1 Introduction

The 2021 edition of the *Conference and Labs of the Evaluation Forum* (CLEF) was hosted by the University "Politehnica" of Bucharest, Romania from 21st to 24th September 2021. Due to the continuing uncertain conditions due to the Covid-19 pandemic, the conference was fully virtual. The conference format remained the same as in past years, and consisted of keynotes, contributed papers, lab sessions, and poster sessions, including reports from other benchmarking initiatives from around the world. All sessions were organized and run online.

CLEF was established in 2000 as a spin-off of the TREC Cross-Language Track with a focus on stimulating research and innovation in multimodal and multilingual information access and retrieval [Ferro, 2019; Ferro and Peters, 2019]. Over the years, CLEF has fostered the creation of language resources in many European and non-European languages, promoted the growth of a vibrant and multidisciplinary research community, provided sizable improvements in the performance of monolingual, bilingual, and multilingual information access systems [Ferro and Silvello, 2017], and achieved a substantial scholarly impact [Larsen, 2019; Tsikrika et al., 2011, 2013] achieving, according to Google Scholar Metrics 2021, h5-index 41 and h5-median 57 and being among the top-20 venues in the "Databases & Information Systems" category.

In its first 10 years, CLEF hosted a series of experimental labs that reported their results at an annual workshop held in conjunction with the European Conference on Digital Libraries (ECDL). In 2010, now a mature and well-respected evaluation forum, CLEF expanded to include a complementary peer-reviewed conference for discussion of advancing evaluation methodologies and reporting the evaluation of information access and retrieval systems regardless of data type, format, language, etc. Moreover, the scope of the evaluation labs was broadened, to comprise not only multilinguality but also multimodality in information access. Multimodality here is intended not only as the ability to deal with information coming in multiple media but also in different modalities, e.g. the Web, social media, news streams, specific domains and so on. Since 2010, the CLEF conference has established a format with keynotes, contributed papers, lab sessions, and poster sessions, including reports from other benchmarking initiatives from around the world. Since 2013, CLEF has been supported by an association, a lightweight not-for-profit legal entity that thanks to the financial support of the CLEF community takes care of the small central coordination needed to operate CLEF on an ongoing basis and makes it a self-sustaining activity [Ferro, 2019].

CLEF 2021 continued the initiative introduced in the 2019 edition during which, the *European Conference for Information Retrieval (ECIR)* and CLEF joined forces: ECIR 2021 hosted a special session dedicated to CLEF Labs where lab organizers present the major outcomes of their Labs and their plans for ongoing activities, followed by a poster session to favour discussion during the conference. This was reflected in the ECIR 2020 proceedings, where CLEF Lab activities and results were reported as short papers. The goal was not only to engage the ECIR community in CLEF activities but also to disseminate the research results achieved during CLEF evaluation cycles as submission of papers to ECIR.

CLEF 2021 ran as an online, free of charge event, thanks to the support of the AI4Media H2020 project[1], AI Multimedia Lab, Politehnica University of Bucharest[2], and sponsorship from

---

[1] https://www.ai4media.eu/
[2] https://www.aimultimedialab.ro/

Facebook AI[3], Keysight Technologies Romania[4], and Siemens Romania R&D[5]. This gave the opportunity to researchers around the globe for remote participation. In total, 618 individuals registered to attend the conference, with approximately 14% coming from Asia, 20% coming from the Americas, 66% from Europe and 7 participants from Africa and 16 from Oceania. The online program was run using Zoom Webinar for the plenary sessions, and Zoom Meetings for the Lab sessions. The number of attendees per plenary was approximately 80 individuals, while the number of attendees for the lab sessions varied, in average being situated between 45-50 participants. The organizers scheduled all Zoom sessions ahead of time, assigning different coordinators to the different sessions. Unfortunately, it was not possible to organize social activities (random encounters, social events, etc.), mainly due to limited solutions available. Several options however have sprung during the recent months that future online editions of conferences could adopt.

