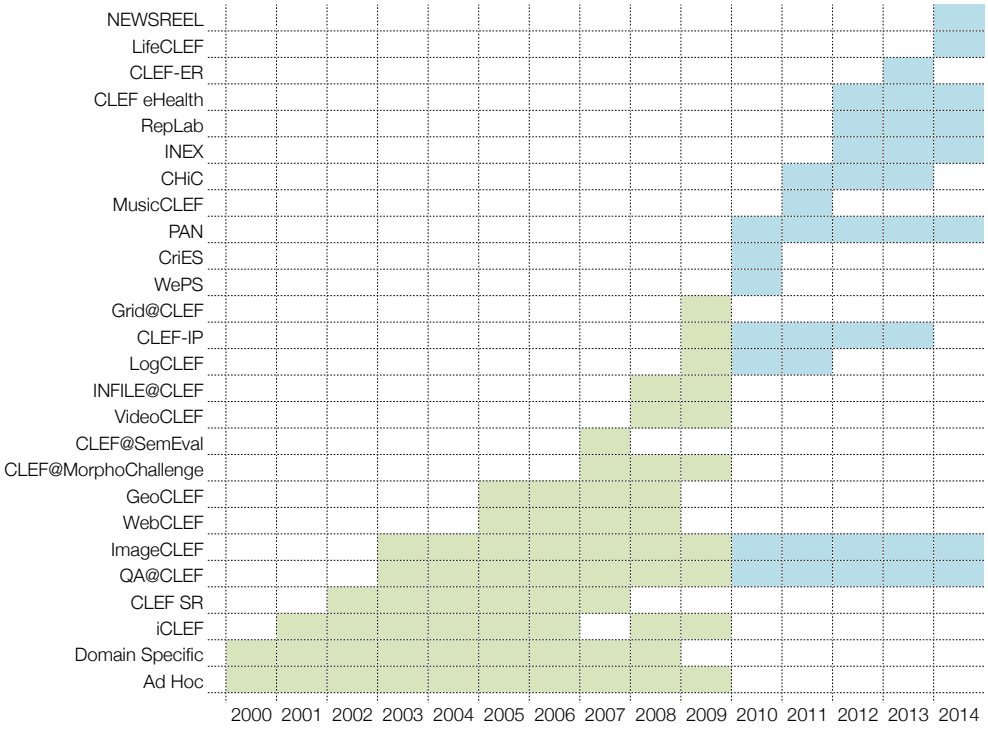

Figure 2: Labs offered by CLEF over the years (CLEF "Classic" period in green; the CLEF Initiative period in blue).

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 [19, 68, 143, 263];

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; and performance trends over the years for multilingual information access [82, 163, 176, 254, 257, 270].

5 Tracks and Labs

Figure 2 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.

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 [2, 3, 41–45, 69–71, 81];

Domain Specific (2000–2008) dealt with multilingual information retrieval on structured scientific data from the social sciences domain [42–44, 134–136, 210, 211, 246];

iCLEF (2001–2006; 2008–2009) explored different aspects of interactive information retrieval on multilingual and multimedia collections, also using gamification techniques [101–105, 130, 180, 181];

CLEF SR (2002–2007) investigated speech retrieval and spoken document retrieval in a monolingual and bilingual setting on automatic speech recognition transcripts [76, 77, 125, 182, 186, 269];

QA@CLEF (2003–2014) examined several aspects of question answering in a multilingual setting on document collections ranging from news, legal documents, medical documents, linked data [55, 88, 96, 154–156, 166, 187–192, 232, 237, 260–262];

ImageCLEF (2003–2014) studied the crosslanguage annotation and retrieval of images to support the advancement of the field of visual media analysis, indexing, classification, and retrieval [22, 50, 51, 57–60, 65–67, 98, 99, 107, 128, 149, 161, 168–173, 177–179, 220, 227, 228, 250, 252, 255, 256, 258, 264, 274];

WebCLEF (2005–2008) addressed multilingual Web search, exploring different faces of navigational queries and known-item search [26, 122, 123, 242];

GeoCLEF (2005–2008) evaluated cross-language geographic information retrieval (GIR) against search tasks involving both spatial and multilingual aspects [94, 95, 158, 160];

CLEF@SemEval (2007) explored the impact of *Word Sense Disambiguation (WSD)* on multilingual information retrieval [1]; 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 [140–142];

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 [145, 146]; it went on to become the MediaEval[5] successful evaluation series, dedicated to evaluating new algorithms for multimedia access and retrieval;

INFILE@CLEF (2008–2009) experimented with cross-language adaptive filtering systems on news corpora [35, 36];

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 [72, 157, 159];

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 [215, 217–219, 231];

[5] `http://www.multimediaeval.org/`

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 [80];

WePS (2010) focused on person name ambiguity and person attribute extraction on Web pages and on online reputation management for organizations [11, 23]; the activity continued in the RepLab lab;

CriES (2010) was run as a brainstorming workshop and addressed the problem of multi-lingual expert search in social media environments [243];

PAN (2010–2014) studied plagiarism, authorship attribution, and social software misuse [16, 20, 100, 121, 126, 221–226];

MusicCLEF (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 [185]; this activity has continued as part of MediaEval;

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 [89, 212, 213];

INEX (2012–2014) was a stand-alone initiative pioneering structured and XML retrieval from 2002[6]; it joined forces with CLEF in 2012 to further promote the evaluation of focused retrieval by providing large test collections of structured documents [28, 29, 54, 137, 236, 253, 267];

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 [12–14];

CLEF eHealth (2012–2014) 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 [131, 247, 248];

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 [229];

LifeCLEF (2014) 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 [124];

NEWSREEL (2014) focused on evaluation of news recommender systems in real-time by offering access to the APIs of a commercial system [118].

[6] `https://inex.mmci.uni-saarland.de/`

6 Trends

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

Figure 3 shows the participation in CLEF over the years. An almost constant growth trend is exhibited, a possible consequence of the capacity of CLEF to renew itself and to attract new communities and expertise in addition to core information retrieval activities.

The final year, 2014, shows a drop in participation which is probably due to both internal and external factors. First and foremost, 2014 represents the beginning of a new challenge for CLEF, as is also discussed in Section 7. For the first time since the beginnings of CLEF, the central organisation of CLEF was not supported by any European project in 2014 but was run by a 100% voluntary effort, striving to find a way to become self-sustainable. Encouragingly, we note that CLEF 2014 was able to attain levels of participation similar to CLEF 2010 and 2011, when CLEF started to benefit from the push of the PROMISE Network of Excellence. With respect to external factors, 2014 has represented a transition between the end of the seventh framework programme of the European Commission and the start of Horizon 2020; this may have caused a gap in the funding for research projects.

Figure 4 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.

Figure 5 shows the number of paper submitted and accepted in the CLEF Conference over the years. We see that the number of accepted papers has changed slightly over the years, almost stabilizing in the last two years, while the number of submitted papers has grown, allowing us to increase the selectivity and quality of the Conference.

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

Assessing the impact of an evaluation activity is a very demanding task. In 2010, TREC conducted a deep study on its economic impact [234]. 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 [251] and some steps in this direction have been performed for CLEF as well [18, 254, 257]. However, analysing the impact of evaluation activities on system performances over the years is still a research challenge, even if initial attempts have been made for both TREC [21] and CLEF [82].

Such rigorous studies are beyond the scope of the present report, here we concentrate on identifying rough indicators with respect to the maturity and liveliness of the scientific production originated by CLEF.

As far as maturity is concerned, an indicator might be found in publications critically analysing,

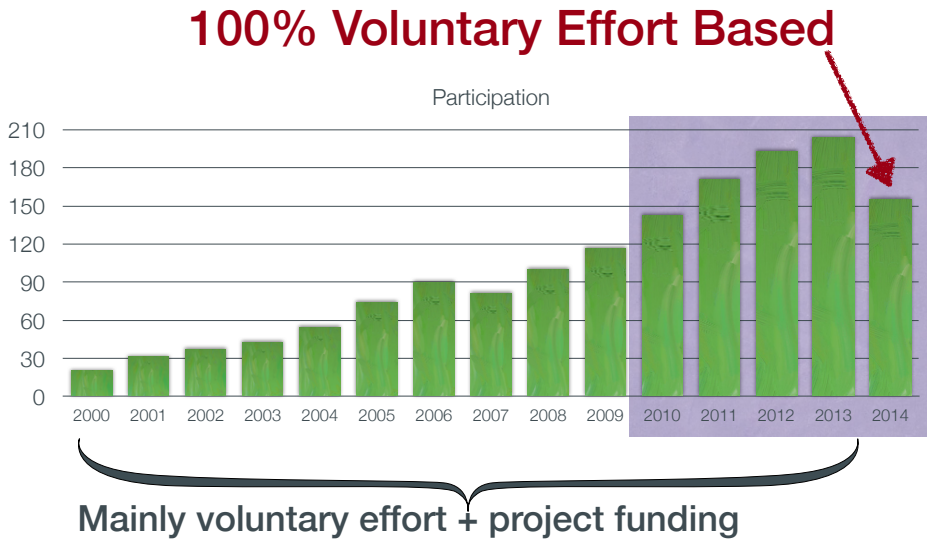

Figure 3: Participation in CLEF over the years (CLEF "Classic" period un-shaded; CLEF Initiative period shaded).

