# Report on the 1st International Workshop on Open Web Search (WOWS 2024) at ECIR 2024

Sheikh Mastura Farzana
German Aerospace Center (DLR)
Germany
sheikh.farzana@dlr.de

Maik Fröbe
Friedrich-Schiller-Universität Jena
Germany
maik.froebe@uni-jena.de

Michael Granitzer
University of Passau
Germany
michael.granitzer@uni-passau.de

Gijs Hendriksen
Radboud University
The Netherlands
gijs.hendriksen@ru.nl

Djoerd Hiemstra
Radboud University
The Netherlands
djoerd.hiemstra@ru.nl

Martin Potthast
University of Kassel, hessian.AI, and ScaDS.AI
Germany
martin.potthast@uni-kassel.de

Arjen P. de Vries
Radboud University
The Netherlands
arjen.devries@ru.nl

Saber Zerhoudi
University of Passau
Germany
saber.zerhoudi@uni-passau.de

**Abstract**

The first International Workshop on Open Web Search (WOWS) was held on Thursday, March 28th, at ECIR 2024 in Glasgow, UK. The full-day workshop had two calls for contributions: the first call aimed at scientific contributions to building, operating, and evaluating search engines cooperatively and the cooperative use of the web as a resource for researchers and innovators. The second call for implementations of retrieval components aimed to gain practical experience with joint, cooperative evaluation of search engines and their components. In total, 2 papers were accepted for the first call, and 11 software components were submitted for the second. The workshop ended with breakout sessions on how the OpenWebSearch.eu project can incorporate collaborative evaluations and a hub of search engines.

**Date:** 28 March 2024.

**Website:** https://opensearchfoundation.org/wows2024/.

**Figure 1.** Group picture of online and in-person participants of the WOWS workshop in Glasgow.

**Figure 2.** Group picture of the hackathon to develop IR components for WOWS at TU Dresden.

# 1 Introduction

Web search is a crucial technology for the digital economy dominated by a few gatekeepers. These gatekeepers control every aspect of search: they crawl the web, build the index, build the search engines, and provide the search applications. Furthermore, these players also control the advertisements next to the search results, which may incentivize them to sacrifice search result quality to show the best-paying results. The gatekeepers have every opportunity to dictate their will to their users: they even control the web browsers and operating systems that users need to access the web [Mager et al., 2023].

The result of this power imbalance is that web publishers have to optimize their content to the web search engines, instead of the search engines optimizing their results to the content. This has resulted in a closed ecosystem of search engines as well as the risk of publishers sacrificing quality. Many users of search engines are aware of this [Lewandowski and Schultheiß, 2022].

The goal of the workshop on Open Web Search [Farzana et al., 2024a] was to encourage and discuss ideas and approaches that would open up the closed search ecosystem that we encounter today. We were particularly interested in approaches that allow organizations to provide search engines cooperatively, for instance by allowing organizations to specialize on one task (crawling, indexing, searching, web search applications) while sharing their results with others. Furthermore, we were interested in ideas and approaches that allow organizations to cooperate on each individual task, in open standards for search and open source information retrieval systems.

Beyond search, the web has demonstrated its critical role as resource, like for training large language models, to analyze human behavioral data or to study the web as a structure itself. However, tapping into such a resource requires large infrastructure, corresponding technical skills as well as large efforts in terms of data cleaning and preprocessing. While a few big organizations can provide those resources, small research groups or early stage startups either lack these capabilities or need to devote large amounts of time and resources outside their core research domain. We were interested in approaches where organizations share their data processing infrastructures and treat their data as open data. Figure 1 shows a group picture at the workshop in Glasgow.

