

From XML Retrieval to Semantic Search and Beyond



The INEX, SBS, and MC2 Labs of CLEF 2012–2018

Jaap Kamps, Marijn Koolen, Shlomo Geva, Ralf Schenkel, Eric SanJuan, and Toine Bogers

Abstract INEX ran as an independent evaluation forum for 10 years before it teamed up with CLEF in 2012. Even before 2012 there was considerable collaboration between INEX and CLEF, and these collaborations increased in intensity when CLEF moved beyond its traditional cross-lingual focus in 2009/2010 shifting to include all experimental IR. This led to the merger of CLEF and INEX, and effectively to the inclusion of INEX as a large track or lab into CLEF in 2012. This chapter details the efforts of the INEX lab in CLEF (2012–2014), as well as the ongoing activities as separate labs, under the labels Social Book Search (2015–2016), and Microblog Contextualization (2016–2018).

J. Kamps (✉)
University of Amsterdam, Amsterdam, Netherlands
e-mail: kamps@uva.nl

M. Koolen
Royal Netherlands Academy of Arts and Sciences, Amsterdam, Netherlands
e-mail: marijn.koolen@di.huc.knaw.nl

S. Geva
QUT, Brisbane, QLD, Australia
e-mail: s.geva@qut.edu.au

R. Schenkel
University of Trier, Trier, Germany
e-mail: schenkel@uni-trier.de

E. SanJuan
University of Avignon, Avignon, France
e-mail: eric.sanjuan@uni-avignon.fr

T. Bogers
Aalborg University Copenhagen, Copenhagen, Denmark
e-mail: toine@hum.aau.dk

1 Introduction

This chapter documents the INEX lab of CLEF, running 2012–2014, including INEX tracks that continued as independent CLEF labs: the CLEF Social Book Search Lab (2015–2016) and the CLEF Cultural Microblog Contextualization Workshop and Lab (2016–2018). The emphasis is on the INEX and Social Book Search Labs as the Microblog Contextualization Lab was still ongoing at the time of writing.

No single chapter can do justice to the massive amount of work done in the INEX and follow up labs, so this chapter is merely meant as a starting point with references to the respective overview papers providing full details on the tracks and resulting test collections. By providing a high level, but comprehensive overview of the wealth of activities spanning many years, we hope to shed some light on important developments in the field during these years, and relations between various activities inside and outside CLEF that may not be immediately apparent for those familiar with only part of the activities. This highlights also two of the key strengths and contributions of INEX in particular, and CLEF in general. First, it is about the people: the open format and volunteer run activities allowed many (young) researchers to get involved in every aspects, from participant discussion to task or track organization, and has educated a generation of researchers now taking up leadership positions in the field. Second, many of the activities had some degree of success inside the respective CLEF campaign, but also have much greater impact by reuse of the benchmarks in publications, including creative reuse in unexpected settings, and by instigating new activities, both happening in future editions of the track, but also outside the lab or outside CLEF.

This chapter is structured as follows. You are now at the end of the introduction in Sect. 1. Next, in Sect. 2, we document the context of INEX joining CLEF, covering the years 2002–2011. The main part of this chapter is in Sect. 3, in which we document the INEX tracks as part of CLEF 2012–2014. This is followed by a discussion of the two follow up labs seeded from INEX: the *Social Book Search* lab as part of CLEF 2014–2016 in Sect. 4, and the *Cultural Microblog Contextualization* workshop and lab as part of CLEF 2016–2018 in Sect. 5. We close off by providing further discussion and reflection in the final Sect. 6.

2 INEX Before CLEF

In this section, we will discuss the context of INEX joining CLEF, very briefly covering the years 2002–2011.

2.1 INEX 2002–2011

The *INitiative for the Evaluation of XML Retrieval (INEX)* was founded in 2002 by a group of people led by Mounia Lalmas and Norbert Fuhr. Collaborations between INEX and *Conference and Labs of the Evaluation Forum (CLEF)* date back to these early days. For a good overview of the first 10 years of INEX, we refer to the proceedings and overview papers, and to the chapters in the Springer Encyclopedia of Database Systems (Kazai 2009, 2018). We restrict our attention here to the relation between the INEX and CLEF evaluation forums before the merge in 2012.

Perhaps not known by many people, the only source of funding ever supporting INEX was a modest contribution from the DELOS Network of Excellence, an EU funded initiative to support and promote digital libraries in the period 1997–2007 (Thanos and Casarosa 2017). And apart from incidental, large, support from dedicated EU projects, this same, small support from DELOS presented the life-line of CLEF throughout these years. Within DELOS, INEX and CLEF were co-responsible for the evaluation activities, and INEX always promoted the inclusion of CLEF in these activities even when DELOS seemed more keen on the relevance of INEX for the DL use case. DELOS was also heavily involved in the *European Conference on Digital Libraries (ECDL)* conference, now continued as the *Theory and Practice of Digital Libraries (TPDL)* conference, that provided a home for CLEF as one of its satellite workshops. Looking back over the DELOS years, Thanos and Casarosa (2017, p. 305) list both CLEF and INEX as the most important spin-offs of the DELOS activities.

Until CLEF 2010, when the CLEF focus was firmly on multi-language retrieval, there was incidental direct collaboration between INEX and CLEF. Let us just mention two of the main examples. First, as INEX started running a large-scale Wikipedia image retrieval task (2006–2007), but had few multimedia retrieval participants, it was decided to move this task to CLEF and make it part of the booming CLEF ImageCLEF track, which turned out to be a very good choice attracting substantial participation to both the WikipediaMM task (continuing 2008–2011) and to ImageCLEF. Second, INEX was worried about the state of IR evaluation when the scale of test collections increased and the retrieval tasks became more specialized. This led to a SIGIR workshop on the Future of Information Retrieval Evaluation run at SIGIR 2009 in Boston (Kamps et al. 2009), which was jointly organized by the main chairs of INEX, CLEF, TREC, and NTCIR. INEX chairs who took the initiative for this workshop were particularly overwhelmed by the support and encouragement they received from CLEF. This was an important workshop that had considerable impact in years to come, by creating a vision and longer term agenda, and the clear conviction that more direct collaboration between the different evaluation forums was urgently needed.

2.2 *INEX Joining CLEF*

In light of the above, there are three major factors that caused the merger of INEX and CLEF in 2012. First, CLEF decided in 2009 to refocus and break free from the earlier multi-language niche, as well as decided to host their own conference as they had simply outgrown the setup as workshop associated with ECDL. As a case in point, CLEF 2009 was far larger than ECDL 2009, even though CLEF was officially still just one of the ECDL satellite workshops. Second, INEX lost its old home of Dagstuhl after 2008, and found it increasingly difficult to organize its own event in December, competing with ADCS, NTCIR, AIRS, . . . , and the holiday season, and was considering moving to a different location and time-slot. Third, with the need for further collaboration between evaluation initiatives in the air, the wish of CLEF to broaden its tracks beyond the traditional multi-language tasks, and the desire of INEX to find a new home, the merger of INEX and CLEF seemed a win-win for everyone involved.

Hence in 2012, INEX and CLEF merged into a single event, and INEX joined as “superlab” with many tracks, similar to ImageCLEF at the time. In 2011, INEX already changed its track structure in anticipation of the CLEF merger. Note that the INEX 2011 workshop was held in December 2011 near Saarbrücken, and the next edition of INEX as part of CLEF was scheduled for September 2012 in Rome. This made INEX 2012 a proverbial nine-months cycle, putting considerable stress on the INEX volunteer-run organization to deliver the baby in time for CLEF 2012.