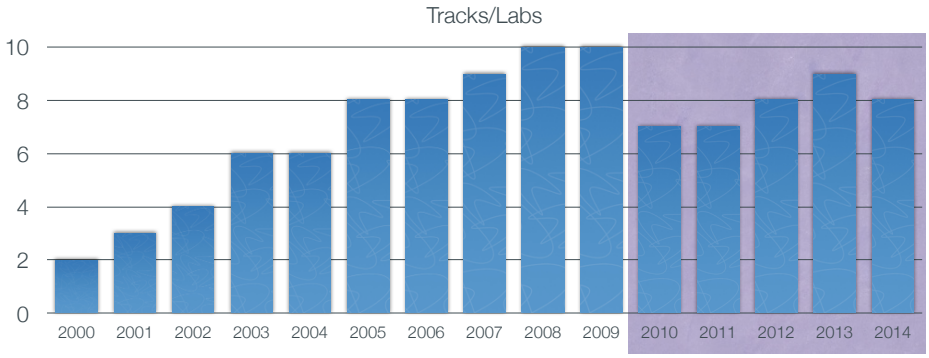

Figure 4: Number of labs offered by CLEF over the years (CLEF "Classic" period un-shaded; the CLEF Initiative period shaded).

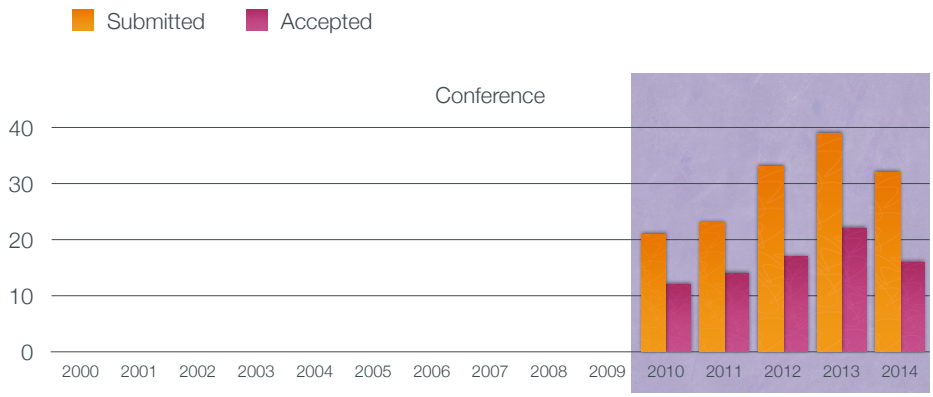

Figure 5: Number of papers submitted and accepted in the CLEF conference over the years (CLEF "Classic" period un-shaded; the CLEF Initiative period shaded).

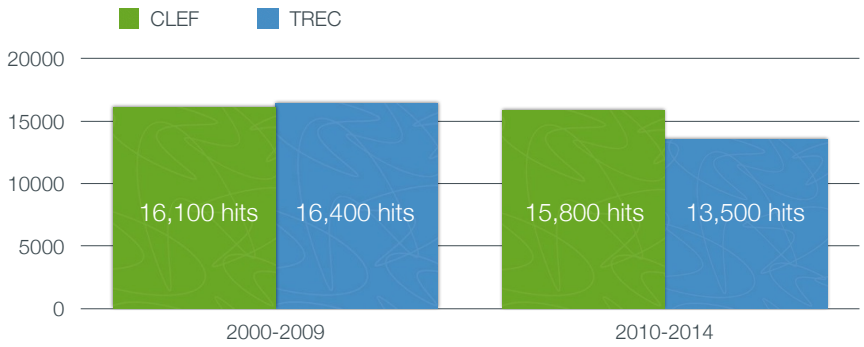

Figure 6: Hits in Google Scholar for the queries "CLEF evaluation" and "TREC evaluation" (CLEF "Classic" period on the left; the CLEF Initiative period on the right).

systematizing, and digesting the achievements, outcomes and experience; this has been done both for TREC [114, 115, 265] and CLEF [47, 167, 194].

When it comes to liveliness, a noisy indicator might be Google Schoolar. Figure 6 shows the number of hits for the two queries "CLEF evaluation" and "TREC evaluation", of course a very rough and coarse-grained estimate of the scientific production produced. The goal is not to compare the two initiatives but just to have an idea of whether CLEF presents trends comparable with a leading initiative in the field. It can be seen that both TREC and CLEF exhibit similar behaviour. However, the number of hits for TREC should be considered as slightly underestimated since TREC has two spin-off activities: TRECVID in 2003 and *Text Analysis Conference (TAC)* in 2008, which are not counted[7].

7 The CLEF Association

The CLEF Association[8] is an independent no-profit legal entity, established in October 2013 as a result of activity of the PROMISE[9] Network of Excellence which backed CLEF from 2010 to 2013.

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

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

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

[7] In particular, the query "TAC evaluation" is extremely noisy bringing in hundreds of thousands of results from the medical domain.

[8] `http://www.clef-initiative.eu/association`

[9] `http://www.promise-noe.eu/`

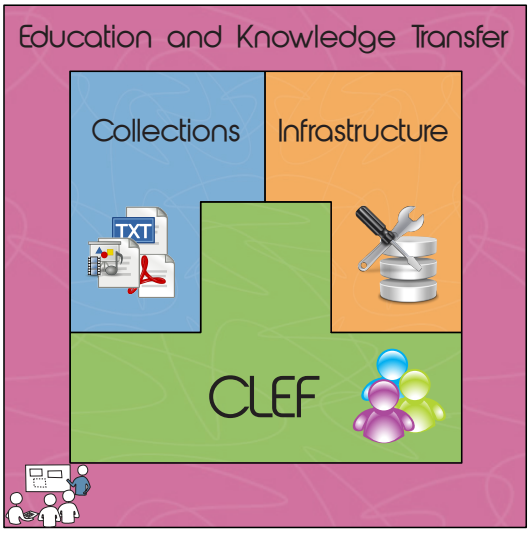

Figure 7: Pillar activities of the CLEF Association.

- *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 7 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 this initial phase, the CLEF Association is 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 [5, 6, 9, 18, 73, 83] 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 [7, 10, 79].

An additional example of the activities carried out by the CLEF Association during its first year to strengthen CLEF and extend its reach is the re-publishing of the entire CLEF Working Notes series [38–40, 46, 49, 85, 87, 174, 175, 202–205, 208, 214] under the CEUR Workshop Proceedings (CEUR-WS.org)[10], which provide permanent identifiers for each volume and better indexing by relevant services such as the DBLP[11] computer science bibliography and Google Scholar[12].

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.

We would like to sincerely and warmly thank Maristella Agosti, Donna Harman, and Carol Peters (Coordinator of CLEF 2000-2009) for their precious and continuous advice and suggestions during this journey into experimental evaluation.

[10] http://ceur-ws.org/

[11] http://www.informatik.uni-trier.de/~ley/db/

[12] http://scholar.google.com/

References

[1] E. Agirre, O. L. de Lacalle, B. Magnini, A. Otegi, G. Rigau, and P. Vossen. SemEval-2007 Task 01: Evaluating WSD on Cross-Language Information Retrieval. In Peters et al. [207], pages 908–917.

[2] E. Agirre, G. M. Di Nunzio, N. Ferro, T. Mandl, and C. Peters. CLEF 2008: Ad Hoc Track Overview. In Peters et al. [200], pages 15–37.

[3] E. Agirre, G. M. Di Nunzio, T. Mandl, and A. Otegi. CLEF 2009 Ad Hoc Track Overview: Robust-WSD Task. In Peters et al. [201], pages 36–49.

[4] M. Agosti, E. Di Buccio, N. Ferro, I. Masiero, S. Peruzzo, and G. Silvello. DIRECTions: Design and Specification of an IR Evaluation Infrastructure. In Catarci et al. [53], pages 88–99.

[5] M. Agosti, G. M. Di Nunzio, and N. Ferro. A Proposal to Extend and Enrich the Scientific Data Curation of Evaluation Campaigns. In T. Sakay, M. Sanderson, and D. K. Evans, editors, *Proc. 1st International Workshop on Evaluating Information Access (EVIA 2007)*, pages 62–73. National Institute of Informatics, Tokyo, Japan, 2007.

[6] M. Agosti, G. M. Di Nunzio, and N. Ferro. Scientific Data of an Evaluation Campaign: Do We Properly Deal With Them? In Peters et al. [198], pages 11–20.

[7] M. Agosti, G. M. Di Nunzio, N. Ferro, D. Harman, and C. Peters. The Future of Large-scale Evaluation Campaigns for Information Retrieval in Europe. In N. Fuhr, L. Kovács, and C. Meghini, editors, *Proc. 11th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2007)*, pages 509–512. Lecture Notes in Computer Science (LNCS) 4675, Springer, Heidelberg, Germany, 2007.

[8] M. Agosti, N. Ferro, C. Peters, M. de Rijke, and A. Smeaton, editors. *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.