2   The CLEF Conference

CLEF 2021 continued the focus of the CLEF conference on "experimental IR", as carried out at evaluation forums (CLEF Labs, TREC, NTCIR, FIRE, MediaEval, RomIP, TAC, etc.), with special attention to the challenges of multimodality, multilinguality, and interactive search. We invited submissions on significant new insights demonstrated on the resulting IR test collections, on analysis of IR test collections and evaluation measures, as well as on concrete proposals to push the boundaries of the Cranfield/TREC/CLEF paradigm [Candan et al., 2021].

Keynotes   The following scholars were invited to give a keynote talk at the CLEF 2021 conference.

Naila Murray (Facebook AI Research, USA) delivered a talk entitled "Unsupervised Meta-Domain Adaptation for Fashion Retrieval" which focused on cross-domain fashion item retrieval where unconstrained consumer images are used to query for fashion items in a collection of high-quality photographs provided by retailers. To perform this task, approaches typically leverage both consumer and shop domains from a given dataset to learn a domain-invariant representation, allowing these images of different nature to be directly compared. When consumer images are not available beforehand, such training is impossible. In the talk, Naila described a recent approach to this challenging and yet practical scenario, which leverages representations learned for cross-domain retrieval from another source dataset and to adapts them to the target dataset for this particular setting.

Mark Sanderson (RMIT University, Australia) gave a speech on "Creating a Conversational Search System", discussing how, in order to achieve a system that works seamlessly, we need research that covers a wide range of disciplines. Mark described some of the work that his PhD students conducted in the last few years to examine different aspects of conversational search including some of the distinct human and interface elements that one needs to consider when building a fully operational conversational search system.

---

[3]https://ai.facebook.com/
[4]https://www.keysight.com/
[5]https://new.siemens.com/ro/

Technical Program   CLEF 2021 received a total of 21 scientific submissions, of which a total of 11 papers (10 long, one short) were accepted. Each submission was reviewed by three program committee members, and the program chairs oversaw the reviewing and follow-up discussions. Twelve countries are represented in the accepted papers, as several of them were a product of international collaboration. This year, researchers addressed the following important challenges in the community: application of neural methods for entity recognition as well as misinformation detection in the health area, skills extraction in job-match databases, stock market prediction using financial news and extraction of audio features for podcast retrieval. Evaluation remains a strong interest with papers on evaluation of 1) pseudo-relevance feedback based on web-based data enrichment, 2) evolving datasets using pivot systems, and 3) multitask learning models for relevance assessment. Creating of shareable open datasets is also a strong focus this year with datasets created for 1) linguistic uncertainty in NLP, 2) an Italian corpus for subjectivity detection in newspapers, and 3) a framework for creating datasets for personalized type-based facet ranking tasks.

Like in previous editions since 2015, CLEF 2021 continued inviting CLEF lab organizers to nominate a "best of the labs" paper that was reviewed as a full paper submission to the CLEF 2021 conference according to the same review criteria and PC. Five full papers were accepted for this section.

3   The CLEF Lab Sessions

Fifteen lab proposals were received and evaluated in peer review based on their innovation potential and the quality of the resources created. To identify the best proposals, well-established criteria from previous editions of CLEF were applied like, for example, topical relevance, novelty, potential impact on future world affairs, likely number of participants, and the quality of the organizing consortium. This year we further stressed the connection to real-life usage scenarios and we tried to avoid overlaps among labs as much as possible in order to promote synergies and integration.

The 12 selected labs represented scientific challenges based on new data sets and real world problems in multimodal and multilingual information access. These data sets provide unique opportunities for scientists to explore collections, develop solutions for these problems, receive feedback on the performance of their solutions, and discuss the issues with peers at the workshops.

The 12 labs running as part of CLEF 2021 comprised mainly labs that continued from previous editions at CLEF (ARQMath, BioASQ, CheckThat!, CheMU, CLEF eHealth, eRisk, ImageCLEF, LifeCLEF, Lilas, PAN, and Touché) and a new pilot/workshop activity (SimpleText). Details of the individual labs are described by the lab organizers in the CLEF Working Notes [Faggioli et al., 2021]. We only provide a brief overview of them here.