From a practical perspective, the workshop emphasized the scientific evaluation of search engines and their components using information retrieval test collections by a shared task. We encouraged participants to submit and evaluate new retrieval approaches or components of retrieval pipelines by using the shared information retrieval evaluation infrastructure TIRA/TIREx [Fröbe et al., 2023a,b]. Modern web search engines use complex pipelines with many components (e.g., query understanding, spam classification, document expansion, etc.), so different organizations may contribute different components in open search ecosystems. Hence, we asked participants to implement retrieval pipelines and/or components of retrieval pipelines in Docker images. Overall, 11 teams participated by implementing retrieval components; notably, 4 teams participated through a dedicated one-week student hackathon at TU Dresden (Figure 2). We have run all submitted components on all test collections available in TIRA/TIREx and made their results (or the intermediate results where applicable) publicly available.[1] The idea behind this setup was that, when the outputs of standard components are available on many test collections, researchers can use them without having to re-execute them, supporting GreenIR [Scells et al., 2022].

To sum up the above, the workshop aims were as follows:

1. Exchanging novel concepts, algorithms, and ideas for building web search engines cooperatively, e.g., via cooperative crawling, cooperative deployment, or cooperative evaluation of search engines, and for tapping the web as a resource for researchers and innovators; and
2. Gaining experience with joint, cooperative evaluation of search engines using TIRA/TIREx.

The workshop was organized by researchers participating in the European project OpenWebSearch.eu (still ongoing), which aims to build a fully open web index, associated infrastructures, algorithms along the whole retrieval pipeline, as well as open machine learning and knowledge representation models [Granitzer et al., 2023; Hendriksen et al., 2024].

## 2 Related Work and Events

The first call of contributions to our Open Web Search workshop, aimed at cooperative approaches to all aspects of web search, may be seen (informally) as a follow-up event of a series of workshops on open source information retrieval, that started in 2005 with the first International Workshop on Open Source Web Information Retrieval; the second International Workshop on Open Source Information Retrieval in 2006 [Yee et al., 2006]; the 2012 Workshop on Open Source Information Retrieval [Trotman et al., 2012]; and the Lucene for information access and retrieval research (LIARR) workshop 2017 [Azzopardi et al., 2017]. The workshop extended beyond the scope of these previous workshops with its specific focus on collaboration.

Our second call for evaluation components is in the line of previous workshops on the reproducibility of information retrieval research, including the workshop on Reproducibility, Inexplicability, and Generalizability of Results (RIGOR) [Arguello et al., 2016] and the Open-Source Information Retrieval Replicability Challenge (OSIRRC) [Clancy et al., 2019]. Here, too, our workshop added to these workshops by our focus on cooperative evaluation.

[1] https://zenodo.org/records/10743990

# 3 Workshop Contributions

The workshop was a full-day workshop divided into three sessions. The first session had a keynote and an overview of the collaborative evaluation efforts of the call for software submissions, the second session consisted of paper presentations, and the third session focused on breakout groups.

## 3.1 Evaluation in the Generative Era

Negar Arabzadeh delivered our keynote, focusing on the evaluation of information access systems that use generative AI methods to produce their results. The talk addressed the challenges of evaluating the performance of these systems, particularly due to the scarcity of human-labeled data, which hinders fair and accurate assessment. In particular, Arabzadeh addressed the question: can generative AI be used to do automatic relevance judgements? The keynote showed evidence that this is indeed the case.

Arabzadeh proposed leveraging the Fréchet Distance to measure the distance between the distributions of relevant judged items and retrieved results, taking inspiration from the success of using Fréchet Inception Distance (FID) in assessing text-to-image generation systems.

The talk highlighted the similarity between evaluating *generated results* and assessing the quality of retrieved results in an ad hoc retrieval setting, where labels are sparse. Arabzadeh explored the use of the Fréchet Distance to quantify the quality of retrieved documents in an ad hoc retrieval system, analogous to how FID is used to evaluate the quality of generated images.

Experimental results using the MS MARCO V1 dataset and TREC Deep Learning Tracks query sets demonstrated that the Fréchet Distance is a suitable metric for evaluating IR systems, particularly in settings where few labels are available [Arabzadeh and Clarke, 2024]. Experiments with varying degree of sparseness in the availability of ground truth labels showed that the Fréchet Distance can distinguish different rankers on their performance, even when the number of queries is low. The metric was found to be robust with respect to the degree of sparsity of the ground truth and to the choice of document embedding representation.