[9] M. Agosti, N. Ferro, and C. Thanos. DESIRE 2011 Workshop on Data infrastructurEs for Supporting Information Retrieval Evaluation. *SIGIR Forum*, 46(1):51–55, June 2012.

[10] J. Allan, W. B. Croft, A. Moffat, and M. Sanderson. Frontiers, Challenges, and Opportunities for Information Retrieval – Report from SWIRL 2012, The Second Strategic Workshop on Information Retrieval in Lorne, February 2012. *SIGIR Forum*, 46(1):2–32, June 2012.

[11] E. Amigó, J. Artiles, J. Gonzalo, D. Spina, B. Liu, and A. Corujo. WePS3 Evaluation Campaign: Overview of the On-line Reputation Management Task. In Braschler et al. [46].

[12] E. Amigó, J. Carrillo de Albornoz, I. Chugur, A. Corujo, J. Gonzalo, T. Martín-Wanton, E. Meij, M. de Rijke, and D. Spina. Overview of RepLab 2013: Evaluating Online Reputation Monitoring Systems. In Forner et al. [86], pages 333–352.

[13] E. Amigó, J. Carrillo de Albornoz, I. Chugur, A. Corujo, J. Gonzalo, E. Meij, M. de Rijke, and D. Spina. Overview of RepLab 2014: Author Profiling and Reputation Dimensions for Online Reputation Management. In Kanoulas et al. [129], pages 307–322.

[14] E. Amigó, A. Corujo, J. Gonzalo, E. Meij, and M. de Rijke. Overview of RepLab 2012: Evaluating Online Reputation Management Systems. In Forner et al. [85].

[15] E. Amigó, J. Gonzalo, and M. F. Verdejo. A Comparison of Evaluation Metrics for Document Filtering. In Forner et al. [84], pages 38–49.

[16] M. Anderka and B. Stein. Overview of the 1th International Competition on Quality Flaw Prediction in Wikipedia. In Forner et al. [85].

[17] M. Angelini, N. Ferro, K. Järvelin, H. Keskustalo, A. Pirkola, G. Santucci, and G. Silvello. Cumulated Relative Position: A Metric for Ranking Evaluation. In Catarci et al. [53], pages 112–123.

[18] M. Angelini, N. Ferro, B. Larsen, H. Müller, G. Santucci, G. Silvello, and T. Tsikrika. Measuring and Analyzing the Scholarly Impact of Experimental Evaluation Initiatives. *Procedia Computer Science*, (in print).

[19] M. Angelini, N. Ferro, G. Santucci, and G. Silvello. Improving Ranking Evaluation Employing Visual Analytics. In Forner et al. [86], pages 29–40.

[20] S. Argamon and P. Juola. Overview of the International Authorship Identification Competition at PAN-2011.

In Petras et al. [214].

[21] T. G. Armstrong, A. Moffat, W. Webber, and J. Zobel. Improvements That Don't Add Up: Ad-Hoc Retrieval Results Since 1998. In D. W.-L. Cheung, I.-Y. Song, W. W. Chu, X. Hu, and J. J. Lin, editors, *Proc. 18th International Conference on Information and Knowledge Management (CIKM 2009)*, pages 601–610. ACM Press, New York, USA, 2009.

[22] T. Arni, P. Clough, M. Sanderson, and M. Grubinger. Overview of the ImageCLEFphoto 2008 Photographic Retrieval Task. In Peters et al. [200], pages 500–511.

[23] J. Artiles, A. Borthwick, J. Gonzalo, S. Sekine, and E. Amigó. WePS-3 Evaluation Campaign: Overview of the Web People Search Clustering and Attribute Extraction Tasks. In Braschler et al. [46].

[24] H. Azarbonyad, A. Shakery, and H. Faili. Exploiting Multiple Translation Resources for English-Persian Cross Language Information Retrieval. In Forner et al. [86], pages 93–99.

[25] L. Azzopardi and K. Balog. Towards a Living Lab for Information Retrieval Research and Development - A Proposal for a Living Lab for Product Search Tasks. In Forner et al. [84], pages 26–37.

[26] K. Balog, L. Azzopardi, J. Kamps, and M. de Rijke. Overview of WebCLEF 2006. In Peters et al. [198], pages 803–819.

[27] H. Baradaran Hashemi, A. Shakery, and H. Feili. Creating a Persian-English Comparable Corpus. In Agosti et al. [8], pages 27–39.

[28] P. Bellot, T. Bogers, S. Geva, M. A. Hall, H. C. Huurdeman, J. Kamps, G. Kazai, M. Koolen, V. Moriceau, J. Mothe, M. Preminger, E. SanJuan, R. Schenkel, M. Skov, X. Tannier, and D. Walsh. Overview of INEX 2014. In Kanoulas et al. [129], pages 212–228.

[29] P. Bellot, A. Doucet, S. Geva, S. Gurajada, J. Kamps, G. Kazai, M. Koolen, A. Mishra, V. Moriceau, J. Mothe, M. Preminger, E. SanJuan, R. Schenkel, X. Tannier, M. Theobald, M. Trappett, and Q. Wang. Overview of INEX 2013. In Forner et al. [86], pages 269–281.

[30] A. Beloborodov, P. Braslavski, and M. Driker. Towards Automatic Evaluation of Health-Related CQA Data. In Kanoulas et al. [129], pages 7–18.

[31] I. Bensalem, P. Rosso, and S. Chikhi. A New Corpus for the Evaluation of Arabic Intrinsic Plagiarism Detection. In Forner et al. [86], pages 53–58.

[32] R. Berendsen, M. Tsagkias, M. de Rijke, and E. Meij. Generating Pseudo Test Collections for Learning to Rank Scientific Articles. In Catarci et al. [53], pages 42–53.

[33] R. Berlanga Llavori, A. Jimeno-Yepes, M. Pérez Catalán, and D. Rebholz-Schuhmann. Context-Dependent Semantic Annotation in Cross-Lingual Biomedical Resources. In Forner et al. [86], pages 120–123.

[34] R. Berlanga Llavori, M. Pérez Catalán, L. Museros Cabedo, and R. Forcada. Semantic Discovery of Resources in Cloud-Based PACS/RIS Systems. In Forner et al. [86], pages 167–178.

[35] R. Besançon, S. Chaudiron, D. Mostefa, O. Hamon, Timimi. I., and K. Choukri. Overview of CLEF 2008 INFILE Pilot Track. In Peters et al. [200], pages 939–946.

[36] R. Besançon, S. Chaudiron, D. Mostefa, Timimi. I., K. Choukri, and M. Laïb. Information Filtering Evaluation: Overview of CLEF 2009 INFILE Track. In Peters et al. [201], pages 342–353.

[37] P. Bhaskar and S. Bandyopadhyay. Language Independent Query Focused Snippet Generation. In Catarci et al. [53], pages 138–140.

[38] F. Borri, A. Nardi, C. Peters, and N. Ferro, editors. *CLEF 2008 Working Notes*. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, `http://ceur-ws.org/Vol-1174/`, 2008.

[39] F. Borri, A. Nardi, C. Peters, and N. Ferro, editors. *CLEF 2009 Working Notes*. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, `http://ceur-ws.org/Vol-1175/`, 2009.

[40] F. Borri, C. Peters, and N. Ferro, editors. *CLEF 2004 Working Notes*. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, `http://ceur-ws.org/Vol-1170/`, 2004.

[41] M. Braschler. CLEF 2000 – Overview of Results. In Peters [193], pages 89–101.

[42] M. Braschler. CLEF 2001 – Overview of Results. In Peters et al. [195], pages 9–26.

[43] M. Braschler. CLEF 2002 – Overview of Results. In Peters et al. [196], pages 9–27.

[44] M. Braschler. CLEF 2003 – Overview of Results. In Peters et al. [197], pages 44–63.

[45] M. Braschler, G. M. Di Nunzio, N. Ferro, and C. Peters. CLEF 2004: Ad Hoc Track Overview and Results

Analysis. In Peters et al. [199], pages 10–26.

[46] M. Braschler, D. K. Harman, E. Pianta, and N. Ferro, editors. *CLEF 2010 Working Notes*. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, `http://ceur-ws.org/Vol-1176/`, 2010.

[47] M. Braschler and C. Peters. Cross-Language Evaluation Forum: Objectives, Results, Achievements. *Information Retrieval*, 7(1–2):7–31, 2004.

[48] G. Cabanac, G. Hubert, M. Boughanem, and C. Chrisment. Tie-Breaking Bias: Effect of an Uncontrolled Parameter on Information Retrieval Evaluation. In Agosti et al. [8], pages 112–123.

[49] L. Cappellato, N. Ferro, M. Halvey, and W. Kraaij, editors. *CLEF 2014 Working Notes*. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, `http://ceur-ws.org/Vol-1180/`, 2014.

[50] B. Caputo, H. Müller, J. Martínez-Gómez, M. Villegas, B. Acar, N. Patricia, N. Barzegar Marvasti, S. Üsküdarli, R. Paredes, M. Cazorla, I. García-Varea, and V. Morell. ImageCLEF 2014: Overview and Analysis of the Results. In Kanoulas et al. [129], pages 192–211.