ARQMath: Answer Retrieval for Mathematical Questions[6] [Mansouri et al., 2021] considers the problem of finding answers to new mathematical questions among posted answers on the community question answering site *Math Stack Exchange*. The goals of the lab are to develop methods for mathematical information retrieval based on both text and formula analysis. Objectives to reach these goals include creating test collections for training and

---

[6]https://www.cs.rit.edu/~dprl/ARQMath

evaluating Math IR systems, establishing a state-of-the-art set of retrieval solutions on these test collection to be used as future baselines, and to promote Math IR to the research community. Compared to the 2020 ARQMath edition, this year the test collection size doubled, the same being observed in the number of participants.

BioASQ[7] [Nentidis et al., 2021] challenges researchers with large-scale biomedical semantic indexing and question answering (QA). The challenges include tasks relevant to hierarchical text classification, machine learning, information retrieval, QA from texts and structured data, multi-document summarization and many other areas. The aim of the BioASQ workshop is to push the research frontier towards systems that use the diverse and voluminous information available online to respond directly to the information needs of biomedical scientists. Four tasks were organized in 2021, two on biomedical semantic indexing, QA, and a new task on COVID-19 QA. For the Spanish semantic indexing task, a continuation of a task introduced in previous years, new data was added that contained Spanish clinical trials and Spanish patents. Participant systems generally outperformed strong baselines, with participant solution clearly shifting towards the use of deep neural approaches. As such approaches, however, require large amounts of training data that were not available for BioASQ tasks, participants were investigating knowledge and model transfer from other resources.

CheckThat!: Detecting Check-Worthy Claims, Previously Fact-Checked Claims, and Fake News[8] [Nakov et al., 2021] aims to foster the development of technologies capable of both spotting and verifying check-worthy claims in short messages and political debates in various languages. This year there were three main shared tasks where participants were to estimate the check-worthiness of a claim in a short message (tweet) and in political texts (debates/speeches), to check if a detected claim was previously verified, retrieve evidence to fact-check a claim, and verify the factuality of a claim. The data in 2021 included more languages than in the previous year, with some teams addressing the challenges for all languages, while others tackling one language, only.

ChEMU: Cheminformatics Elsevier Melbourne University[9] [Li et al., 2021] proposes two key information extraction tasks over chemical reactions from patent texts. The ChEMU corpus builds on the one used in the previous lab edition, being extended to provide data for two distinct 2021 tasks: reference resolution for chemical reactions, and anaphora resolution to identify relationships (i.e. coreference and bridging relationships) between expressions in descriptions of chemical reactions. Out of 19 originally registered teams, only two manged to submit experiments. The tasks proved to be complex, the submitted experiments barely over-performing the baseline results.

CLEF eHealth[10] [Suominen et al., 2021] aims to support the development of techniques to aid laypeople, clinicians and policy-makers in easily retrieving and making sense of medical content to support their decision making. The goals of the lab are to develop processing

---

[7]http://www.bioasq.org/workshop2021
[8]https://sites.google.com/view/clef2021-checkthat
[9]http://chemu2021.eng.unimelb.edu.au/
[10]https://clefehealth.imag.fr/

methods and resources in a multilingual setting to enrich difficult-to-understand eHealth texts and provide valuable documentation. Organized since 2012, the CLEF eHealth labs have provided a recurring contribution to the creation and dissemination of text analytics resources, methods, test collection, and evaluation benchmarks that support both medical professionals and laypersons when dealing with health-related information. The 2021 CLEF eHealth edition organized two tasks. The first one is a multilingual Information Extraction task, focusing on Spanish language ultrasound reports. The second task, Consumer Health Search, is a continuation of previous CLEF eHealth IR tasks with a new representative web corpus and layperson medical queries. From the 67 teams that originally registered, 11 submitted runs to the two tasks.

eRisk: Early Risk Prediction on the Internet[11] [Parapar et al., 2021] explores challenges of evaluation methodology, effectiveness metrics and other processes related to early mental health risk detection. Early detection technologies can be employed in different areas, particularly those related to health and safety. Along the years this evaluations took place, it has become evident that the interplay between psychological disorders and the users' expression through language is a very challenging task, with currently available solutions not reaching satisfactory performance levels. The 2021 edition of the lab contained three tasks, two being continuations of tasks organized in the previous years (self-harm and depression severity detection), and a new one on the topic of pathological gambling. The data provided to campaign participants consisted of texts written in social media. From the 76 teams that originally registered to this lab, 18 had submitted experiments, with a total of 117 runs (26 for Task 1, 55 for Task 2, and 36 for Task 3).