Arabzadeh concluded by emphasizing the potential of the Fréchet Distance in evaluating IR systems in real-world settings, where obtaining comprehensive ground truth labels can be challenging and expensive. She suggested that future research could utilize the Fréchet Distance to evaluate different generative models, expanding the scope of evaluation in IR systems and allowing for the comparison of generated and retrieved results in the same context.

## 3.2 Current Topics

Two scientific articles were accepted for the first call of papers. Dinzinger et al. presented a survey of web content control standards related to generative AI [Dinzinger et al., 2024]. Content ownership and copyright have become more and more important with the rise of large AI models trained on massive amounts of data scraped from the Web. While several standards have been proposed, no one-size-fits-all approach exists yet, with only poor adoption of ad hoc standards. Workshop discussion addressed how using crawled data for training AI differs from indexing it to power a search engine. As a result, more fine-grained protection of digital assets or content might be necessary, e.g. by including metadata on the image level.

Wiegmann et al. introduced their new Mastodon collection, containing posts of over 1,0000 Mastodon instances [Wiegmann et al., 2024]. In the past years, the federated microblogging service has become more widely adopted due to people wanting to move away from centralized platforms such as Twitter/X. However, due to its federated nature, search remains a challenging open problem. We discussed what search for a federated microblogging infrastructure might look like. *Is fully federated search feasible and efficient? Can we repurpose somewhat centralized "relay" nodes for building bigger search indexes?* From our discussion, it became apparent that a shared task setting for this dataset would greatly benefit the extent to which we can develop and evaluate search in the Fediverse (the collection of federated social networks like Mastodon). However, microblog search users might have information needs that differ from those in other types of search, meaning it might still be challenging to construct good topics for this shared task; we might be better off using anonymized query logs from a larger instance.

## 3.3 Collaborative Development and Evaluation of Search Components

The second call for retrieval components attracted submissions by 11 teams. The idea was to collect components of retrieval pipelines to execute them on all collections available in TIREx and make their outputs publicly available. Retrieval components were submitted as Docker images to TIRA and were implemented against ir_datasets [MacAvaney et al., 2021]. All outputs on datasets that allow for public sharing are available on Zenodo, and we improved the TIRA python API so that it can directly load cached outputs from Zenodo in declarative PyTerrier pipelines [Macdonald et al., 2021]. We especially see a big value of this in teaching scenarios, as students in IR courses can re-use components without having to re-execute them [Fröbe et al., 2024].

Team `dossier` submitted two query intent classifiers developed with Snorkel [Ratner et al., 2017; Alexander et al., 2022]. Experiments showed that the effectiveness of retrieval models can substantially vary between different query intents, indicating that intent prediction can be an important component of retrieval pipelines [Alexander et al., 2024].

Team `fschlatt` submitted two health classification models that classify how health-related a query respectively a document is (using termhood scores [Schlatt et al., 2022]). Filtering non-health documents in the scenario of the TREC Health Misinformation 2019 task [Abualsaud et al., 2019] improved the effectiveness of retrieval models like BM25 and monoT5 [Schlatt, 2024].

Team `marcel-gohsen` submitted a query entity linking approach and an approach to generate plausible query interpretations based on entities [Kasturia et al., 2022]. Entities in the query interpretation are linked to Wikipedia articles, and experiments showed that many of the TREC topics contain entities [Gohsen and Stein, 2024].

Team `naverlabseurope` submitted SPLADE [Formal et al., 2021b,a] as a re-ranker component. For each query–document pair of some first-stage retrieval model, this SPLADE component outputs the query representation, the document representation, and the re-ranking score. The query respectively document representations are the tokens with their predicted score by SPLADE and were included in the output to allow integration of these token scores into other retrieval components.

Team `salamander` submitted a classifier that predicts if a given query has a comparative information need (e.g., `should I buy a playstation or xbox`) or not [Bondarenko et al., 2020a].

The classifier is precision-oriented and especially focused on retrieval scenarios like in the Touché shared task [Bondarenko et al., 2020b] that provides many comparative information needs.