[51] B. Caputo, H. Müller, B. Thomee, M. Villegas, R. Paredes, D. Zellhöfer, H. Goëau, A. Joly, P. Bonnet, J. Martínez-Gómez, I. García-Varea, and M. Cazorla. ImageCLEF 2013: The Vision, the Data and the Open Challenges. In Forner et al. [86], pages 250–268.

[52] T. Cassidy, H. Ji, H. Deng, J. Zheng, and J. Han. Analysis and Refinement of Cross-Lingual Entity Linking. In Catarci et al. [53], pages 1–12.

[53] T. Catarci, P. Forner, D. Hiemstra, A. Peñas, and G. Santucci, editors. *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.

[54] T. Chappell and S. Geva. Overview of the INEX 2012 Relevance Feedback Track. In Forner et al. [85].

[55] P. Cimiano, V. Lopez, C. Unger, E. Cabrio, A.-C. Ngonga Ngomo, and S. Walter. Multilingual Question Answering over Linked Data (QALD-3): Lab Overview. In Forner et al. [86], pages 321–332.

[56] P. Clough and P. Goodale. Selecting Success Criteria: Experiences with an Academic Library Catalogue. In Forner et al. [86], pages 59–70.

[57] P. Clough, M. Grubinger, T. Deselaers, A. Hanbury, and H. Müller. Overview of the ImageCLEF 2006 Photographic Retrieval and Object Annotation Tasks. In Peters et al. [198], pages 223–256.

[58] P. Clough, H. Müller, T. Deselaers, M. Grubinger, T. M. Lehmann, J. R. Jensen, and W. R. Hersh. The CLEF 2005 Cross-Language Image Retrieval Track. In Peters et al. [206], pages 535–557.

[59] P. Clough, H. Müller, and M. Sanderson. The CLEF 2004 Cross-Language Image Retrieval Track. In Peters et al. [199], pages 597–613.

[60] P. Clough and M. Sanderson. The CLEF 2003 Cross Language Image Retrieval Track. In Peters et al. [197], pages 581–593.

[61] R. Corezola Pereira, V. Pereira Moreira, and R. Galante. A New Approach for Cross-Language Plagiarism Analysis. In Agosti et al. [8], pages 15–26.

[62] S. de L. Pertile and V. Pereira Moreira. A Test Collection to Evaluate Plagiarism by Missing or Incorrect References. In Catarci et al. [53], pages 141–143.

[63] S. de L. Pertile, P. Rosso, and V. Pereira Moreira. Counting Co-occurrences in Citations to Identify Plagiarised Text Fragments. In Forner et al. [86], pages 150–154.

[64] M. de Rijke, K. Balog, T. Bogers, and A. van den Bosch. On the Evaluation of Entity Profiles. In Agosti et al. [8], pages 94–99.

[65] T. Deselaers and T. M. Deserno. Medical Image Annotation in ImageCLEF 2008. In Peters et al. [200], pages 523–530.

[66] T. Deselaers and A. Hanbury. The Visual Concept Detection Task in ImageCLEF 2008. In Peters et al. [200], pages 531–538.

[67] T. Deselaers, A. Hanbury, V. Viitaniemi, A. A. Benczúr, M. Brendel, B. Daróczy, H. J. Escalante Balderas, T. Gevers, C. A. Hernández-Gracidas, S. C. H. Hoi, J. Laaksonen, M. Li, H. M. Marín Castro, H. Ney, X. Rui, N. Sebe, J. Stöttinger, and L. Wu. Overview of the ImageCLEF 2007 Object Retrieval Task. In Peters et al. [207], pages 445–471.

[68] E. Di Buccio, M. Dussin, N. Ferro, I. Masiero, G. Santucci, and G. Tino. To Re-rank or to Re-query: Can

Visual Analytics Solve This Dilemma? In Forner et al. [84], pages 119–130.

[69] G. M. Di Nunzio, N. Ferro, G. J. F. Jones, and C. Peters. CLEF 2005: Ad Hoc Track Overview. In Peters et al. [206], pages 11–36.

[70] G. M. Di Nunzio, N. Ferro, T. Mandl, and C. Peters. CLEF 2006: Ad Hoc Track Overview. In Peters et al. [198], pages 21–34.

[71] G. M. Di Nunzio, N. Ferro, T. Mandl, and C. Peters. CLEF 2007: Ad Hoc Track Overview. In Peters et al. [207], pages 13–32.

[72] G. M. Di Nunzio, J. Leveling, and T. Mandl. LogCLEF 2011 Multilingual Log File Analysis: Language Identification, Query Classification, and Success of a Query. In Petras et al. [214].

[73] M. Dussin and N. Ferro. Managing the Knowledge Creation Process of Large-Scale Evaluation Campaigns. In M. Agosti, J. Borbinha, S. Kapidakis, C. Papatheodorou, and G. Tsakonas, editors, *Proc. 13th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2009)*, pages 63–74. Lecture Notes in Computer Science (LNCS) 5714, Springer, Heidelberg, Germany, 2009.

[74] F. Esuli and F. Sebastiani. Evaluating Information Extraction. In Agosti et al. [8], pages 100–111.

[75] A. M. Ezzeldin, M. H. Kholief, and Y. El-Sonbaty. ALQASIM: Arabic Language Question Answer Selection in Machines. In Forner et al. [86], pages 100–103.

[76] M. Federico, N. Bertoldi, G.-A. Levow, and G. J. F. Jones. CLEF 2004 Cross-Language Spoken Document Retrieval Track. In Peters et al. [199], pages 816–820.

[77] M. Federico and G. J. F. Jones. The CLEF 2003 Cross-Language Spoken Document Retrieval Track. In Peters et al. [197], page 646.

[78] M. Ferrante, N. Ferro, and M. Maistro. Rethinking How to Extend Average Precision to Graded Relevance. In Kanoulas et al. [129], pages 19–30.

[79] N. Ferro, R. Berendsen, A. Hanbury, M. Lupu, V. Petras, M. de Rijke, and G. Silvello. PROMISE Retreat Report – Prospects and Opportunities for Information Access Evaluation. *SIGIR Forum*, 46(2):60–84, December 2012.

[80] N. Ferro and D. Harman. CLEF 2009: Grid@CLEF Pilot Track Overview. In Peters et al. [201], pages 552–565.

[81] N. Ferro and C. Peters. CLEF 2009 Ad Hoc Track Overview: TEL & Persian Tasks. In Peters et al. [201], pages 13–35.

[82] N. Ferro and G. Silvello. CLEF 15th Birthday: What Can We Learn From Ad Hoc Retrieval? In Kanoulas et al. [129], pages 31–43.

[83] N. Ferro and G. Silvello. Making it Easier to Discover, Re-Use and Understand Search Engine Experimental Evaluation Data. *ERCIM News*, 96:26–27, January 2014.

[84] P. Forner, J. Gonzalo, J. Kekäläinen, M. Lalmas, and M. de Rijke, editors. *Multilingual and Multimodal Information Access Evaluation. Proceedings of the Second International Conference of the Cross-Language Evaluation Forum (CLEF 2011)*. Lecture Notes in Computer Science (LNCS) 6941, Springer, Heidelberg, Germany, 2011.

[85] P. Forner, J. Karlgren, C. Womser-Hacker, and N. Ferro, editors. *CLEF 2012 Working Notes*. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, `http://ceur-ws.org/Vol-1178/`, 2012.

[86] P. Forner, H. Müller, R. Paredes, P. Rosso, and B. Stein, editors. *Information Access Evaluation meets Multilinguality, Multimodality, and Visualization. Proceedings of the Fourth International Conference of the CLEF Initiative (CLEF 2013)*. Lecture Notes in Computer Science (LNCS) 8138, Springer, Heidelberg, Germany, 2013.

[87] P. Forner, R. Navigli, D. Tufis, and N. Ferro, editors. *CLEF 2013 Working Notes*. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, `http://ceur-ws.org/Vol-1179/`, 2013.

[88] P. Forner, A. Peñas, E. Agirre, I. Alegria, C. Forascu, N. Moreau, P. Osenova, P. Prokopidis, P. Rocha, B. Sacaleanu, R. F. E. Sutcliffe, and E. F. T. K. Sang. Overview of the Clef 2008 Multilingual Question Answering Track. In Peters et al. [200], pages 262–295.

[89] M. Gäde, N. Ferro, and M. Lestari Paramita. CHiC 2011 – Cultural Heritage in CLEF: From Use Cases to Evaluation in Practice for Multilingual Information Access to Cultural Heritage. In Petras et al. [214].

[90] M. Gäde, J. Stiller, and V. Petras. Which Log for Which Information? Gathering Multilingual Data from

Different Log File Types. In Agosti et al. [8], pages 70–81.

[91] P. Galuscáková, P. Pecina, and J. Hajic. Penalty Functions for Evaluation Measures of Unsegmented Speech Retrieval. In Catarci et al. [53], pages 100–111.

[92] D. Ganguly, J. Leveling, and G. J. F. Jones. Simulation of Within-Session Query Variations Using a Text Segmentation Approach. In Forner et al. [84], pages 89–94.