ImageCLEF: Multimedia Retrieval[12] [Ionescu et al., 2021] provides an evaluation forum for visual media analysis, indexing, classification/learning, and retrieval in medical, nature, and social media applications with a focus on multimodal data, that is data from a variety of sources and media. The 2021 ImageCLEF edition consisted of four main tasks dedicated to multimedia retrieval in four areas: medical, nature, identification of hand-drawn components, and social media, with the latter being newly introduced this year. The first task consisted of three subtasks related to radiology images (Visual Question Answering, CT-based tuberculosis evaluation, and captioning concepts across radiology images). The nature related task contained training and test data to form 3D reconstructions of coral environments. The task on hand-drawn images focused on user interface drawings as well as screenshot images that, by segmentation and labeling steps, are to provide additional support for code developers. The social media related task aimed to assess the vulnerability potential and real-life effects of users sharing personal visual data. 42 participating groups submitted over 250 experiments to these tasks, with results varying in performance improvements over previous task editions.

LifeCLEF: Multimedia Life Species Identification[13] [Joly et al., 2021] aims at boosting research on the identification and prediction of living organisms in order to solve the taxo-

---

[11] https://erisk.irlab.org/
[12] https://www.imageclef.org/2021
[13] https://www.imageclef.org/LifeCLEF2021

nomic gap and improve our knowledge of biodiversity. Through its biodiversity informatics related challenges, LifeCLEF is intended to push the boundaries of the state-of-the-art in several research directions at the frontier of multimedia information retrieval, machine learning and knowledge engineering. LifeCLEF in 2021 organized four challenges (PlantCLEF, BirdCLEF, GeoLifeCLEF, and SnakeCLEF) involving image data, audio data, and geolocations. In terms of participating teams that submitted runs, the BirdCLEF task (a bird sound recognition task) stands out with over 800 teams submitting experimental results. The main LifeCLEF outcome is that, taken together, the solutions used by the participants to solve the lab tasks provide a new snapshot of the state-of-the-art system performances in computer vision, audio analysis techniques, and machine learning algorithms that can be part of a real-world biodiversity monitoring system.

LiLAS: Living Labs for Academic Search[14] [Schaer et al., 2021] aims to bring together researchers interested in the online evaluation of academic search systems. The long term goal is to foster knowledge on improving the search for academic resources like literature, research data, and the interlinking between these resources in fields from the life sciences and the social sciences. The immediate goal of this lab is to develop ideas, best practices, and guidelines for a full online evaluation campaign at CLEF 2022. The first LiLAS iteration as a workshop-lab provided participants exclusive access to real-world academic data search systems, LIVIVO for scientific literature search and GESIS Search for data sets and open access publication search, for each of which a use case was defined. STELLA was introduced as the living lab framework to assess participant submissions which were provided either as static search results sets or as Docker images to be integrated in the live search systems. Nine experimental systems were evaluated with metrics designed for assessing interleaved results, combining results from the participants with baseline results provided by the search systems.

PAN: Digital Text Forensics and Stylometry[15] [Bevendorff et al., 2021] is a networking initiative for the digital text forensics, where researchers and practitioners study technologies that analyze texts with regard to originality, authorship, and trustworthiness. PAN provides evaluation resources consisting of large-scale corpora, performance measures and web services that allow for meaningful evaluations. The main goal is to provide for sustainable and reproducible evaluations, to get a clear view of the capabilities of state-of-the-art algorithms. This year, PAN organized three shared tasks: detecting authors of hate speech spreaders, authorship verification, and multi-author writing style analysis. Each of the tasks made use of its own specifically designed collection of documents. For the first task, focusing on profiling hate speech spreaders, a data set of social media postings (i.e. Twitter) was created, with manually annotated tweets as hater/not-hater labels. For the second task, authorship verification, the lab organizers aimed to a scaled up benchmark setting using fan-fiction literature. Finally, the multi-author style analysis task, a task that has evolved along the years PAN was organized as a lab, used a collection of Q&A postings from StackExchange where paragraphs from different answers were joined into one text with, thus, multiple authors.