Team `seanmacavaney` submitted two components, DocT5Query [Nogueira and Lin, 2019] and the Corpus Graph [MacAvaney et al., 2022]. Both components are recall-oriented and rather compute expensive, hence, re-using cached outputs help reduce the environmental footprint of IR experiments. However, this only holds for cases where the out-of-the-box model is used, while changes to the model would still require the experiment to be re-run.

The teams `tu-dresden-01`, `tu-dresden-02`, `tu-dresden-03`, and `tu-dresden-04` participated in the workshop during a one-week hackathon that took place at TU Dresden. For this hackathon, 12 students registered (covering both bachelor and master courses) and got ECTS credits for their work. We prepared a pool of potential project ideas from which student teams could choose, and students then had one week for preparation before the hackathon with one or two papers related to their selected topic. The projects covered (1) web page genre classification, (2) snippet extraction from long documents, (3) query expansion and variants with large language models, and (4) syntactic and semantic document features [Erben et al., 2024].

Team `QPPTK` submitted a set of 12 query performance predictors that were available in the QPPTK[2] framework [Faggioli et al., 2021], showing that the effectiveness of performance predictors varies substantially between test collections [Zendel et al., 2024]. The submission uses a PyTerrier index and an ir_datasets ID as input to predict the performance with all 12 available predictors for every query in the passed dataset, using the processing pipeline specified for the index.

Team `qspell` submitted a set of query spelling correction components and compared the effectiveness of vocabulary-based, embedding-based and LLM-based spelling correction methods [Zelch et al., 2024]. Interestingly, the embedding-based methods have a lower rate of false spelling corrections than the other methods, but are limited to changing misspellings that also appear in the embedding vocabulary. As queries in TREC-style test collections rarely have spelling errors, a discussion emerged on how one could enrich existing test collections with realistic spelling errors.

## 3.4 Breakout Discussions

The third and last session of the workshop had two parallel breakout groups. First, both breakout groups brainstormed on their disjoint topics independently, and finally, all outcomes and the main results of the brainstorming were presented in a joint session to conclude the workshop.

### 3.4.1 Collaborative Evaluation in the OpenWebSearch.EU Project

If successful, the OpenWebSearch.eu project provides a shared infrastructure to enable a diverse set of downstream search applications. The discussion group focused on the challenges and opportunities that collaborative evaluation can provide in this scenario.

The discussion started with the observation that the Open Web Index builds the basis for all derived search applications. Hence, design decisions in the index impact what downstream applications are possible. Downstream applications could provide Cranfield-style metadata (i.e., topics and relevance judgments, possibly derived from click data) and minimal retrieval pipelines to continuously evaluate the quality of the index, ensuring that potential problems are already

[2] https://github.com/Zendelo/QPP-EnhancedEval/tree/main/code/qpptk

identified early on in the pipeline. Ideally, a pull of the updated index could provide warnings in case the estimated quality of a derived search application would substantially decrease so that the search application can maybe prefer to still operate an older version of the index in production while the underlying problem can be resolved manually. The discussion continued on the different stakeholders in this scenario, focusing on how they could be motivated to collaborate, how they can trust each other, and how they could provide mutual benefit. The main stakeholders are (1) the provider of the Open Web Index, (2) operators of downstream applications, (3) users of downstream applications, and (4) researchers aiming to study/support new interesting search applications. In this scenario, operators of downstream applications and their users can provide valuable data to evaluate their search scenarios that are useful to the providers of the Open Web Index and researchers. Especially in cases where important search scenarios do not yet work sufficiently, researchers could try to find potential solutions.