[93] D. Ganguly, J. Leveling, and G. J. F. Jones. A Case Study in Decompounding for Bengali Information Retrieval. In Forner et al. [86], pages 108–119.

[94] F. Gey, R. Larson, M. Sanderson, K. Bischoff, T. Mandl, K. Womser-Hacker, D. Santos, P. Rocha, G. M. Di Nunzio, and N. Ferro. GeoCLEF 2006: the CLEF 2006 Cross-Language Geographic Information Retrieval Track Overview. In Peters et al. [198], pages 852–876.

[95] F. C. Gey, R. R. Larson, M. Sanderson, H. Joho, P. Clough, and V. Petras. GeoCLEF: The CLEF 2005 Cross-Language Geographic Information Retrieval Track Overview. In Peters et al. [206], pages 908–919.

[96] D. Giampiccolo, P. Forner, J. Herrera, A. Peñas, C. Ayache, C. Forascu, V. Jijkoun, P. Osenova, P. Rocha, B. Sacaleanu, and R. F. E. Sutcliffe. Overview of the CLEF 2007 Multilingual Question Answering Track. In Peters et al. [207], pages 200–236.

[97] D. G. Glinos. Discovering Similar Passages within Large Text Documents. In Kanoulas et al. [129], pages 98–109.

[98] H. Goëau, P. Bonnet, A. Joly, N. Boujemaa, D. Barthelemy, J.-F. Molino, P. Birnbaum, E. Mouysset, and M. Picard. The CLEF 2011 Plant Images Classification Task. In Petras et al. [214].

[99] H. Goëau, P. Bonnet, A. Joly, I. Yahiaoui, D. Barthelemy, N. Boujemaa, and J.-F. Molino. The ImageCLEF 2012 Plant Identification Task. In Forner et al. [85].

[100] T. Gollub, M. Potthast, A. Beyer, M. Busse, F. Rangel Pardo, P. Rosso, E. Stamatatos, and B. Stenno. Recent Trends in Digital Text Forensics and Its Evaluation - Plagiarism Detection, Author Identification, and Author Profiling. In Forner et al. [86], pages 282–302.

[101] J. Gonzalo, P. Clough, and J. Karlgren. Overview of iCLEF 2008: Search Log Analysis for Multilingual Image Retrieval. In Peters et al. [200], pages 227–235.

[102] J. Gonzalo, P. Clough, and A. Vallin. Overview of the CLEF 2005 Interactive Track. In Peters et al. [206], pages 251–262.

[103] J. Gonzalo and D. W. Oard. The CLEF 2002 Interactive Track. In Peters et al. [196], pages 372–382.

[104] J. Gonzalo and D. W. Oard. iCLEF 2004 Track Overview: Pilot Experiments in Interactive Cross-Language Question Answering. In Peters et al. [199], pages 310–322.

[105] J. Gonzalo, V. Peinado, P. Clough, and J. Karlgren. Overview of iCLEF 2009: Exploring Search Behaviour in a Multilingual Folksonomy Environment. In Peters et al. [209], pages 13–20.

[106] T. Goodwin and S. M. Harabagiu. The Impact of Belief Values on the Identification of Patient Cohorts. In Forner et al. [86], pages 155–166.

[107] M. Grubinger, P. Clough, A. Hanbury, and H. Müller. Overview of the ImageCLEFphoto 2007 Photographic Retrieval Task. In Peters et al. [207], pages 433–444.

[108] P. Gupta, A. Barrón-Cedeño, and P. Rosso. Cross-Language High Similarity Search Using a Conceptual Thesaurus. In Catarci et al. [53], pages 67–75.

[109] M. Hagen and C. Glimm. Supporting More-Like-This Information Needs: Finding Similar Web Content in Different Scenarios. In Kanoulas et al. [129], pages 50–61.

[110] M. Hall and E. Toms. Building a Common Framework for IIR Evaluation. In Forner et al. [86], pages 17–28.

[111] H. Hammarström. Automatic Annotation of Bibliographical References for Descriptive Language Materials. In Forner et al. [84], pages 62–73.

[112] A. Hanbury and H. Müller. Automated Component-Level Evaluation: Present and Future. In Agosti et al. [8], pages 124–135.

[113] A. Hanbury, H. Müller, G. Langs, M.-A. Weber, B. H. Menze, and T. Salas Fernandez. Bringing the Algorithms to the Data: Cloud-Based Benchmarking for Medical Image Analysis. In Catarci et al. [53], pages 24–29.

[114] D. K. Harman. *Information Retrieval Evaluation*. Morgan & Claypool Publishers, USA, 2011.

[115] D. K. Harman and E. M. Voorhees, editors. *TREC. Experiment and Evaluation in Information Retrieval*. MIT Press, Cambridge (MA), USA, 2005.

[116] C. G. Harris and T. Xu. The Importance of Visual Context Clues in Multimedia Translation. In Forner et al. [84], pages 107–118.

[117] D. Hiemstra and C. Hauff. MapReduce for Information Retrieval Evaluation: "Let's Quickly Test This on 12 TB of Data". In Agosti et al. [8], pages 64–69.

[118] F. Hopfgartner, B. Kille, A. Lommatzsch, T. Plumbaum, T. Brodt, and T. Heintz. Benchmarking News Recommendations in a Living Lab. In Kanoulas et al. [129], pages 250–267.

[119] D. A. Hull and D. W. Oard. 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.

[120] B. Huurnink, K. Hofmann, M. de Rijke, and M. Bron. Validating Query Simulators: An Experiment Using Commercial Searches and Purchases. In Agosti et al. [8], pages 40–51.

[121] G. Inches and F. Crestani. Overview of the International Sexual Predator Identification Competition at PAN-2012. In Forner et al. [85].

[122] V. Jijkoun and M. de Rijke. Overview of WebCLEF 2007. In Peters et al. [207], pages 725–731.

[123] V. Jijkoun and M. de Rijke. Overview of WebCLEF 2008. In Peters et al. [200], pages 787–793.

[124] A. Joly, H. Goëau, H. Glotin, C. Spampinato, P. Bonnet, W.-P. Vellinga, R. Planquè, A. Rauber, R. B. Fisher, and H. Müller. LifeCLEF 2014: Multimedia Life Species Identification Challenges. In Kanoulas et al. [129], pages 229–249.

[125] G. J. F. Jones and M. Federico. CLEF 2002 Cross-Language Spoken Document Retrieval Pilot Track Report. In Peters et al. [196], pages 446–457.

[126] P. Juola. An Overview of the Traditional Authorship Attribution Subtask. In Forner et al. [85].

[127] J. Jürgens, P. Hansen, and C. Womser-Hacker. Going beyond CLEF-IP: The 'Reality' for Patent Searchers. In Catarci et al. [53], pages 30–35.

[128] J. Kalpathy-Cramer, H. Müller, S. Bedrick, I. Eggel, A. Garcia Seco de Herrera, and T. Tsikrika. Overview of the CLEF 2011 Medical Image Classification and Retrieval Tasks. In Petras et al. [214].

[129] E. Kanoulas, M. Lupu, P. Clough, M. Sanderson, M. Hall, A. Hanbury, and E. Toms, editors. *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.

[130] J. Karlgren, J. Gonzalo, and P. Clough. iCLEF 2006 Overview: Searching the Flickr WWW Photo-Sharing Repository. In Peters et al. [198], pages 186–194.

[131] L. Kelly, L. Goeuriot, H. Suominen, T. Schreck, G. Leroy, D. L. Mowery, S. Velupillai, W. Webber Chapman, D. Martínez, G. Zuccon, and J. R. M. Palotti. Overview of the ShARe/CLEF eHealth Evaluation Lab 2014. In Kanoulas et al. [129], pages 172–191.

[132] A. Keszler, L. Kovács, and T. Szirányi. The Appearance of the Giant Component in Descriptor Graphs and Its Application for Descriptor Selection. In Catarci et al. [53], pages 76–81.

[133] S.-J. Kim and J.-H. Lee. Subtopic Mining Based on Head-Modifier Relation and Co-occurrence of Intents Using Web Documents. In Forner et al. [86], pages 179–191.

[134] M. Kluck. The Domain-Specific Track in CLEF 2004: Overview of the Results and Remarks on the Assessment Process. In Peters et al. [199], pages 260–270.

[135] M. Kluck and F. C. Gey. The Domain-Specific Task of CLEF – Specific Evaluation Strategies in Cross-Language Information Retrieval. In Peters [193], pages 48–56.

[136] M. Kluck and M. Stempfhuber. Domain-Specific Track CLEF 2005: Overview of Results and Approaches, Remarks on the Assessment Analysis. In Peters et al. [206], pages 212–221.

[137] M. Koolen, G. Kazai, J. Kamps, M. Preminger, A. Doucet, and M. Landoni. Overview of the INEX 2012 Social Book Search Track. In Forner et al. [85].

[138] A. Kosmopoulos, G. Paliouras, and I. Androutsopoulos. The Effect of Dimensionality Reduction on Large Scale Hierarchical Classification. In Kanoulas et al. [129], pages 160–171.

[139] N. K. Kumar, G. S. K. Santosh, and V. Varma. A Language-Independent Approach to Identify the Named Entities in Under-Resourced Languages and Clustering Multilingual Documents. In Forner et al. [84], pages 74–82.