---

[14]https://clef-lilas.github.io/
[15]http://pan.webis.de/

SimpleText: (Re)Telling right scientific stories to non-specialists via text simplification[16] [Ermakova et al., 2021] aims to create a community interested in generating a simplified summary of scientific documents and to contribute in making the science really open and accessible for everyone. The goal is to generate a simplified abstract of multiple scientific documents based on a given query. SimpleText was organized as a workshop that discussed three pilot tasks on text simplification for scientific information access, all contributing steps towards arriving to a simplified text summary of an input scientific text. The first pilot task addresses the passage selection challenge, i.e. which parts of a document are appropriate for inclusion into a simplified summary. The second pilot task aimed to decide which terms in a selected passage require a simplifying explanation and contextualisation. Finally, the last pilot task discussed aims to obtain simplified text passages derived from input scientific text passages. The document collection to be used for these tasks is to be compiled from preprint and open access repositories, Wikipedia, science journalism article resources.

Touché: Argument retrieval[17] [Bondarenko et al., 2021] is the first shared task on the topic of argument retrieval. Decision making processes, be it at the societal or at the personal level, eventually come to a point where one side will challenge the other with a why-question, which is a prompt to justify one's stance. Thus, technologies for argument mining and argumentation processing are maturing at a rapid pace, giving rise for the first time to argument retrieval. In its second year, Touché has organized two shared tasks: an argument retrieval for controversial questions task and an argument retrieval for comparative questions. The two tasks used different document collections: the first task provided the args.me corpus, while for the second one argument retrieval was to be done on the ClueWeb12 collection. Out of 36 registered teams, 27 have sent in their retrieval experiments, where relevance judgements from the 2020 lab edition could be used for training.

As a group, the 152 lab organizers were based in 22 countries, with Germany, and France leading the distribution. Despite CLEF's traditionally Europe-based audience, 44 (28.9%) organizers were affiliated with international institutions outside of Europe. The gender distribution was biased towards 75% male organizers.

More information on the CLEF 2021 conference, the CLEF initiative and the CLEF Association is provided on the Web:

- CLEF 2020: http://clef2021.clef-initiative.eu/
- CLEF initiative: http://www.clef-initiative.eu/
- CLEF Association: http://www.clef-initiative.eu/association

4   CLEF 2022 and Beyond

CLEF 2022 will be hosted by the University Bologna, Italy on 5-8 September 2022.
  More information on CLEF 2022, the call for papers and the ongoing labs are available at:

---

[16]https://www.irit.fr/simpleText/
[17]https://touche.webis.de/

- https://clef2022.clef-initiative.eu/

As far as labs are concerned, CLEF 2022 will run 14 evaluation activities out of 15 proposals received: 10 will be a continuation of the labs running during CLEF 2021 and 4 will be new pilot labs.

The continued activities are:

- ARQMath: Answer Retrieval for Questions on Math
  https://www.cs.rit.edu/~dprl/ARQMath;
- BioASQ: Large-scale Biomedical Semantic Indexing and Question Answering
  http://www.bioasq.org/workshop2022;
- CheckThat! Lab on Fighting the COVID-19 Infodemic and Fake News Detection
  https://sites.google.com/view/clef2022-checkthat;
- ChEMU: Cheminformatics Elsevier Melbourne University lab
  http://chemu2022.eng.unimelb.edu.au/;
- eRisk: Early Risk Prediction on the Internet
  https://erisk.irlab.org/;
- ImageCLEF: Multimedia Retrieval Challenge in CLEF
  https://www.imageclef.org/2022;
- LifeCLEF: Biodiversity Identification and Prediction Challenges
  https://www.imageclef.org/LifeCLEF2022;
- PAN: Lab on Digital Text Forensics and Stylometry
  https://pan.webis.de/.
- SimpleText: Automatic Simplification of Scientific Texts
  http://simpletext-project.com/;
- Touché: Argument Retrieval
  https://touche.webis.de/.