The core of INEX had always been its *Ad hoc search* track, running keyword and structured queries against a structured corpus, allowing to retrieve any document part as a result, but evaluated as a proper IR or search task against topical relevance. The Ad hoc track also consumed most of the resources, and had the largest number of participants, effectively competing with an increasing number of other tracks. To promote the other activities, the general chairs, in consultation with the INEX steering committee, made the difficult decision to stop the ad hoc track in 2011, and focus on five activities. The last pre-CLEF edition INEX 2011 (Bellot et al. 2012b) featured the following tracks:

Books and Social Search Track *investigating techniques to support users in searching and navigating books, metadata and complementary social media. The Social Search for Best Books Task studies the relative value of authoritative metadata and user-generated content using a collection based on data from Amazon and LibraryThing. The Prove It Task asks for pages confirming or refuting a factual statement, using a corpus of the full texts of 50k digitized books.*

Data Centric Track *investigating retrieval over a strongly structured collection of documents based on IMDb. The Ad Hoc Search Task has informational requests to be answered by the entities in IMDb (movies, actors, directors, etc.). The Faceted Search Task asks for a restricted list of facets and facet-values that will optimally guide the searcher toward relevant information.*

Question Answering Track *investigating tweet contextualization, answering questions of the form “what is this tweet about?” with a synthetic summary of contextual information grasped from Wikipedia and evaluated by both the relevant text retrieved, and the “last point of interest.”*

Relevance Feedback Track *investigating the utility of incremental passage level relevance feedback by simulating a searcher’s interaction. An unconventional evaluation track where submissions are executable computer programs rather than search results.*

Snippet Retrieval Track *investigating how to generate informative snippets for search results. Such snippets should provide sufficient information to allow the user to determine the relevance of each document, without needing to view the document itself.*

Where as the INEX Relevance Feedback and Snippet Retrieval tracks were mature and continued from earlier years, the other three tracks were completely redesigned in anticipation of the CLEF merger: the social data of the INEX Book Track was added, the tweet contextualization focus emerged in the INEX QA Track, and work on highly structured data was started at the INEX Data Centric Track.

3 INEX at CLEF

In this section, we document the INEX labs as part of CLEF 2012–2014. Table 1 gives an overview of the INEX activities over the year, including the specific tracks, with references to the respective overview papers.

3.1 CLEF 2012

INEX 2012 (Bellot et al. 2012a) was the first INEX held as part of CLEF 2012, which was the eleventh annual edition of INEX. Recall that the last independent

Table 1 CLEF INEX tracks over the years

Tracks	2012	2013	2014
INEX overview	Bellot et al. (2012a)	Bellot et al. (2013a)	Bellot et al. (2014a)
Linked data	Wang et al. (2012)	Gurajada et al. (2013)	
Relevance feedback	Chappell and Geva (2012)		
Snippet retrieval	Trappett et al. (2012)	Trappett et al. (2013)	
Social book search	Koolen et al. (2012)	Koolen et al. (2013)	Koolen et al. (2014)
Interactive SBS			Hall et al. (2014)
Tweet contextualization	SanJuan et al. (2012)	Bellot et al. (2013b)	Bellot et al. (2014b)

INEX was held in December 2011, making the cycle to CLEF 2012 a particularly short year to run the full cycle of test collection development. INEX 2012 had again five tracks, all based on or derived from, the INEX 2011 tracks, which we will describe in this section.

Linked Data (LD) (Wang et al. 2012) was a new track but a direct descendant of the INEX 2011 Data-Centric Track, moving from the highly structured IMDb data to a rich textual corpus (Wikipedia) with rich semantic annotation (DBpedia).

Linked Data Track *investigating retrieval over a strongly structured collection of documents based on DBpedia and Wikipedia. The Ad Hoc Search Task has informational requests to be answered by the entities in DBpedia/Wikipedia. The Faceted Search Task asks for a restricted list of facets and facet-values that will optimally guide the searcher toward relevant information.*

LD did amazing efforts to create novel benchmarks at scale bringing linked data within the scope of IR. Perhaps this was too early, and the short nine-months cycle of INEX 2012 didn't help, as participation was small, and there were significant issues in creating the massive corpus merging Wikipedia with DBpedia and Yago linked data (but these issues were fixed in the next year).

Relevance Feedback (RF) (Chappell and Geva 2012) was a continuation of the track in INEX 2011, and the final edition of the track. The track studied (incremental) relevance feedback on the INEX Wikipedia Corpus from 2009 by reusing topics and judgments from earlier year. As it outgrew the Cranfield style evaluation effort, due to continuous submissions and scoring, it also lost some of the momentum of an annual cycle with clear deadlines, and it was decided not to continue it into 2013 but to offer it online to any interested party.

Snippet Retrieval (SR) (Trappett et al. 2012) was a continuing track of INEX 2011, investigating how to generate informative snippets for search results. This track also used the INEX Wikipedia Corpus from 2009, and reused some old ad hoc search topics plus created novel topics and judgments at both summary and document level. As the snippet retrieval track ran late in 2012—due to the short nine-month cycle of 2012—it was decided to carry it over to 2013, and build one final test collection for SR over the 2 years together.

Social Book Search (SBS) (Koolen et al. 2012) was another new track but a direct descendant of the INEX 2011 Books and Social Search Track, and the earlier Book Search Tracks (since 2007). Although the traditional out-of-copyright, full-text books still were continued as a task by special demand of those interested, the focus clearly shifted to the social book data, coming from Amazon and LibraryThing and originally constructed to support the INEX Interactive Track (running 2004–2010), including real-world complex book search requests, user profiles and personal book catalogues.

Tweet Contextualization (TC) (SanJuan et al. 2012) was also a new track, but directly derived from the INEX 2011 Question Answering Track, which focused on more NLP-oriented tasks and moved to multidocument summarization.

Tweet Contextualization Track *investigating tweet contextualization, answering questions of the form “what is this tweet about?” with a synthetic summary*

of contextual information grasped from Wikipedia and evaluated by both the relevant text retrieved, and the “last point of interest.”

The use case was based on short tweets or posts, which cannot be fully comprehensive, and to look for background information in Wikipedia about relevant entities, concepts, events, products, etc., referred to by the tweet or post. As the INEX Wikipedia corpus was getting dated, a new Wikipedia dump of 2012 was used. Topics were 1000 tweets, of which 63 were evaluated, both in terms of a synthetic measure against reference summaries, as well as by a human judgment on the readability of the whole summary.

At CLEF 2012 in Rome, the collocation with or embedding in the larger CLEF family worked out very well for INEX. There was a considerable new interest in the INEX tracks from other CLEF attendees, as INEX helped CLEF break free from the earlier focus on cross-language retrieval. During the CLEF organizers meeting there was active discussion on more direct collaboration between the different CLEF labs, which was strongly supported by the INEX organizers.

There are two major activities related to INEX that took place outside CLEF. First, there was the spin-off *Exploiting Semantic Annotations for Information Retrieval (ESAIR)* workshop that took place at CIKM 2012 at Maui (Kamps et al. 2013), featuring a range of papers and great keynotes on Knowledge Graphs (Evgeniy Gabrilovich), and Conversational Search (Ron Kaplan). This ESAIR workshop was continued from CIKM 2011 in Glasgow, and CIKM 2010 in Toronto. Second, a new track on *Contextual Suggestion* was started, running at the *Text REtrieval Conference (TREC)* 2012 (Dean-Hall et al. 2012), rather than at CLEF, to attract a wider attendance. This track was a direct result of the SIGIR 2011 workshop on *Supporting Complex Search Tasks: Entertain Me* (Belkin et al. 2011), which in turn was a spin-off of discussion at INEX 2010 (Beckers et al. 2010).

3.2 CLEF 2013

INEX 2013 (Bellot et al. 2013a) was the twelfth annual run of INEX, and the second as part of the CLEF family. With all the changes instigated in 2011 and 2012, INEX 2013 was a year of continuity, with training data for all tasks being available from 2012 and new, additional test collections being developed in 2013.

To better align with the general CLEF structure, INEX 2013 featured three *themes*: (1) searching professional and user generated data (SBS); searching structured or semantic data (LD); and focused retrieval (SR and TC). That is, INEX 2013 featured a total of four tasks, which we will discuss in the following.