The discussion continued with how this scenario could be prototyped to gain first experience. An idea was that a teaching initiative, like in the hackathon for developing retrieval components that took place at TU Dresden in preparation for the WOWS workshop (Figure 2), could be a suitable concept to test this approach. In such a teaching initiative, students could develop and evaluate search applications for use cases of their choice. Students are often very motivated to work on practical implementations, and if the integration to the Open Web Index works well and the OpenWebSearch.eu project finds long-term funding, it might even be possible that resulting search engines could be operated and updated without much effort. Especially in the light that the development of new search applications might yield interesting new problems that do not work sufficiently, such efforts might even yield new future research projects, potentially helping to reduce the gap between research and teaching [Bauer et al., 2023a,b; Fröbe et al., 2024]

### 3.4.2 A Hub for Declarative Search Engines

To make it easy for anyone to deploy a search engine, the OpenWebSearch.eu project is working on the so-called Open Web Search Engine Hub: a hub for *declarative* search engine specifications. The core idea is that one can declare a specification for their search engine (or reuse an existing one), and a set of tools would spin up the corresponding search engine. The main questions we discussed in this breakout group were: *What should such a specification look like? What components should it contain? How can we go from a specification to a running search application?*

During our discussion, it became clear that three components are crucial for a search specification: 1) the document collection or index files, 2) a choice of search engine to use as backend (e.g. Lucene, Terrier, ElasticSearch), and 3) a frontend to serve the search application (e.g. classical SERP, conversational UI). On top of that, it should be possible to define *middleware*, that can run common components in a search pipeline, such as re-ranking, query expansion, document classification/enrichment, and result set filtering. These components could nicely tie in with the components overview of TIREx,[3] where we could imagine someone searching for middleware in TIREx and wanting to integrate it into their own search specification.

We also discussed 'deployment' of these specifications; specifically, how to go from an abstract specification into a running search application. For that, we agreed that the components supported by the specifications (collection, backend, frontend, and middleware) can be run as containerized

[3] https://tira.io/tirex/components

software, and they should communicate using pre-defined API standards. Finally, we considered how updates to the document collection should be handled in a simple manner, but were unable to reach a conclusion due to time constraints.

# 4 Conclusion

The goal of the first WOWS workshop was to unite researchers passionate about open source and collaborative development and evaluation of web search components. We are happy with the number of submissions and the large amount of interest shown during the workshop. We obtained interesting insights from using TIREx as a platform for evaluation of the different components, which can help us improve this setup for possible future editions of the workshop.

# Acknowledgments

We are thankful to the organizers of ECIR 2024 for their assistance in hosting the workshop. We would also like to thank our keynote speaker, Negar Arabzadeh. Furthermore, we thank all paper authors, participants in the shared task, and the program committee for making the workshop a success: Daria Alexander, Alexander Bondarenko, Michael Dinzinger, Linda Erben, Maximilian Ernst, Guglielmo Faggioli, Hui Fang, Noor Afshan Fathima, Thibault Formal, Marcel Gohsen, Julius Gonsior, Christian Guetl, Lander Haag, Matthias Hagen, Maria Hampel, Claudia Hauff, Florian Heß, Malte-Christian Kuns, Wojciech Kusa, Gustav Lahmann, Carlos Lassance, Sean MacAvaney, Katja Mankinen, Vincent Melisch, Per Natzschka, Wilhelm Pertsch, Lina Razouk, Jan Heinrich Reimer, Anja Reusch, Ferdinand Schlatt, Benno Stein, Reiner Stolle, Nandan Thakur, Robert Thomas Thoss, Tuan Giang Trinh, Andrew Trotman, Jan van Acken, Matti Wiegmann, Ines Zelch, and Oleg Zendel. Finally, we thank Christine Plote of the Open Search Foundation for hosting and managing the workshop website.

The organizers of the workshop have received funding from the European Union's Horizon Europe research and innovation program under grant agreement No 101070014 (OpenWebSearch.EU)[4]. We are grateful for the EU's support.

# References

Mustafa Abualsaud, Mark D. Smucker, Christina Lioma, Maria Maistro, and Guido Zuccon. Overview of the TREC 2019 Decision track. In *Proceedings of TREC*, 2019.