The new activities are:

- HIPE: Named Entity Recognition and Linking in Multilingual Historical Documents
  https://hipe-eval.github.io/HIPE-2022/;
- iDPP: Intelligent Disease Progression Prediction
  https://brainteaser.health/open-evaluation-challenges/idpp-2022/;
- JokeR: Automatic Wordplay and Humour Translation
  http://joker-project.com/;
- LeQua: Learning to Quantify
  https://lequa2022.github.io/.

CLEF 2023 will be hosted by CERTH-ITI, Greece, in early September 2023.

Finally, bids for hosting CLEF 2024 are now open and will close around July 2022. Proposals can be sent to the CLEF Steering Committee Chair at chair@clef-initiative.eu.

Acknowledgments

References

J. Bevendorff, B. Chulvi, G. L. De La Peña Sarracén, M. Kestemont, E. Manjavacas, I. Markov, M. Mayerl, M. Potthast, F. Rangel, P. Rosso, E. Stamatatos, B. Stein, M. Wiegmann, M. Wolska, and E. Zangerle. Overview of PAN 2021: Authorship Verification, Profiling Hate Speech Spreaders on Twitter, and Style Change Detection. In Candan et al. [2021], pages 419–431.

A. Bondarenko, L. Gienapp, M. Fröbe, M. Beloucif, Y. Ajjour, A. Panchenko, C. Biemann, B. Stein, J. Wachsmuth, M. Potthast, and M. Hagen. Overview of Touché 2021: Argument Retrieval. In Candan et al. [2021], pages 450–467.

K. S. Candan, B. Ionescu, L. Goeuriot, B. Larsen, H. Müller, A. Joly, M. Maistro, F. Piroi, G. Faggioli, and N. Ferro, editors. *Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Twelfth International Conference of the CLEF Association (CLEF 2021)*, 2021. Lecture Notes in Computer Science (LNCS) 12880, Springer, Heidelberg, Germany.

L. Ermakova, P. Bellot, P. Braslavski, J. Kamps, J. Mothe, D. Nurbakova, I. Ovchinnikova, and E. SanJuan. Overview of SimpleText 2021 - CLEF Workshop on Text Simplification for Scientific Information Access. In Candan et al. [2021], pages 432–449.

G. Faggioli, N. Ferro, A. Joly, M. Maistro, and F. Piroi, editors. *CLEF 2021 Working Notes*, 2021. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, http://ceur-ws.org/Vol-2936/.

---

[18]http://www.clef-initiative.eu/association
[19]https://www.ai4media.eu/
[20]https://www.aimultimedialab.ro/
[21]https://ai.facebook.com/
[22]https://www.keysight.com/
[23]https://new.siemens.com/ro/

N. Ferro. What Happened in CLEF... For a While? In F. Crestani, M. Braschler, J. Savoy, A. Rauber, H. Müller, D. E. Losada, G. Heinatz Bürki, L. Cappellato, and N. Ferro, editors, *Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Tenth International Conference of the CLEF Association (CLEF 2019)*, pages 3–45. Lecture Notes in Computer Science (LNCS) 11696, Springer, Heidelberg, Germany, 2019.

N. Ferro and C. Peters, editors. *Information Retrieval Evaluation in a Changing World – Lessons Learned from 20 Years of CLEF*, volume 41 of *The Information Retrieval Series*, 2019. Springer International Publishing, Germany.

N. Ferro and G. Silvello. 3.5K runs, 5K topics, 3M assessments and 70M measures: What trends in 10 years of Adhoc-*ish* CLEF? *Information Processing & Management*, 53(1):175–202, January 2017.

B. Ionescu, H. Müller, R. Péteri, A. B. Abacha, M. Sarrouti, D. Demner-Fushman, S. A. Hasan, S. Kozlovski, V. Liauchuk, Y. D. Cid, V. Kovalev, O. Pelka, A. García Seco de Herrera, J. Jacutprakart, C. M. Friedrich, R. Berari, A. Tauteanu, D. Fichou, P. Brie, M. Dogariu, L. D. Ştefan, M. G. Constantin, J. Chamberlain, A. Campello, A. Clark, T. A. Oliver, H. Moustahfid, A. Popescu, and J. Deshayes-Chossart. Overview of the ImageCLEF 2021: Multimedia Retrieval in Medical, Nature, Internet and Social Media Applications. In Candan et al. [2021], pages 345–370.