Linked Data (LD) (Gurajada et al. 2013) managed to address the corpus issues of 2012, when a large part of the corpus was ultimately removed as it caused validity problems with the strict schema. This produced a massive linked data search corpus covering the core of the linked data graph (constituted by DBpedia) and rich textual sources with semantic annotations (the corresponding Wikipedia). The Faceted

Search task received less attention, due to the need for general queries with very large answer sets, rather than very selective queries expressing detailed information need. Of special mention is the new Jeopardy task, attempting to encourage writing or generating rich SPARQL style queries expressing complex needs. Although the LD track did not continue in 2014, the resulting corpus and test collections are a key resource and have been widely reused inside IR and beyond.

As explained in the section above, *Snippet Retrieval (SR)* (Trappett et al. 2013) was carried over from INEX 2012 as it ran so late for 2012 that it was decided to view this as early for 2013, effectively creating a 21 month cycle for INEX 2013. So the same topic set was used, additional runs were requested, and all judgments on snippet and document relevance were completed. As now a wealth of data for evaluating snippet retrieval was available, INEX 2013 presented the last year of the SR track.

Social Book Search (SBS) (Koolen et al. 2013) continued strong with the social book data. In particular the detailed statements of request, detailed user profiles and personal book catalogues, and expert book recommendations from the LibraryThing forums, proved very valuable data for research, as one of the few examples of social data being of general interest, and less fraught with privacy concerns. By popular demand of some participants, a small scale continuation of the “Prove It” task on the scanned book corpus was permitted by the track organizers, but the main focus was on the social book search tasks.

Tweet Contextualization (TC) (Bellot et al. 2013b) was picking up momentum, creating a test collection with 120 tweets over INEX 2012 and 2013. This was reflected by advance approaches, with the better systems combining NLP pipelines, twitter specific processing, and IR finesse, realizing the broader track’s goals to bridge IR and NLP approaches.

At CLEF 2013 in Valencia, there was a clear feeling of return-on-investments: all tracks were now stable and rerunning, and clearly attracting new participants from the broader CLEF community. At the same time, the increasing number of labs and tracks at CLEF led to an increasing competition for attention, and to thinning down the participation per track and task. This already led to the introduction of (or rather read: the restriction to) three themes in 2013 mentioned above, and continued in 2014 with pressure to reduce the number of tracks within INEX. This led the main organizers of INEX to discussions on whether to continue the INEX as a track, or split up the INEX tracks into 2–3 separate CLEF tracks, hence effectively discontinuing INEX. On a more positive note, intense discussion with the *Cultural Heritage in CLEF (CHiC)* lab led to the decision to fold CHiC into the Social Book Search Track of INEX in 2014, as an *Interactive Social Book Search (iSBS)* track.

There are three major activities related to INEX that took place outside CLEF. First, the spin-off ESAIR workshop continued at CIKM 2013 in San Francisco (Bennett et al. 2014), with a range of papers and keynotes on Wikification (Dan Roth), Reading difficulty annotation (Kevyn Collins-Thompson) and UI/UX for semantic annotations (Marti Hearst). Second, the spin-off *Contextual Suggestion Track* continued at TREC in 2013 (Dean-Hall et al. 2013). Third, a spin-off of the Social Book Search Track was the Structure Extraction (SE) task ran at ICDAR 2013

(Doucet et al. 2013), with the aim of evaluating automatic techniques for deriving structure from OCR and building hyperlinked table of contents.

3.3 CLEF 2014

INEX 2014 (Bellot et al. 2014a) was the thirteenth annual run of INEX, and the third as part of the CLEF family. Although INEX originally merged into CLEF as a “super lab” having a large number of tracks like ImageCLEF at the time, and now also absorbed the CLEF’s CHiC lab, it fully complied with the CLEF structure and limited itself to three tracks. Hence, INEX 2014 was more focused and featured three tasks, described now.

Interactive Social Book Search (iSBS) (Hall et al. 2014) was an exciting merge of the CHiC and INEX communities, with overwhelming support of the broader information science community, in order to run a large scale interactive track.

Interactive Social Book Search Track *investigating user information seeking behavior when interacting with various sources of information, for realistic task scenarios, and how the user interface impacts search and the search experience.*

The CHiC track (Petras et al. 2012, 2013) struggled with Europeana data, and with clear connections to the system-centric part of IR. INEX had run successful Interactive Tracks between 2004–2010 (Nordlie and Pharo 2012), but this line of activity lost momentum although the Amazon/LibraryThing corpus constructed for INEX iTrack experiments was still used in the SBS Track. This discussion led to federated effort of a large number of organizers to revive a user-centric interactive track at INEX, and to ensure it was seeded by the insights and challenges encountered in the system-centric SBS track. In particular, a novel type of multistage UI was developed that showed different functionality depending on the information seeking stage. Due to the relatively heavy system development in this inaugural year, a relatively modest number of 41 test persons participated in the track’s user studies.

Social Book Search (SBS) (Koolen et al. 2014) finally really ended the earlier scanned books tasks, and fully focused on the social book data also used in the iSBS track. The track boosted the number of user profiles and personal catalogues made available, far beyond those occurring in the search requests, in order to satisfy the taste of recommender system approaches. Of particular note is the seamless collaboration between the SBS and iSBS tracks, with extensive discussion between the two tracks, and the feeling that this discussion was interesting and helpful for both sides, realizing the dream of IR to fully integrate its computer science and information science parts.

Tweet Contextualization (TC) (Bellot et al. 2014b) continued strong, keeping the Wikipedia corpus stable since 2012 and 2013, but using a selection of tweets from the CLEF RepLab 2013, exploiting the extensive annotated data available for the RepLab tasks. The track was running in a similar way to 2013, except for the impact

of the changing character of the tweets used as topic, moving from general news to the specific product and services as used in the real-world social media monitoring part of RepLab. One of the main upshots of the track was that effective approaches for news tended to generalize to the specific realistic tweets derived from RepLab, demonstrating the clear value of the track's test collections. In addition to the main task in English, there was a pilot task in Spanish.

A fourth track, a continuation of the INEX Linked Data Track in collaboration with, and part of, the CLEF QA Track's QALD (QA from Linked Data) was announced, but the massive INEX corpus proved out of reach of the QALD participants and the desired collaboration didn't materialize in the end.

At CLEF 2014 in Sheffield, there were countless meetings between the iSBS and SBS track organizers, remindful of the endless discussions during the original INEX workshops at Dagstuhl. This led to an ambitious plan to boost interactive studies in a truly collaborative effort involving the largest group of organizers ever seen on any track, in any year, in any evaluation forum. . .

There are three major activities related to INEX that took place outside CLEF. First, the spin-off ESAIR workshop continued at CIKM 2014 in Shanghai (Alonso et al. 2015b), with a range of papers and keynotes on Semantic Search (Peter Mika) and Entity Linking (Silviu-Petru Cucerzan). Second, the spin-off *Contextual Suggestion Track* continued at TREC in 2014 (Dean-Hall et al. 2014). Third, a spin-off *New Trends in Content-Based Recommender Systems* (CBRecSys) workshop was held at RecSys 2014 (Bogers et al. 2014). As the SBS had content, ratings and catalogues but failed to attract significant numbers of researchers working on recommender systems to INEX and CLEF, a workshop was invented to feature the same data and tasks at RecSys. Although not many participated in the data challenge, the workshop was a resounding success and put content-based recommendation back on the agenda of RecSys as a first-class citizen.

3.4 End of INEX?

The INEX chairs decided not to submit a lab proposal for CLEF 2015. Instead, they strongly encouraged the remaining INEX tracks to submit their own proposals and become directly embedded into CLEF. First, thanks to the fruitful collaboration with CHiC and the former INEX *Interactive Track*, organizers, the INEX SBS track had great momentum and continued as a CLEF 2015 lab (to be further discussed in Sect. 4). Second, the INEX TC Track organizers were strongly advised to resubmit a significantly updated track proposal as a new CLEF lab, but this revision required more time, and it was decided to take a leap year for planning and reflection, and resubmit as a CLEF 2016 lab (to be further discussed in Sect. 5).