Daria Alexander, Wojciech Kusa, and Arjen P. de Vries. ORCAS-I: queries annotated with intent using weak supervision. In Enrique Amigó, Pablo Castells, Julio Gonzalo, Ben Carterette, J. Shane Culpepper, and Gabriella Kazai, editors, *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022*, pages 3057–3066. ACM, 2022. doi: 10.1145/3477495.3531737. URL https://doi.org/10.1145/3477495.3531737.

[4] https://doi.org/10.3030/101070014

Daria Alexander, Wojciech Kusa, and Arjen P. de Vries. ORCAS-I query intent predictor as component of TIRA. In Farzana et al. [2024b], pages 23–29. URL https://ceur-ws.org/Vol-3689/WOWS_2024_paper_3.pdf.

Negar Arabzadeh and Charles LA Clarke. Fréchet distance for offline evaluation of information retrieval systems with sparse labels. *arXiv preprint arXiv:2401.17543*, 2024.

J. Arguello, M. Crane, F. Diaz, J. Lin, and A. Trotman. Report on the SIGIR 2015 workshop on reproducibility, inexplicability, and generalizability of results (RIGOR). *SIGIR Forum*, 49(2): 107–116, 2016.

Leif Azzopardi, Matt Crane, Hui Fang, Grant Ingersoll, Jimmy Lin, Yashar Moshfeghi, Harrisen Scells, Peilin Yang, and Guido Zuccon. The Lucene for information access and retrieval research (LIARR) workshop at SIGIR 2017. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1429–1430. Association for Computing Machinery, 2017. ISBN 9781450350228.

Christine Bauer, Ben Carterette, Nicola Ferro, Norbert Fuhr, Joeran Beel, Timo Breuer, Charles L. A. Clarke, Anita Crescenzi, Gianluca Demartini, Giorgio Maria Di Nunzio, Laura Dietz, Guglielmo Faggioli, Bruce Ferwerda, Maik Fröbe, Matthias Hagen, Allan Hanbury, Claudia Hauff, Dietmar Jannach, Noriko Kando, Evangelos Kanoulas, Bart P. Knijnenburg, Udo Kruschwitz, Meijie Li, Maria Maistro, Lien Michiels, Andrea Papenmeier, Martin Potthast, Paolo Rosso, Alan Said, Philipp Schaer, Christin Seifert, Damiano Spina, Benno Stein, Nava Tintarev, Julián Urbano, Henning Wachsmuth, Martijn C. Willemsen, and Justin Zobel. Report on the dagstuhl seminar on frontiers of information access experimentation for research and education. *SIGIR Forum*, 57(1):7:1–7:28, 2023a. doi: 10.1145/3636341.3636351. URL https://doi.org/10.1145/3636341.3636351.

Christine Bauer, Maik Fröbe, Dietmar Jannach, Udo Kruschwitz, Paolo Rosso, Damiano Spina, and Nava Tintarev. Overcoming methodological challenges in information retrieval and recommender systems through awareness and education. *Bauer et al.[2023a]*, pages 51–67, 2023b.

Alexander Bondarenko, Pavel Braslavski, Michael Völske, Rami Aly, Maik Fröbe, Alexander Panchenko, Chris Biemann, Benno Stein, and Matthias Hagen. Comparative Web Search Questions. In James Caverlee, Xia Ben Hu, Mounia Lalmas, and Wei Wang, editors, *13th ACM International Conference on Web Search and Data Mining (WSDM 2020)*, pages 52–60. ACM, February 2020a. doi: 10.1145/3336191.3371848.

Alexander Bondarenko, Maik Fröbe, Meriem Beloucif, Lukas Gienapp, Yamen Ajjour, Alexander Panchenko, Chris Biemann, Benno Stein, Henning Wachsmuth, Martin Potthast, and Matthias Hagen. Overview of Touché 2020: Argument Retrieval. In Avi Arampatzis, Evangelos Kanoulas, Theodora Tsikrika, Stefanos Vrochidis, Hideo Joho, Christina Lioma, Carsten Eickhoff, Aurélie Névéol, Linda Cappellato, and Nicola Ferro, editors, *Experimental IR Meets Multilinguality, Multimodality, and Interaction. 11th International Conference of the CLEF Association (CLEF 2020)*, volume 12260 of *Lecture Notes in Computer Science*, pages 384–395, Berlin Heidelberg New York, September 2020b. Springer. doi: 10.1007/978-3-030-58219-7_26.