[140] M. Kurimo, M. Creutz, and M. Varjokallio. Morpho Challenge Evaluation Using a Linguistic Gold Standard. In Peters et al. [207], pages 864–872.

[141] M. Kurimo, V. T. Turunen, and M. Varjokallio. Overview of Morpho Challenge 2008. In Peters et al. [200], pages 951–966.

[142] M. Kurimo, S. Virpioja, V. T. Turunen, G. W. Blackwood, and W. Byrne. Overview and Results of Morpho Challenge 2009. In Peters et al. [201], pages 587–597.

[143] J. Kürsten and M. Eibl. Comparing IR System Components Using Beanplots. In Catarci et al. [53], pages 136–137.

[144] M. Kvist and S. Velupillai. SCAN: A Swedish Clinical Abbreviation Normalizer - Further Development and Adaptation to Radiology. In Kanoulas et al. [129], pages 62–73.

[145] M. Larson, E. Newman, and G. J. F. Jones. Overview of VideoCLEF 2008: Automatic Generation of Topic-Based Feeds for Dual Language Audio-Visual Content. In Peters et al. [200], pages 906–917.

[146] M. Larson, E. Newman, and G. J. F. Jones. Overview of VideoCLEF 2009: New Perspectives on Speech-Based Multimedia Content Enrichment. In Peters et al. [209], pages 354–368.

[147] L. A. Leiva, M. Villegas, and R. Paredes. Relevant Clouds: Leveraging Relevance Feedback to Build Tag Clouds for Image Search. In Forner et al. [86], pages 143–149.

[148] C. W. Leong, S. Hassan, M. E. Ruiz, and M. Rada. Improving Query Expansion for Image Retrieval via Saliency and Picturability. In Forner et al. [84], pages 137–142.

[149] M. Lestari Paramita, M. Sanderson, and P. Clough. Diversity in Photo Retrieval: Overview of the ImageCLEFPhoto Task 2009. In Peters et al. [209], pages 45–59.

[150] A. Lipani, F. Piroi, L. Andersson, and A. Hanbury. An Information Retrieval Ontology for Information Retrieval Nanopublications. In Kanoulas et al. [129], pages 44–49.

[151] A. Loponen and K. Järvelin. A Dictionary- and Corpus-Independent Statistical Lemmatizer for Information Retrieval in Low Resource Languages. In Agosti et al. [8], pages 3–14.

[152] S. Mackie, R. McCreadie, C. Macdonald, and I. Ounis. Comparing Algorithms for Microblog Summarisation. In Kanoulas et al. [129], pages 153–159.

[153] W. Magdy and G. J. F. Jones. Examining the Robustness of Evaluation Metrics for Patent Retrieval with Incomplete Relevance Judgements. In Agosti et al. [8], pages 82–93.

[154] B. Magnini, D. Giampiccolo, P. Forner, C. Ayache, V. Jijkoun, P. Osenova, A. Peñas, P. Rocha, B. Sacaleanu, and R. F. E. Sutcliffe. Overview of the CLEF 2006 Multilingual Question Answering Track. In Peters et al. [198], pages 223–256.

[155] B. Magnini, S. Romagnoli, A. Vallin, J. Herrera, A. Peñas, V. Peinado, M. F. Verdejo, and M. de Rijke. The Multiple Language Question Answering Track at CLEF 2003. In Peters et al. [197], pages 471–486.

[156] B. Magnini, A. Vallin, C. Ayache, G. Erbach, A. Peñas, M. de Rijke, P. Rocha, K. I. Simov, and R. F. E. Sutcliffe. Overview of the CLEF 2004 Multilingual Question Answering Track. In Peters et al. [199], pages 371–391.

[157] T. Mandl, M. Agosti, G. M. Di Nunzio, A. S. Yeh, I. Mani, C. Doran, and J. M. Schulz. LogCLEF 2009: The CLEF 2009 Multilingual Logfile Analysis Track Overview. In Peters et al. [201], pages 508–517.

[158] T. Mandl, P. Carvalho, G. M. Di Nunzio, F. Gey, R. Larson, D. Santos, and C. Womser-Hacker. GeoCLEF 2008: The CLEF 2008 Cross-Language Geographic Information Retrieval Track Overview. In Peters et al. [200], pages 808–821.

[159] T. Mandl, G. M. Di Nunzio, and J. M. Schulz. LogCLEF 2010: the CLEF 2010 Multilingual Logfile Analysis Track Overview. In Braschler et al. [46].

[160] T. Mandl, F. Gey, G. M. Di Nunzio, N. Ferro, R. Larson, M. Sanderson, D. Santos, C. Womser-Hacker, and X. Xie. GeoCLEF 2007: the CLEF 2007 Cross-Language Geographic Information Retrieval Track Overview. In Peters et al. [207], pages 745–772.

[161] J. Martínez-Gómez, I. García-Varea, and B. Caputo. Overview of the ImageCLEF 2012 Robot Vision Task. In Forner et al. [85].

[162] J. Mayfield, D. Lawrie, P. McNamee, and D. W. Oard. Building a Cross-Language Entity Linking Collection in Twenty-One Languages. In Forner et al. [84], pages 3–13.

[163] R. McCreadie, C. Macdonald, I. Ounis, and J. Brassey. A Study of Personalised Medical Literature Search. In Kanoulas et al. [129], pages 74–85.

[164] M. R. Mirsarraf and N. Dehghani. A Dependency-Inspired Semantic Evaluation of Machine Translation Systems. In Forner et al. [86], pages 71–74.

[165] A. Molina, E. SanJuan, and J.-M. Torres-Moreno. A Turing Test to Evaluate a Complex Summarization Task. In Forner et al. [86], pages 75–80.

[166] R. Morante and W. Daelemans. Overview of the QA4MRE Pilot Task: Annotating Modality and Negation for a Machine Reading Evaluation. In Petras et al. [214].

[167] H. Müller, P. Clough, T. Deselaers, and B. Caputo, editors. *ImageCLEF – Experimental Evaluation in Visual Information Retrieval*. Springer-Verlag, Heidelberg, Germany, 2010.

[168] H. Müller, T. Deselaers, T. M. Deserno, P. Clough, E. Kim, and W. R. Hersh. Overview of the ImageCLEFmed 2006 Medical Retrieval and Medical Annotation Tasks. In Peters et al. [198], pages 595–608.

[169] H. Müller, T. Deselaers, T. M. Deserno, J. Kalpathy-Cramer, E. Kim, and W. R. Hersh. Overview of the ImageCLEFmed 2007 Medical Retrieval and Medical Annotation Tasks. In Peters et al. [207], pages 472–491.

[170] H. Müller, A. Garcia Seco de Herrera, J. Kalpathy-Cramer, D. Demner-Fushman, S. Antani, and I. Eggel. Overview of the ImageCLEF 2012 Medical Image Retrieval and Classification Tasks. In Forner et al. [85].

[171] H. Müller, J. Kalpathy-Cramer, I. Eggel, S. Bedrick, S. Radhouani, B. Bakke, C. E. Khan Jr., and W. R. Hersh. Overview of the CLEF 2009 Medical Image Retrieval Track. In Peters et al. [209], pages 72–84.

[172] H. Müller, J. Kalpathy-Cramer, I. Eggel, S. Bedrick, J. Reisetter, C. E. Khan Jr., and W. R. Hersh. Overview of the CLEF 2010 Medical Image Retrieval Track. In Braschler et al. [46].

[173] H. Müller, J. Kalpathy-Cramer, C. E. Kahn, W. Hatt, S. Bedrick, and W. Hersh. Overview of the ImageCLEFmed 2008 Medical image Retrieval Task. In Peters et al. [200], pages 512–522.

[174] A. Nardi, C. Peters, and N. Ferro, editors. *CLEF 2007 Working Notes*. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, `http://ceur-ws.org/Vol-1173/`, 2007.

[175] A. Nardi, C. Peters, J. L. Vicedo, and N. Ferro, editors. *CLEF 2006 Working Notes*. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, `http://ceur-ws.org/Vol-1172/`, 2006.

[176] R. Nordlie and N. Pharo. Seven Years of INEX Interactive Retrieval Experiments - Lessons and Challenges. In Catarci et al. [53], pages 13–23.

[177] S. Nowak and P. Dunker. Overview of the CLEF 2009 Large-Scale Visual Concept Detection and Annotation Task. In Peters et al. [209], pages 94–109.

[178] S. Nowak and M. J. Huiskes. New Strategies for Image Annotation: Overview of the Photo Annotation Task at ImageCLEF 2010. In Braschler et al. [46].

[179] S. Nowak, K. Nagel, and J. Liebetrau. The CLEF 2011 Photo Annotation and Concept-based Retrieval Tasks. In Petras et al. [214].

[180] D. W. Oard and J. Gonzalo. The CLEF 2001 Interactive Track. In Peters et al. [195], pages 308–319.

[181] D. W. Oard and J. Gonzalo. The CLEF 2003 Interactive Track. In Peters et al. [197], pages 425–434.