On a historical note, the INEX organizers already planned to step out and pass on the baton to the SBS track in 2014, as there was discussion (and some sense of agreement) to merge the TC track with CLEF RepLab. As in the end the RepLab/TC

merge didn't materialize, it was decided to continue with INEX for another year as the main CLEF lab label in 2014.

4 Social Book Search at CLEF

In this section, we document the *Social Book Search (SBS)* labs as part of CLEF 2015 and 2016. Table 2 gives an overview of the SBS activities over the year, including the specific tracks running, with references to the respective overview papers.

As detailed in the previous sections, the CLEF SBS lab stands on the shoulders of giants. It is a direct continuation of the SBS Track run at INEX since 2011, and has an even longer prehistory as the Book Search Track at INEX since 2007.

4.1 CLEF 2015

The CLEF 2015 *Social Book Search (SBS)* lab (Koolen et al. 2015a) was the first edition of the lab running as independent CLEF lab, and the fifth edition of SBS as part of the INEX family. As the CLEF SBS lab was a direct continuation of the INEX 2014 iSBS and SBS Tracks, there were two tracks running, which we will describe in further detail now.

The *Interactive Track* (Gäde et al. 2015a) was a direct continuation of the INEX iSBS track running as an extensive, concerted online user experiment.

Interactive Track *this is a user-centred track investigating how searchers use different types of metadata at various stages in the search process and how a search interface can support each stage in that process.*

An advanced system based on the Amazon/LibraryThing book collection was made available, with a baseline faceted search interface and multi-stage search interface exhibiting different functionalities depending on the information seeking stage.

The multistage interface worked with three different user interface configurations. In the start-up stage of search (pre-focus in the Kulthau/Vakkari model) where users explore the available information, the UI displays a *browse view*

Table 2 CLEF social book search tracks over the years

Tracks	2015	2016
SBS overview	Koolen et al. (2015a)	Koolen et al. (2016a)
Suggestion	Koolen et al. (2015b)	Koolen et al. (2016b)
Interactive	Gäde et al. (2015a)	Gäde et al. (2016)
Mining		Bogers et al. (2016a)

providing a query specific overview of the collection and navigation by Amazon subject classification as well as dense search results with title and ratings. In the main stage of search (focus in the Kulthau/Vakkari model) where users do an in-depth search and collect the relevant information, the UI displays the *search view* with a rich faceted search interface with more detail on each book result—this view is corresponding to advanced UIs as evolved in e-commerce and professional applications. In the final stage of search (post-focus in the Kulthau/Vakkari model) where users review and refine the selected information, and backtrack when needed, the UI displays the *book bag view* with all selected results and notes, plus providing a display of detailed information about each selected book.

In addition to the extensive, and innovative, system design and system building, about 200 test persons took part in the online user study comparing a tradition UI with the multi-stage UI, for both purposeful search and non-goal oriented search tasks, providing a very rich set of data for further analysis.

The *Suggestion Track* Koolen et al. (2015b) was the direct continuation of the INEX SBS track, focusing on IR for dealing with professional and user-generated data, and for exploring search that combines aspects of retrieval and content-based recommendation in a natural way.

Suggestion Track *this is a system-centred track focused on the comparative evaluation of systems in terms of how well they rank search results for complex book search requests that consist of both extensive natural language expressions of information needs as well as example books that reflect important aspects of those information needs, using a large collection of book descriptions with both professional metadata and user-generated content.*

The topics of 2015 were selected to include both a narrative statement of request (satisfying the needs of retrieval approaches) as well as one or more example books (satisfying the needs of recommender systems approaches). In addition, topics and book recommendations were humanly annotated, e.g., whether the example books or suggested books were positive (the requester wanted more like this, or responded positively recommended this), or negative, or neutral (no clear value judgement is expressed), in order to facilitate further analysis of the results. This led to an incredibly rich test collection, with a very high degree of realism as all requests and judgments are derived from LibraryThing forum data, supporting a wide range of experiments and deep analysis. Far more attention was given to appropriate recommender system evaluation, in order to attract those researchers to the track.

At CLEF 2015 in Toulouse, a large fraction of the many organizers were present, and there was a general sense of pride and content about the large efforts but also large achievements of this year. Of further special notice is the keynote on Polyrepresentation in Complex Book Search (Ingo Frommholz). The decision to spin off SBS as an independent CLEF lab clearly gave new impetus, and both tracks grew considerably.

There are four major activities related to SBS that took place outside CLEF.¹ First, the spin-off ESAIR workshop continued at CIKM 2015 in Melbourne (Balog et al. 2016). Second, the spin-off *Contextual Suggestion Track* continued at TREC 2015 (Dean-Hall et al. 2015). Third, the spin-off *New Trends in Content-Based Recommender Systems (CBRecSys)* workshop continued at RecSys 2015 (Bogers and Koolen 2015). Fourth, as mentioned above, a spin-off *Supporting Complex Search Tasks (SCST)* workshop was run at *European Conference on Information Retrieval (ECIR)* in Vienna (Gäde et al. 2015b). This workshop was a result of the CHiC and INEX interactive activities as part of the iSBS track, which wanted a mid-cycle deadline around ECIR, in order to finish the first round of activities timely, and reflect on the broader discussion and ways to take it forward.

4.2 CLEF 2016

The CLEF 2016 *Social Book Search (SBS)* lab (Koolen et al. 2016a) was the second run as independent CLEF lab, and sixth edition of SBS as part of the INEX family. The CLEF 2016 SBS lab continued the two existing tracks, and added a new track with a data mining/NLP focus. Hence, there were three tracks running, which we will describe in further detail now.

The *Interactive Track* Gäde et al. (2016) could cash in on all the development investments of the last 2 years, and ran a very similar track with minor refinements in the system and experimental setup, but with a large number of additional test persons in the user studies. One of the innovations was that also some of the LibraryThing requests from the forums (as used as topics in the Suggestion Track in the year before) were made part of the interactive experiments. Again a wealth of data was created (questionnaires and logs), but also again time to do a proper analysis of this rich data was running short, and regrettably only initial analysis was done and reported in the overview papers.

The *Mining Track* Bogers et al. (2016a) was a new addition to the track instigated by the observations in earlier years on the selection of suitable forum topics for use in the track, aiming to extract more information from the narrative part of the forums (or add other social media data).

Mining Track *this is a new track focused on detecting book search requests in forum posts for automatic book recommendation, as well as detecting and linking book titles in online book discussion forums.*

This led to two tasks, the first being the *book search request identification* task, in which the goal is to identify which threads on online forums are book search requests and locate the opening post with the actual request. This task directly

¹In fact, there was also a related *Graph Search and Beyond (GSB)* workshop at SIGIR 2015 in Santiago de Chili (Alonso et al. 2015a), but this was a further spin-off of INEX rather than of SBS.

serviced a need of the track, as the process of selecting suitable topics, e.g., those containing a book request proper and enough book recommendations, was always done with a combination of scripts and manual inspection, that felt suboptimal. The data of earlier years presented suitable training data for effective classifiers. The second task was the *book linking* task, in which the goal is to recognize book titles in forum posts and link them to the corresponding metadata record through their unique book ID. This task was also in response to a direct need of the track, as it was observed that most, but not all, book suggestions can be extracted from the narrative part of the forum discussions, as not all book recommendations are properly annotated in the forum data. Again, the data of earlier years, hiding the explicitly annotated instances, provided ample training data for effective classifiers. These tasks would allow the track to go beyond the LibraryThing forums, and a large dump of Reddit book related discussion was part of the track, enabling the automatic classification of raw forum data into the format required by, or most useful for, the SBS track.