A. Joly, H. Goëau, S. Kahl, L. Picek, T. Lorieul, E. Cole, B. Deneu, M. Servajean, A. Durso, I. Bolon, H. Glotin, R. Planqué, R. L. Ruiz de Castañeda, W.-P. Vellinga, H. Klinck, T. Denton, I. Eggel, P. Bonnet, and H. Müller. Overview of LifeCLEF 2021: An Evaluation of Machine-Learning Based Species Identification and Species Distribution Prediction. In Candan et al. [2021], pages 371–393.

B. Larsen. The Scholarly Impact of CLEF 2010-2017. In Ferro and Peters [2019], pages 547–554.

Y. Li, B. Fang, J. He, H. Yoshikawa, S. A. Akhondi, C. Druckenbrodt, C. Thorne, Z. Afzal, Z. Zhai, T. Baldwin, and K. Verspoor. Overview of ChEMU 2021: Reaction Reference Resolution and Anaphora Resolution in Chemical Patents. In Candan et al. [2021], pages 292–307.

B. Mansouri, R. Zanibbi, D. W. Oard, and A. Agarwal. Overview of ARQMath-2 (2021): Second CLEF Lab on Answer Retrieval for Questions on Math. In Candan et al. [2021], pages 215–238.

P. Nakov, G. Da San Martino, T. Elsayed, A. Barrón-Cedeño, R. Míguez, S. Shaar, F. Alam, F. Haouari, M. Hasanain, W. Mansour, B. Hamdan, Z. S. Ali, N. Babulkov, A. Nikolov, G. K. Shahi, J. M. Struß, T. Mandl, M. Kutlu, and Y. S. Kartal. Overview of the CLEF–2021 CheckThat! Lab on Detecting Check-Worthy Claims, Previously Fact-Checked Claims, and Fake News. In Candan et al. [2021], pages 264–291.

A. Nentidis, G. Katsimpras, E. Vandorou, A. Krithara, L. Gasco, M. Krallinger, and G. Paliouras. Overview of BioASQ 2021: The Ninth BioASQ Challenge on Large-Scale Biomedical Semantic Indexing and Question Answering. In Candan et al. [2021], pages 239–263.

J. Parapar, P. Martín-Rodilla, D. E. Losada, and F. Crestani. Overview of eRisk 2021: Early Risk Prediction on the Internet. In Candan et al. [2021], pages 324–344.

P. Schaer, T. Breuer, L. J. Castro, B. Wolff, J. Schaible, and N. Tavakolpoursaleh. Overview of LiLAS 2021 – Living Labs for Academic Search. In Candan et al. [2021], pages 394–418.

H. Suominen, L. Goeuriot, L. Kelly, L. Alonso Alemany, E. Bassani, N. Brew-Sam, V. Cotik, D. Filippo, G. González-Sáez, F. Luque, P. Mulhem, G. Pasi, R. Roller, S. Seneviratne, R. Upadhyay, J. Vivaldi, M. Viviani, and C. Xu. Overview of the CLEF eHealth Evaluation Lab 2021. In Candan et al. [2021], pages 308–323.

T. Tsikrika, A. Garcia Seco de Herrera, and H. Müller. Assessing the Scholarly Impact of Image-CLEF. In P. Forner, J. Gonzalo, J. Kekäläinen, M. Lalmas, and M. de Rijke, editors, *Multilingual and Multimodal Information Access Evaluation. Proceedings of the Second International Conference of the Cross-Language Evaluation Forum (CLEF 2011)*, pages 95–106. Lecture Notes in Computer Science (LNCS) 6941, Springer, Heidelberg, Germany, 2011.

T. Tsikrika, B. Larsen, H. Müller, S. Endrullis, and E. Rahm. The Scholarly Impact of CLEF (2000–2009). In P. Forner, H. Müller, R. Paredes, P. Rosso, and B. Stein, editors, *Information Access Evaluation meets Multilinguality, Multimodality, and Visualization. Proceedings of the Fourth International Conference of the CLEF Initiative (CLEF 2013)*, pages 1–12. Lecture Notes in Computer Science (LNCS) 8138, Springer, Heidelberg, Germany, 2013.