[182] D. W. Oard, J. Wang, G. J. F. Jones, R. W. White, P. Pecina, D. Soergel, X. Huang, and I. Shafran. Overview of the CLEF-2006 Cross-Language Speech Retrieval Track. In Peters et al. [198], pages 744–758.

[183] M.-D. Olvera-Lobo and J. Gutiérrez-Artacho. Multilingual Question-Answering System in Biomedical Domain on the Web: An Evaluation. In Forner et al. [84], pages 83–88.

[184] N. Orio, C. C. S. Liem, G. Peeters, and M. Schedl. MusiClef: Multimodal Music Tagging Task. In Catarci et al. [53], pages 36–41.

[185] N. Orio and D. Rizo. Overview of MusiCLEF 2011. In Petras et al. [214].

[186] P. Pecina, P. Hoffmannová, G. J. F. Jones, Y. Zhang, and D. W. Oard. Overview of the CLEF-2007 Cross-Language Speech Retrieval Track. In Peters et al. [207], pages 674–686.

[187] A. Peñas, P. Forner, A. Rodrigo, R. F. E. Sutcliffe, C. Forascu, and C. Mota. Overview of ResPubliQA 2010: Question Answering Evaluation over European Legislation. In Braschler et al. [46].

[188] A. Peñas, P. Forner, R. F. E. Sutcliffe, A. Rodrigo, C. Forascu, I. Alegria, D. Giampiccolo, N. Moreau, and P. Osenova. Overview of ResPubliQA 2009: Question Answering Evaluation over European Legislation. In Peters et al. [201], pages 174–196.

[189] A. Peñas, E. H. Hovy, P. Forner, A. Rodrigo, R. F. E. Sutcliffe, C. Forascu, and C. Sporleder. Overview of QA4MRE at CLEF 2011: Question Answering for Machine Reading Evaluation. In Petras et al. [214].

[190] A. Peñas, E. H. Hovy, P. Forner, A. Rodrigo, R. F. E. Sutcliffe, and R. Morante. QA4MRE 2011-2013: Overview of Question Answering for Machine Reading Evaluation. In Forner et al. [86], pages 303–320.

[191] A. Peñas, E. H. Hovy, P. Forner, A. Rodrigo, R. F. E. Sutcliffe, C. Sporleder, C. Forascu, Y. Benajiba, and P. Osenova. Overview of QA4MRE at CLEF 2012: Question Answering for Machine Reading Evaluation. In Forner et al. [85].

[192] A. Peñas, C. Unger, and A.-C. and Ngonga Ngomo. Overview of CLEF Question Answering Track 2014. In Kanoulas et al. [129], pages 300–306.

[193] C. Peters, editor. *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.

[194] C. Peters, M. Braschler, and P. Clough. *Multilingual Information Retrieval*. Springer-Verlag, Heidelberg, Germany, 2011.

[195] C. Peters, M. Braschler, J. Gonzalo, and M. Kluck, editors. *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.

[196] C. Peters, M. Braschler, J. Gonzalo, and M. Kluck, editors. *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.

[197] C. Peters, M. Braschler, J. Gonzalo, and M. Kluck, editors. *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.

[198] C. Peters, P. Clough, F. C. Gey, J. Karlgren, B. Magnini, D. W. Oard, M. de Rijke, and M. Stempfhuber, editors. *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.

[199] C. Peters, P. Clough, J. Gonzalo, G. J. F. Jones, M. Kluck, and B. Magnini, editors. *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.

[200] C. Peters, T. Deselaers, N. Ferro, J. Gonzalo, G. J. F. Jones, M. Kurimo, T. Mandl, and A. Peñas, editors. *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.

[201] C. Peters, G. M. Di Nunzio, M. Kurimo, T. Mandl, D. Mostefa, A. Peñas, and G. Roda, editors. *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.

[202] C. Peters and N. Ferro, editors. *CLEF 2000 Working Notes*. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, `http://ceur-ws.org/Vol-1166/`, 2000.

[203] C. Peters and N. Ferro, editors. *CLEF 2001 Working Notes*. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, `http://ceur-ws.org/Vol-1167/`, 2001.

[204] C. Peters and N. Ferro, editors. *CLEF 2002 Working Notes*. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, `http://ceur-ws.org/Vol-1168/`, 2002.

[205] C. Peters and N. Ferro, editors. *CLEF 2003 Working Notes*. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, `http://ceur-ws.org/Vol-1169/`, 2003.

[206] C. Peters, F. C. Gey, J. Gonzalo, G. J. F. Jones, M. Kluck, B. Magnini, H. Müller, and M. de Rijke,

editors. *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.

[207] C. Peters, V. Jijkoun, T. Mandl, H. Müller, D. W. Oard, A. Peñas, V. Petras, and D. Santos, editors. *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.

[208] C. Peters, V. Quochi, and N. Ferro, editors. *CLEF 2005 Working Notes*. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, `http://ceur-ws.org/Vol-1171/`, 2005.

[209] C. Peters, T. Tsikrika, H. Müller, J. Kalpathy-Cramer, G. J. F. Jones, J. Gonzalo, and B. Caputo, editors. *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.

[210] V. Petras and S. Baerisch. The Domain-Specific Track at CLEF 2008. In Peters et al. [200], pages 186–198.

[211] V. Petras, S. Baerisch, and M. Stempfhuber. The Domain-Specific Track at CLEF 2007. In Peters et al. [207], pages 160–173.

[212] V. Petras, T. Bogers, M. Hall, J. Savoy, P. Malak, A. Pawlowski, N. Ferro, and I. Masiero. Cultural Heritage in CLEF (CHiC) 2013. In Forner et al. [86], pages 192–211.

[213] V. Petras, N. Ferro, M. Gäde, A. Isaac, M. Kleineberg, I. Masiero, M. Nicchio, and J. Stiller. Cultural Heritage in CLEF (CHiC) Overview 2012. In Forner et al. [85].

[214] V. Petras, P. Forner, P. Clough, and N. Ferro, editors. *CLEF 2011 Working Notes*. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, `http://ceur-ws.org/Vol-1177/`, 2011.

[215] F. Piroi. CLEF-IP 2010: Retrieval Experiments in the Intellectual Property Domain. In Braschler et al. [46].

[216] F. Piroi, M. Lupu, and A. Hanbury. Effects of Language and Topic Size in Patent IR: An Empirical Study. In Catarci et al. [53], pages 54–66.

[217] F. Piroi, M. Lupu, and A. Hanbury. Overview of CLEF-IP 2013 Lab - Information Retrieval in the Patent Domain. In Forner et al. [86], pages 232–249.

[218] F. Piroi, M. Lupu, A. Hanbury, A. P. Sexton, W. Magdy, and I. V. Filippov. CLEF-IP 2012: Retrieval Experiments in the Intellectual Property Domain. In Forner et al. [85].

[219] F. Piroi, M. Lupu, A. Hanbury, and V. Zenz. CLEF-IP 2011: Retrieval in the Intellectual Property Domain. In Petras et al. [214].

[220] A. Popescu, T. Tsikrika, and J. Kludas. Overview of the Wikipedia Retrieval Task at ImageCLEF 2010. In Braschler et al. [46].

[221] M. Potthast, A. Barrón-Cedeño, A. Eiselt, B. Stein, and P. Rosso. Overview of the 2nd International Competition on Plagiarism Detection. In Braschler et al. [46].

[222] M. Potthast, A. Eiselt, A. Barrón-Cedeño, B. Stein, and P. Rosso. Overview of the 3rd International Competition on Plagiarism Detection. In Petras et al. [214].

[223] M. Potthast, T. Gollub, M. Hagen, J. Kiesel, M. Michel, A. Oberländer, M. Tippmann, A. Barrón-Cedeño, P. Gupta, P. Rosso, and B. Stein. Overview of the 4th International Competition on Plagiarism Detection. In Forner et al. [85].

[224] M. Potthast, T. Gollub, F. Rangel Pardo, P. Rosso, E. Stamatatos, and B. Stein. Improving the Reproducibility of PAN's Shared Tasks: Plagiarism Detection, Author Identification, and Author Profiling. In Kanoulas et al. [129], pages 268–299.

[225] M. Potthast and T. Holfeld. Overview of the 2nd International Competition on Wikipedia Vandalism Detection. In Petras et al. [214].

[226] M. Potthast, B. Stein, and T. Holfeld. Overview of the 1st International Competition on Wikipedia Vandalism Detection. In Braschler et al. [46].

[227] A. Pronobis, M. Fornoni, H. I. Christensen, and B. Caputo. The Robot Vision Track at ImageCLEF 2010. In Braschler et al. [46].

[228] A. Pronobis, L. Xing, and B. Caputo. Overview of the CLEF 2009 Robot Vision Track. In Peters et al.

[209], pages 110–119.

[229] D. Rebholz-Schuhmann, S. Clematide, F. Rinaldi, S. Kafkas, E. M. van Mulligen, Q.-C. Bui, J. Hellrich, J. Lewin, D. Milward, M. Poprat, A. Jimeno-Yepes, U. Hahn, and J. A. Kors. Entity Recognition in Parallel Multi-lingual Biomedical Corpora: The CLEF-ER Laboratory Overview. In Forner et al. [86], pages 353–367.