The *Suggestion Track* Koolen et al. (2016b) could also cash in on all the development investments of the last 2 years, and ran a very similar track with minor refinements in the setup, but with a large number of additional run and new groups joining the track. The innovations in the track setup were minor things making the setup even more perfect. To give an example, the removal of the few out of corpus recommendations: books recommended in the forum that are not in the Amazon collection, which doesn't affect the relative system ranking but gives slightly more accurate scores. Perhaps the main innovation was in the submitted systems, where the very rich data setup was used for the first time to train very specific word embeddings, that proved to be very effective.

The SBS sessions at CLEF 2016 in Évora featured strong attendance of participants and organizers, and included a keynote on the reception of literature and book search (Pertti Vakkari). In addition to the lab at CLEF, there are two major activities related to SBS that took place outside CLEF. First, the spin-off *Contextual Suggestion Track* continued at TREC 2016 (Hashemi et al. 2016). Second, the spin-off *New Trends in Content-Based Recommender Systems* (CBRecSys) workshop continued at RecSys 2016 (Bogers et al. 2016b).

4.3 To Be Continued?

At CLEF 2016 in Évora, there was considerable discussion on the future of the SBS lab, as no new lab or track proposal was submitted to CLEF. It was decided to take a sabbatical year for the track in 2017: with no track or lab running, but a year of reflection, with the intent to come back with a new lab proposal in 2018.

There was however, a second instance of the *Supporting Complex Search Tasks* (SCST) workshop, now held at the ACM SIGIR CHIIR Conference in Oslo (Koolen et al. 2017). The CHIIR workshop generate many new plans, and clearly

demonstrated the need for a track as iSBS which promotes direct collaboration between system-centric and user-centric researchers across the field of IR.

However, despite the many great plans, no concrete proposal was made for a 2018 track at CLEF, mostly due to staffing issues, as the same group of organizers was coordinating most of the work continuously for over a decade. The work on promoting and facilitating interactive IR systems and experiments does however continue with full force at the *Barriers to Interactive IR Resources Re-use (BIIRRR)* workshops held at the *Conference on Human Information Interaction and Retrieval (CHIIR)* in 2018 (Bogers et al. 2018) and 2019 (Bogers et al. 2019).

5 Cultural Microblog Contextualization at CLEF

In this section, we document the *Microblog Contextualization (MC2)* labs as part of CLEF 2016–2018. Table 3 gives an overview of the MC2 activities over the year, including the specific tracks running, and with references to the respective overview papers.

We include a brief discussion of the MC2 lab as it has a long prehistory as the INEX Tweet Contextualization track (2012–2014), which in turn was derived from the pre-CLEF INEX QA track since 2011.

5.1 CLEF 2016

The CLEF 2016 *Cultural Microblog Contextualization (MC2)* lab (Goeriot et al. 2016) was the first edition run as independent CLEF lab, and sixth edition as part of the INEX family.

After a year of reflection amongst the organizers of the INEX Tweet Contextualization Track in 2015, the track organizers submitted a significantly changed lab proposal to CLEF 2016, which was accepted as a workshop for CLEF 2016.

It is important to stress that *Cultural Microblog Contextualization (MC2)* lab was run as a workshop, and not as a regular CLEF track or lab. This despite its heavy focus on gathering, organizing, and delivering a relevant social data related to events

Table 3 CLEF cultural microblog contextualization tracks over the years

Tracks	2016	2017	2018
MC2 overview	Goeriot et al. (2016)	Ermakova et al. (2017a)	Hajjem et al. (2018)
Workshop report	Ermakova et al. (2016)		
Content analysis		Ermakova et al. (2017b)	
Search and timeline		Goeriot et al. (2017)	
Cross language/argumentative mining			Cossu et al. (2018)

generating a large number of micro-blog posts and web documents (such as, and in particular, cultural festivals). This resulted in an impressive amount of data and a pilot task, which are described in detail in (Ermakova et al. 2016).

This extensive corpus created in 2016 was used to support a proper MC2 lab in the following year.

5.2 CLEF 2017

The CLEF 2017 *Cultural Microblog Contextualization (MC2)* lab (Ermakova et al. 2017a) was the second edition run as independent CLEF lab, and seventh as part of the INEX family.

Microblog Contextualization (MC2) deals with how cultural context of a microblog affects its social impact at large. This involves microblog search, classification, filtering, language recognition, localization, entity extraction, linking open data and summarization. MC2 at CLEF 2017 featured three tasks.

Content Analysis (Ermakova et al. 2017b) was a new track on the NLP end of the spectrum, dealing with a number of classification tasks that are a prerequisite for other tasks involving noisy social media data.

Content Analysis Track *Given a stream of microblogs, filter out microblogs dealing with festivals and perform language(s) identification, event localization, author categorization, DBpedia entities recognition and automatic summarization of linked wikipedia pages in four languages.*

Some of these subtasks were based on the filtering and priority tasks of RepLab 2014.

Microblog Search (Goeuriot et al. 2017) was a new task to locate the most relevant microblog posts in the corpus, in response to a cultural query about festivals in Arabic, English, French, or Spanish.

Microblog Search Track *Given a cultural entity as a set of Wikipedia pages: i) retrieve relevant microblogs for an entity; or ii) summarize the most informative microblogs.*

The topics and queries were extracted in various ways from social media or review corpora.

Timeline Illustration, also documented in (Goeuriot et al. 2017), had the goal of retrieving all relevant tweets dedicated to each event of a festival, according to the program provided.

Time Line Illustration Track *The goal of the Timeline illustration based on Microblogs is to provide, for each event of a cultural festival, the most interesting tweets.*

The track focused on four large festivals, two music and two theater festivals in France and the UK.

5.3 CLEF 2018

The *Cultural Microblog Contextualization (MC2)* lab continues at CLEF 2018 (Hajjem et al. 2018), which is its third run as a CLEF lab, and eighth edition as part of the INEX family.

In 2018, the *Microblog Contextualization (MC2)* lab shifted its focus to *Multilingual Cultural Mining and Retrieval*. The lab promoted developing processing methods and resources to mine the social media sphere surrounding cultural events such as festivals. This requires to deal with almost all languages and dialects as well as informal expressions. A total of three tracks were organized in 2018.

The *cross language* task (Cossu et al. 2018) was specific to movies.

Cross Language Cultural Retrieval over MicroBlogs Track *investigating: (a) small microblogs multilingual information retrieval in Arabic, English, French and Latin languages; (b) microblogs bilingual information retrieval for tuning systems running on language pairs; (c) microblog monolingual information retrieval based on 2017 language identification.*

Topics were extracted from the French VodKaster website that allows readers to get personal short comments (or “microcritics”) about movies. The challenge of the task was to find related microblogs in four different languages in a large archive.

The *argumentation mining* task (Cossu et al. 2018) aimed to automatically identify reason-conclusion structures from text, which can model the position, stance or attitude (as expressed via Twitter microblogs) of a social web user about a cultural event.

Mining Opinion Argumentation Track *investigating: (a) polarity detection in microblogs; (b) automatic identification of argumentation elements over microblogs and Wikipedia; (c) classification and summarization of arguments in texts.*

The idea was to perform a search process on a massive microblog collection that focuses on claims about a given festival.

In addition, there was a new pilot task on dialect or language variation detection using a new corpus, and extending the earlier 2017 language recognition task.

Dialectal Focus Retrieval Track *investigating: (a) Arabic dialects in blogs, microblogs and video news transcriptions; (b) Spanish language variations in blogs, microblog and journals.*

5.4 To Be Continued?

At CLEF 2018 in Avignon, there was extensive discussion on the future of the *Cultural Microblog Contextualization (MC2)* lab as no new lab or track proposal was submitted to CLEF 2019. The track addressed an important area of research, and

managed to attract a strong base of participants with ongoing research and interest in this area, and plans for follow-up activities within CLEF or elsewhere are ongoing. The main result of the *Microblog Contextualization (MC2)* tracks, however, is a great number of unique test collections that can be used for future experiments.

6 Discussion and Conclusions

This completes our rundown of the *INitiative for the Evaluation of XML Retrieval (INEX)* lab of CLEF 2012–2014, and follow up CLEF *Social Book Search (SBS)* lab (2015–2016) and CLEF *Cultural Microblog Contextualization (MC2)* lab (2016–2018). We emphasize the INEX and SBS activities, as it's too early to look back on the MC2 lab which is still in full swing at the time of writing.

As we promised in the introduction, this chapter doesn't provide a definite account of these labs, as no single chapter can do justice to the massive amount of work done in the INEX and follow up labs. Rather, we hope this chapter to be useful as a starting point with references to the respective overview papers providing full details on the labs, tracks and resulting test collections. We hope that the comprehensive high level overview helps convey the impressive breadth and scope of activities spanning many years, trying to highlight some relations between various activities inside and outside CLEF that may not be immediately apparent, thereby also highlighting some of the important developments in the field during these years.

There is one aspect in which this chapter is significantly lacking: one of the main impacts of the labs is in all the research papers it enabled and encouraged: the numerous track participation papers in the CLEF working notes, the many conference papers derived from track participation, both in the CLEF conference proceedings as in other proceedings and journals, as well as all the other papers that use some of benchmarks. As a case in point, Google Scholar lists over well over four thousand papers that mention INEX. Understanding the impact of INEX in particular, and CLEF in general, would need to take this research uptake into account.

One of INEX's key contributions to IR is that it has been completely volunteer run since 2002, with all organization and activities crowdsourced to the participants: from the start (proposal of tracks and tasks) until the end (topic creation, topic assessment). This created a generation of researchers that were touched by INEX, and all contributed to its great success. Just to mention a selection of key organizers involved in 2012–2018: *P. Bellot, T. Bogers, T. Chappell, J. Cossu, A. Doucet, L. Ermakova, S. Geva, L. Goeuriot, S. Gurajada, M. Gäde, M. Hajjem, M. A. Hall, I. Hendrickx, H. C. Huurdeman, J. Kamps, G. Kazai, M. Koolen, M. Landoni, M. Marx, A. Mishra, V. Moriceau, J. Mothe, P. Mulhem, J.-N. Nie, M. Preminger, G. Ramírez, E. SanJuan, M. Sanderson, E. Sanjuan, R. Schenkel, F. Scholer, A. Schuh, M. Skov, X. Tannier, M. Theobald, E. Toms, M. Trappett, A. Trotman, S. Verberne, D. Walsh, and Q. Wang.* But there were countless more participant volunteers

helping out with the tracks, and becoming part of the extended INEX family. It is through this generation of researchers, and their follow-up students, that INEX is making a lasting impact on the field for many years to come.

Acknowledgements Thanks to the CLEF Association, CLEF Conference, and CLEF Labs organizers for the wonderful support over the years, that greatly facilitated all the work reported in this chapter. Special thanks to the editors of this volume for support, encouragement and great flexibility.

References

- Alonso O, Hearst MA, Kamps J (2015a) Report on the first SIGIR workshop on graph search and beyond (GSB'15). SIGIR Forum 49(2):89–97. <http://doi.acm.org/10.1145/2888422.2888436>
- Alonso O, Kamps J, Karlgren J (2015b) Report on the seventh workshop on exploiting semantic annotations in information retrieval (ESAIR'14). SIGIR Forum 49(1):27–34. <http://doi.acm.org/10.1145/2795403.2795412>
- Balog K, Dalton J, Doucet A, Ibrahim Y (2016) Report on the eighth workshop on exploiting semantic annotations in information retrieval (ESAIR '15). SIGIR Forum 50(1):49–57. <http://doi.acm.org/10.1145/2964797.2964806>
- Beckers T, Bellot P, Demartini G, Denoyer L, Vries CMD, Doucet A, Fachry KN, Fuhr N, Gallinari P, Geva S, Huang WC, Iofciu T, Kamps J, Kazai G, Koolen M, Kutty S, Landoni M, Lehtonen M, Moriceau V, Nayak R, Nordlie R, Pharo N, SanJuan E, Schenkel R, Tannier X, Theobald M, Thom JA, Trotman A, de Vries AP (2010) Report on INEX 2009. SIGIR Forum 44(1):38–57. <http://doi.acm.org/10.1145/1842890.1842897>
- Belkin NJ, Clarke CLA, Gao N, Kamps J, Karlgren J (2011) Report on the SIGIR workshop on “entertain me”: supporting complex search tasks. SIGIR Forum 45(2):51–59. <http://doi.acm.org/10.1145/2093346.2093354>
- Bellot P, Chappell T, Doucet A, Geva S, Gurajada S, Kamps J, Kazai G, Koolen M, Landoni M, Marx M, Mishra A, Moriceau V, Mothe J, Preminger M, Ramírez G, Sanderson M, Sanjuan E, Scholer F, Schuh A, Tannier X, Theobald M, Trappett M, Trotman A, Wang Q (2012a) Report on INEX 2012. SIGIR Forum 46(2):50–59. <http://doi.acm.org/10.1145/2422256.2422264>
- Bellot P, Chappell T, Doucet A, Geva S, Kamps J, Kazai G, Koolen M, Landoni M, Marx M, Moriceau V, Mothe J, Ramírez G, Sanderson M, SanJuan E, Scholer F, Tannier X, Theobald M, Trappett M, Trotman A, Wang Q (2012b) Report on INEX 2011. SIGIR Forum 46(1):33–42. <http://doi.acm.org/10.1145/2215676.2215679>
- 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 (2013a) Overview of INEX 2013. In: 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, pp 269–281
- Bellot P, Moriceau V, Mothe J, SanJuan E, Tannier X (2013b) Overview of INEX tweet contextualization 2013 track. In: Forner P, Navigli R, Tufis D, Ferro N (eds) Working notes for CLEF 2013 conference, Valencia, Spain, September 23–26, 2013. CEUR-WS.org. CEUR workshop proceedings, vol 1179. <http://ceur-ws.org/Vol-1179/CLEF2013wn-INEX-BellotEt2013.pdf>