[230] N. Rekabsaz and M. Lupu. A Real-World Framework for Translator as Expert Retrieval. In Kanoulas et al. [129], pages 141–152.

[231] G. Roda, J. Tait, F. Piroi, and V. Zenz. CLEF-IP 2009: Retrieval Experiments in the Intellectual Property Domain. In Peters et al. [201], pages 385–409.

[232] A. Rodrigo, A. Peñas, and M. F. Verdejo. Overview of the Answer Validation Exercise 2008. In Peters et al. [200], pages 296–313.

[233] R. Roller and M. Stevenson. Self-supervised Relation Extraction Using UMLS. In Kanoulas et al. [129], pages 116–127.

[234] B. R. Rowe, D. W. Wood, A. L. Link, and D. A. Simoni. *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.

[235] S. Sabetghadam, R. Bierig, and A. Rauber. A Hybrid Approach for Multi-faceted IR in Multimodal Domain. In Kanoulas et al. [129], pages 86–97.

[236] E. SanJuan, V. Moriceau, X. Tannier, P. Bellot, and J. Mothe. Overview of the INEX 2012 Tweet Contextualization Track. In Forner et al. [85].

[237] D. Santos and L. M. Cabral. GikiCLEF: Expectations and Lessons Learned. In Peters et al. [201], pages 212–222.

[238] D. Savenkov, P. Braslavski, and M. Lebedev. Search Snippet Evaluation at Yandex: Lessons Learned and Future Directions. In Forner et al. [84], pages 14–25.

[239] P. Schaer. Better than Their Reputation? On the Reliability of Relevance Assessments with Students. In Catarci et al. [53], pages 124–135.

[240] P. Schaüble and P. Sheridan. Cross-Language Information Retrieval (CLIR) Track Overview. In E. M. Voorhees and D. K. Harman, editors, *The Sixth Text REtrieval Conference (TREC-6)*, pages 31–44. National Institute of Standards and Technology (NIST), Special Publication 500-240, Washington, USA., 1997.

[241] A. Schuth and M. Marx. Evaluation Methods for Rankings of Facetvalues for Faceted Search. In Forner et al. [84], pages 131–136.

[242] B. Sigurbjörnsson, J. Kamps, and M. de Rijke. Overview of WebCLEF 2005. In Peters et al. [206], pages 810–824.

[243] P. Sorg, P. Cimiano, A. Schultz, and S. Sizov. Overview of the Cross-lingual Expert Search (CriES) Pilot Challenge. In Braschler et al. [46].

[244] D. Spina, E. Amigó, and J. Gonzalo. Filter Keywords and Majority Class Strategies for Company Name Disambiguation in Twitter. In Forner et al. [84], pages 38–49.

[245] V. Stefanov, A. Sachs, M. Kritz, M. Samwald, M. Gschwandtner, and A. Hanbury. A Formative Evaluation of a Comprehensive Search System for Medical Professionals. In Forner et al. [86], pages 81–92.

[246] M. Stempfhuber and S. Baerisch. The Domain-Specific Track at CLEF 2006: Overview of Approaches, Results and Assessment. In Peters et al. [198], pages 163–169.

[247] H. Suominen. CLEFeHealth2012 - The CLEF 2012 Workshop on Cross-Language Evaluation of Methods, Applications, and Resources for eHealth Document Analysis. In Forner et al. [85].

[248] H. Suominen, S. Salanterä, S. Velupillai, W. Webber Chapman, G. K. Savova, N. Elhadad, S. Pradhan, B. R. South, D. L. Mowery, G. J. F. Jones, J. Leveling, L. Kelly, L. Goeuriot, D. Martínez, and G. Zuccon. Overview of the ShARe/CLEF eHealth Evaluation Lab 2013. In Forner et al. [86], pages 212–231.

[249] W. Tannebaum and A. Rauber. Mining Query Logs of USPTO Patent Examiners. In Forner et al. [86], pages 136–142.

[250] B. Thomee and A. Popescu. Overview of the ImageCLEF 2012 Flickr Photo Annotation and Retrieval Task. In Forner et al. [85].

[251] C. V. Thornley, A. C. Johnson, A. F. Smeaton, and H. Lee. The Scholarly Impact of TRECVid (2003–2009). *Journal of the American Society for Information Science and Technology (JASIST)*, 62(4):613–627, April

2011.

[252] T. Tommasi, B. Caputo, P. Welter, M. O. Güld, and T. M. Deserno. Overview of the CLEF 2009 Medical Image Annotation Track. In Peters et al. [209], pages 85–93.

[253] M. Trappett, S. Geva, A. Trotman, F. Scholer, and M. Sanderson. Overview of the INEX 2012 Snippet Retrieval Track. In Forner et al. [85].

[254] T. Tsikrika, A. Garcia Seco de Herrera, and H. Müller. Assessing the Scholarly Impact of ImageCLEF. In Forner et al. [84], pages 95–106.

[255] T. Tsikrika and J. Kludas. Overview of the WikipediaMM Task at ImageCLEF 2008. In Peters et al. [200], pages 539–550.

[256] T. Tsikrika and J. Kludas. Overview of the WikipediaMM Task at ImageCLEF 2009. In Peters et al. [209], pages 60–71.

[257] T. Tsikrika, B. Larsen, H. Müller, S. Endrullis, and E. Rahm. The Scholarly Impact of CLEF (2000–2009). In Forner et al. [86], pages 1–12.

[258] T. Tsikrika, A. Popescu, and J. Kludas. Overview of the Wikipedia Image Retrieval Task at ImageCLEF 2011. In Petras et al. [214].

[259] M. Turchi, J. Steinberger, M. Alexandrov Kabadjov, and R. Steinberger. Using Parallel Corpora for Multilingual (Multi-document) Summarisation Evaluation. In Agosti et al. [8], pages 52–63.

[260] J. Turmo, P. Comas, S. Rosset, O. Galibert, N. Moreau, D. Mostefa, P. Rosso, and D. Buscaldi. Overview of QAST 2009. In Peters et al. [201], pages 197–211.

[261] J. Turmo, P. Comas, S. Rosset, L. Lamel, N. Moreau, and D. Mostefa. Overview of QAST 2008. In Peters et al. [200], pages 296–313.

[262] A. Vallin, B. Magnini, D. Giampiccolo, L. Aunimo, C. Ayache, P. Osenova, A. Peñas, M. de Rijke, B. Sacaleanu, D. Santos, and R. F. E. Sutcliffe. Overview of the CLEF 2005 Multilingual Question Answering Track. In Peters et al. [206], pages 307–331.

[263] F. Valverde-Albacete, J. Carrillo de Albornoz, and C. Peláez-Moreno. A Proposal for New Evaluation Metrics and Result Visualization Technique for Sentiment Analysis Tasks. In Forner et al. [86], pages 41–42.

[264] M. Villegas and R. Paredes. Overview of the ImageCLEF 2012 Scalable Web Image Annotation Task. In Forner et al. [85].

[265] E. M. Voorhees. TREC: Continuing Information Retrieval's Tradition of Experimentation. *Communications of the ACM (CACM)*, 50(11):51–54, November 2007.

[266] A. Walker, A. Starkey, J. Z. Pan, and A. Siddharthan. Making Test Corpora for Question Answering More Representative. In Kanoulas et al. [129], pages 1–6.

[267] Q. Wang, J. Kamps, G. Ramírez Camps, M. Marx, A. Schuth, M. Theobald, S. Gurajada, and A. Mishra. Overview of the INEX 2012 Linked Data Track. In Forner et al. [85].

[268] X. Wang, Wang. X., and Q. Zhang. A Web-Based CLIR System with Cross-Lingual Topical Pseudo Relevance Feedback. In Forner et al. [86], pages 104–107.

[269] R. W. White, D. W. Oard, G. J. F. Jones, D. Soergel, and X. Huang. Overview of the CLEF-2005 Cross-Language Speech Retrieval Track. In Peters et al. [206], pages 744–759.

[270] T. Wilhelm-Stein and M. Eibl. A Quantitative Look at the CLEF Working Notes. In Forner et al. [86], pages 13–16.

[271] T. Wilhelm-Stein, R. Herms, M. Ritter, and M. Eibl. Improving Transcript-Based Video Retrieval Using Unsupervised Language Model Adaptation. In Kanoulas et al. [129], pages 110–115.

[272] X. Yan, G. Gao, X. Su, H. Wei, X. Zhang, and Q. Lu. Hidden Markov Model for Term Weighting in Verbose Queries. In Catarci et al. [53], pages 82–87.

[273] H. Zamani, H. N. Esfahani, P. Babaie, S. Abnar, M. Dehghani, and A. Shakery. Authorship Identification Using Dynamic Selection of Features from Probabilistic Feature Set. In Kanoulas et al. [129], pages 128–140.

[274] D. Zellhöfer. Overview of the Personal Photo Retrieval Pilot Task at ImageCLEF 2012. In Forner et al. [85].

[275] L. Zhang, A. Rettinger, M. Färber, and M. Tadic. A Comparative Evaluation of Cross-Lingual Text Annotation Techniques. In Forner et al. [86], pages 124–135.