- Bellot P, Bogers T, Geva S, Hall MA, Huurdeman HC, Kamps J, Kazai G, Koolen M, Moriceau V, Mothe J, Preminger M, SanJuan E, Schenkel R, Skov M, Tannier X, Walsh D (2014a) Overview of INEX 2014. In: 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, pp 212–228
- Bellot P, Moriceau V, Mothe J, SanJuan E, Tannier X (2014b) Overview of INEX tweet contextualization 2014 track. In: Cappellato L, Ferro N, Halvey M, Kraaij W (eds) Working notes for CLEF 2014 conference, Sheffield, UK, September 15–18, 2014. CEUR-WS.org. CEUR Workshop proceedings, vol 1180, pp 494–500. <http://ceur-ws.org/Vol-1180/CLEF2014wn-Inex-BellotEt2014.pdf>
- Bennett PN, Gabrilovich E, Kamps J, Karlgren J (2014) Report on the sixth workshop on exploiting semantic annotations in information retrieval (ESAIR'13). SIGIR Forum 48(1):13–20. <http://doi.acm.org/10.1145/2641383.2641387>
- Bogers T, Koolen M (2015) Second workshop on new trends in content-based recommender systems (cbrecsys 2015). In: Werthner H, Zanker M, Golbeck J, Semeraro G (eds) Proceedings of the 9th ACM conference on recommender systems, RecSys 2015, Vienna, Austria, September 16–20, 2015. ACM, New York, pp 339–340. <http://doi.acm.org/10.1145/2792838.2798718>
- Bogers T, Koolen M, Cantador I (2014) Workshop on new trends in content-based recommender systems: (cbrecsys 2014). In: Kobsa A, Zhou MX, Ester M, Koren Y (eds) Eighth ACM conference on recommender systems, RecSys '14, Foster City, Silicon Valley, CA, USA - October 06–10, 2014. ACM, New York, pp 379–380. <http://doi.acm.org/10.1145/2645710.2645784>
- Bogers T, Hendrickx I, Koolen M, Verberne S (2016a) Overview of the SBS 2016 mining track. In: Balog K, Cappellato L, Ferro N, Macdonald C (eds) Working notes of CLEF 2016 - conference and labs of the evaluation forum, Évora, Portugal, 5–8 September, 2016. CEUR-WS.org, CEUR workshop proceedings, vol 1609, pp 1053–1063. <http://ceur-ws.org/Vol-1609/16091053.pdf>
- Bogers T, Koolen M, Musto C, Lops P, Semeraro G (2016b) Third workshop on new trends in content-based recommender systems (cbrecsys 2016). In: Sen S, Geyer W, Freyne J, Castells P (eds) Proceedings of the 10th ACM conference on recommender systems, Boston, MA, USA, September 15–19, 2016. ACM, New York, pp 419–420. <http://doi.acm.org/10.1145/2959100.2959200>
- Bogers T, Gäde M, Hall MM, Freund L, Koolen M, Petras V, Skov M (2018) Report on the workshop on barriers to interactive IR resources re-use (BIIRRR 2018). SIGIR Forum 52(1):119–128. <https://doi.org/10.1145/3274784.3274795>
- Bogers T, Dodson S, Freund L, Gäde M, Hall MM, Koolen M, Petras V, Pharo N, Skov M (2019) Workshop on barriers to interactive IR resources re-use (BIIRRR 2019). In: Azzopardi L, Halvey M, Ruthven I, Joho H, Murdock V, Qvarfordt P (eds) Proceedings of the 2019 conference on human information interaction and retrieval, CHIIR 2019, Glasgow, Scotland, UK, March 10–14, 2019. ACM, New York, pp 389–392. <https://doi.org/10.1145/3295750.3298965>
- Chappell T, Geva S (2012) Overview of the INEX 2012 relevance feedback track. In: 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/>
- Cossu J, Gonzalo J, Hajjem M, Hamon O, Latiri C, SanJuan E (2018) CLEF MC2 2018 lab technical overview of cross language microblog search and argumentative mining. In: Cappellato L, Ferro N, Nie J, Soulier L (eds) Working notes of CLEF 2018 - conference and labs of the evaluation forum, Avignon, France, September 10–14, 2018. CEUR-WS.org. CEUR workshop proceedings, vol 2125, http://ceur-ws.org/Vol-2125/invited_paper_7.pdf
- Dean-Hall A, Clarke CLA, Kamps J, Thomas P, Voorhees EM (2012) Overview of the TREC 2012 contextual suggestion track. In: Voorhees EM, Buckland LP (eds) Proceedings of the twenty-first text retrieval conference, TREC 2012, Gaithersburg, Maryland, USA, November 6–9, 2012. National Institute of Standards and Technology (NIST), Gaithersburg, vol Special Publication 500-298, <http://trec.nist.gov/pubs/trec21/papers/CONTEXTUAL12.overview.pdf>

- Dean-Hall A, Clarke CLA, Simone N, Kamps J, Thomas P, Voorhees EM (2013) Overview of the TREC 2013 contextual suggestion track. In: Voorhees EM (ed) Proceedings of the twenty-second text retrieval conference, TREC 2013, Gaithersburg, Maryland, USA, November 19–22, 2013. National Institute of Standards and Technology (NIST), Gaithersburg, vol Special Publication 500-302 <http://trec.nist.gov/pubs/trec22/papers/CONTEXT.OVERVIEW.pdf>
- Dean-Hall A, Clarke CLA, Kamps J, Thomas P, Voorhees EM (2014) Overview of the TREC 2014 contextual suggestion track. In: Voorhees EM, Ellis A (eds) Proceedings of the twenty-third text retrieval conference, TREC 2014, Gaithersburg, Maryland, USA, November 19–21, 2014. National Institute of Standards and Technology (NIST), Gaithersburg, vol Special Publication 500-308, <http://trec.nist.gov/pubs/trec23/papers/overview-context.pdf>
- Dean-Hall A, Clarke CLA, Kamps J, Kiseleva J, Voorhees EM (2015) Overview of the TREC 2015 contextual suggestion track. In: Voorhees EM, Ellis A (eds) Proceedings of the twenty-fourth text retrieval conference, TREC 2015, Gaithersburg, Maryland, USA, November 17–20, 2015. National Institute of Standards and Technology (NIST), Gaithersburg, vol Special Publication 500-319 <http://trec.nist.gov/pubs/trec24/papers/Overview-CX.pdf>
- Doucet A, Kazai G, Colutto S, Mühlberger G (2013) ICDAR 2013 competition on book structure extraction. In: 12th international conference on document analysis and recognition, ICDAR 2013, Washington, DC, USA, August 25–28, 2013, IEEE Computer Society, Washington, pp 1438–1443. <https://doi.org/10.1109/ICDAR.2013.290>
- Ermakova L, Goeriot L, Mothe J, Mulhem P, Nie J, SanJuan E (2016) Cultural micro-blog contextualization 2016 workshop overview: data and pilot tasks. In: Balog K, Cappellato L, Ferro N, Macdonald C (eds) Working notes of CLEF 2016 - conference and labs of the evaluation forum, Évora, Portugal, 5–8 September, 2016. CEUR-WS.org. CEUR workshop proceedings, vol 1609, pp 1197–1200. <http://ceur-ws.org/Vol-1609/16091197.pdf>
- Ermakova L, Goeriot L, Mothe J, Mulhem P, Nie J, SanJuan E (2017a) CLEF 2017 microblog cultural contextualization lab overview. In: Jones GJF, Lawless S, Gonzalo J, Kelly L, Goeriot L, Mandl T, Cappellato L, Ferro N (eds) Experimental IR meets multilinguality, multimodality, and interaction - 8th international conference of the CLEF association, CLEF 2017, Dublin, Ireland, September 11–14, 2017, Proceedings. Lecture Notes in Computer Science, vol 10456, pp 304–314. Springer, Basel. https://doi.org/10.1007/978-3-319-65813-1_27
- Ermakova L, Mothe J, SanJuan E (2017b) CLEF 2017 microblog cultural contextualization content analysis task overview. In: Cappellato L, Ferro N, Goeriot L, Mandl T (eds) Working notes of CLEF 2017 - conference and labs of the evaluation forum, Dublin, Ireland, September 11–14, 2017. CEUR-WS.org, CEUR workshop proceedings, vol 1866. http://ceur-ws.org/Vol-1866/invited_paper_14.pdf
- Gäde M, Hall MM, Huurdeman HC, Kamps J, Koolen M, Skov M, Toms E, Walsh D (2015a) Overview of the SBS 2015 interactive track. In: Cappellato L, Ferro N, Jones GJF, SanJuan E (eds) Working notes of CLEF 2015 - conference and labs of the evaluation forum, Toulouse, France, September 8–11, 2015. CEUR-WS.org, CEUR workshop proceedings, vol 1391. <http://ceur-ws.org/Vol-1391/78-CR.pdf>
- Gäde M, Hall MM, Huurdeman HC, Kamps J, Koolen M, Skov M, Toms E, Walsh D (2015b) Report on the first workshop on supporting complex search tasks. SIGIR Forum 49(1):50–56. <http://doi.acm.org/10.1145/2795403.2795415>
- Gäde M, Hall MM, Huurdeman HC, Kamps J, Koolen M, Skov M, Bogers T, Walsh D (2016) Overview of the SBS 2016 interactive track. In: Balog K, Cappellato L, Ferro N, Macdonald C (eds) Working notes of CLEF 2016 - conference and labs of the evaluation forum, Évora, Portugal, 5–8 September, 2016. CEUR-WS.org. CEUR workshop proceedings, vol 1609, pp 1024–1038. <http://ceur-ws.org/Vol-1609/16091024.pdf>
- Goeriot L, Mothe J, Mulhem P, Murtagh F, SanJuan E (2016) Overview of the CLEF 2016 cultural micro-blog contextualization workshop. In: 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, pp 371–378

- Goeriot L, Mulhem P, SanJuan E (2017) CLEF 2017 MC2 search and time line tasks overview. In: Cappellato L, Ferro N, Goeriot L, Mandl T (eds) Working notes of CLEF 2017 - conference and labs of the evaluation forum, Dublin, Ireland, September 11–14, 2017. CEUR-WS.org, CEUR workshop proceedings, vol 1866. http://ceur-ws.org/Vol-1866/invited_paper_15.pdf
- Gurajada S, Kamps J, Mishra A, Schenkel R, Theobald M, Wang Q (2013) Overview of the INEX 2013 linked data track. In: Forner P, Navigli R, Tufis D, Ferro N (eds) Working notes for CLEF 2013 conference, Valencia, Spain, September 23–26, 2013. CEUR-WS.org, CEUR workshop proceedings, vol 1179. <http://ceur-ws.org/Vol-1179/CLEF2013wn-INEX-GurajadaEt2013.pdf>
- Hajjem M, Cossu J, Latiri C, SanJuan E (2018) CLEF MC2 2018 lab overview. In: Bellot P, Trabelsi C, Mothe J, Murtagh F, Nie J, Soulier L, SanJuan E, Cappellato L, Ferro N (eds) Experimental IR meets multilinguality, multimodality, and interaction - 9th international conference of the CLEF association, CLEF 2018, Avignon, France, September 10–14, 2018, Proceedings. Lecture Notes in Computer Science, vol 11018, pp 302–308. Springer. https://doi.org/10.1007/978-3-319-98932-7_27
- Hall MM, Huurdeman HC, Koolen M, Skov M, Walsh D (2014) Overview of the INEX 2014 interactive social book search track. In: Cappellato L, Ferro N, Halvey M, Kraaij W (eds) Working notes for CLEF 2014 conference, Sheffield, UK, September 15–18, 2014. CEUR-WS.org, CEUR workshop proceedings, vol 1180, pp 480–493. <http://ceur-ws.org/Vol-1180/CLEF2014wn-Inex-HallEt2014.pdf>
- Hashemi SH, Clarke CLA, Kamps J, Kiseleva J, Voorhees EM (2016) Overview of the TREC 2016 contextual suggestion track. In: Voorhees EM, Ellis A (eds) Proceedings of the twenty-fifth text retrieval conference, TREC 2016, Gaithersburg, Maryland, USA, November 15–18, 2016, National Institute of Standards and Technology (NIST), Gaithersburg, vol Special Publication 500-321. <http://trec.nist.gov/pubs/trec25/papers/Uamsterdam-CX.pdf>
- Kamps J, Geva S, Peters C, Sakai T, Trotman A, Voorhees EM (2009) Report on the SIGIR 2009 workshop on the future of IR evaluation. SIGIR Forum 43(2):13–23. <http://doi.acm.org/10.1145/1670564.1670567>
- Kamps J, Karlgren J, Mika P, Murdock V (2013) Report on the fifth workshop on exploiting semantic annotations in information retrieval (ESAIR'12). SIGIR Forum 47(1):38–45. <http://doi.acm.org/10.1145/2492189.2492196>
- Kazai G (2009, 2018) INEX. In: Liu L, Özsu MT (eds) Encyclopedia of database systems. Springer, New York, p 1472. https://doi.org/10.1007/978-0-387-39940-9_2846 https://doi.org/10.1007/978-0-387-39940-9_2846
- Koolen M, Kazai G, Kamps J, Preminger M, Doucet A, Landoni M (2012) Overview of the INEX 2012 social book search track. In: 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/>
- Koolen M, Kazai G, Preminger M, Doucet A (2013) Overview of the INEX 2013 social book search track. In: Forner P, Navigli R, Tufis D, Ferro N (eds) Working notes for CLEF 2013 conference, Valencia, Spain, September 23–26, 2013. CEUR-WS.org, CEUR workshop proceedings, vol 1179. <http://ceur-ws.org/Vol-1179/CLEF2013wn-INEX-KoolenEt2013b.pdf>
- Koolen M, Bogers T, Kamps J, Kazai G, Preminger M (2014) Overview of the INEX 2014 social book search track. In: Cappellato L, Ferro N, Halvey M, Kraaij W (eds) Working notes for CLEF 2014 conference, Sheffield, UK, September 15–18, 2014. CEUR-WS.org, CEUR workshop proceedings, vol 1180, pp 462–479. <http://ceur-ws.org/Vol-1180/CLEF2014wn-Inex-KoolenEt2014.pdf>
- Koolen M, Bogers T, Gäde M, Hall MA, Huurdeman HC, Kamps J, Skov M, Toms E, Walsh D (2015a) Overview of the CLEF 2015 social book search lab. In: Mothe J, Savoy J, Kamps J, Pinel-Sauvagnat K, Jones GJF, 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, pp 545–564

- Koolen M, Bogers T, Kamps J (2015b) Overview of the SBS 2015 suggestion track. In: Cappellato L, Ferro N, Jones GJF, SanJuan E (eds) Working notes of CLEF 2015 - conference and labs of the evaluation forum, Toulouse, France, September 8–11, 2015. CEUR-WS.org. CEUR workshop proceedings, vol 1391. <http://ceur-ws.org/Vol-1391/76-CR.pdf>
- Koolen M, Bogers T, Gäde M, Hall M, Hendrickx I, Huurdeman HC, Kamps J, Skov M, Verberne S, Walsh D (2016a) Overview of the CLEF 2016 social book search lab. In: 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, pp 351–370
- Koolen M, Bogers T, Kamps J (2016b) Overview of the SBS 2016 suggestion track. In: Balog K, Cappellato L, Ferro N, Macdonald C (eds) Working notes of CLEF 2016 - conference and labs of the evaluation forum, Évora, Portugal, 5–8 September, 2016. CEUR-WS.org, CEUR workshop proceedings, vol 1609, pp 1039–1052. <http://ceur-ws.org/Vol-1609/16091039.pdf>
- Koolen M, Kamps J, Bogers T, Belkin NJ, Kelly D, Yilmaz E (2017) Report on the second workshop on supporting complex search tasks. SIGIR Forum 51(1):58–66. <http://doi.acm.org/10.1145/3130332.3130343>
- Nordlie R, Pharo N (2012) Seven years of INEX interactive retrieval experiments - lessons and challenges. In: 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, pp 13–23
- Petras V, Ferro N, Gäde M, Isaac A, Kleineberg M, Masiero I, Nicchio M, Stiller J (2012) Cultural heritage in CLEF (CHiC) overview 2012. In: 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/>
- Petras V, Bogers T, Hall M, Savoy J, Malak P, Pawlowski A, Ferro N, Masiero I (2013) Cultural heritage in CLEF (CHiC) 2013. In: 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, pp 192–211
- SanJuan E, Moriceau V, Tannier X, Bellot P, Mothe J (2012) Overview of the INEX 2012 tweet contextualization track. In: 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/>
- Thanos C, Casarosa V (2017) The key role of the DELOS network of excellence in establishing digital libraries as a research field in Europe. LIBER Quart 26(4):296–307. <http://doi.org/10.18352/lq.10165>
- Trappett M, Geva S, Trotman A, Scholer F, Sanderson M (2012) Overview of the INEX 2012 snippet retrieval track. In: 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/>
- Trappett M, Geva S, Trotman A, Scholer F, Sanderson M (2013) Overview of the INEX 2013 snippet retrieval track. In: Forner P, Navigli R, Tufis D, Ferro N (eds) Working notes for CLEF 2013 conference, Valencia, Spain, September 23–26, 2013. CEUR-WS.org, CEUR workshop proceedings, vol 1179. <http://ceur-ws.org/Vol-1179/CLEF2013wn-INEX-TrappettEt2013.pdf>
- Wang Q, Kamps J, Ramírez Camps G, Marx M, Schuth A, Theobald M, Gurajada S, Mishra A (2012) Overview of the INEX 2012 linked data track. In: 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/>