Ryan Clancy, Nicola Ferro, Claudia Hauff, Jimmy Lin, Tetsuya Sakai, and Ze Zhong Wu. The SIGIR 2019 open-source IR replicability challenge (OSIRRC 2019). In Benjamin Piwowarski, Max Chevalier, Éric Gaussier, Yoelle Maarek, Jian-Yun Nie, and Falk Scholer, editors, *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, July 21-25, 2019*, pages 1432–1434. ACM, 2019.

Michael Dinzinger, Florian Heß, and Michael Granitzer. A Survey of Web Content Control for Generative AI. In Farzana et al. [2024b], pages 2–13. URL https://ceur-ws.org/Vol-3689/WOWS_2024_paper_1.pdf.

Linda Erben, Maria Hampel, Malte-Christian Kuns, Vincent Melisch, Per Natzschka, Wilhelm Pertsch, Lina Razouk, Reiner Stolle, Robert Thomas Thoss, Tuan Giang Trinh, Julius Gonsior, and Anja Reusch. Assembling four Open Web Search Components: TU Dresden at WOWS 2024. In Farzana et al. [2024b], pages 73–93. URL https://ceur-ws.org/Vol-3689/WOWS_2024_paper_8.pdf.

Guglielmo Faggioli, Oleg Zendel, J. Shane Culpepper, Nicola Ferro, and Falk Scholer. An enhanced evaluation framework for query performance prediction. In Djoerd Hiemstra, Marie-Francine Moens, Josiane Mothe, Raffaele Perego, Martin Potthast, and Fabrizio Sebastiani, editors, *Advances in Information Retrieval - 43rd European Conference on IR Research, ECIR 2021, Virtual Event, March 28 - April 1, 2021, Proceedings, Part I*, volume 12656 of *Lecture Notes in Computer Science*, pages 115–129. Springer, 2021. doi: 10.1007/978-3-030-72113-8_8. URL https://doi.org/10.1007/978-3-030-72113-8_8.

Sheikh Mastura Farzana, Maik Fröbe, Michael Granitzer, Gijs Hendriksen, Djoerd Hiemstra, Martin Potthast, and Saber Zerhoudi. The First International Workshop on Open Web Search (WOWS). In Nazli Goharian, Nicola Tonellotto, Yulan He, Aldo Lipani, Graham McDonald, Craig Macdonald, and Iadh Ounis, editors, *Advances in Information Retrieval - 46th European Conference on Information Retrieval, ECIR 2024, Glasgow, UK, March 24-28, 2024, Proceedings, Part V*, volume 14612 of *Lecture Notes in Computer Science*, pages 426–431. Springer, 2024a. doi: 10.1007/978-3-031-56069-9_58. URL https://doi.org/10.1007/978-3-031-56069-9_58.

Sheikh Mastura Farzana, Maik Fröbe, Gijs Hendriksen, Michael Granitzer, Djoerd Hiemstra, Martin Potthast, and Saber Zerhoudi, editors. *Proceedings of the first International Workshop on Open Web Search co-located with the 46th European Conference on Information Retrieval ECIR 2024*, number 3689 in CEUR Workshop Proceedings, Aachen, 2024b. URL https://ceur-ws.org/Vol-3689/.

Thibault Formal, Carlos Lassance, Benjamin Piwowarski, and Stéphane Clinchant. SPLADE v2: Sparse lexical and expansion model for information retrieval. *CoRR*, abs/2109.10086, 2021a. URL https://arxiv.org/abs/2109.10086.

Thibault Formal, Benjamin Piwowarski, and Stéphane Clinchant. SPLADE: sparse lexical and expansion model for first stage ranking. In Fernando Diaz, Chirag Shah, Torsten Suel, Pablo

Castells, Rosie Jones, and Tetsuya Sakai, editors, *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, pages 2288–2292. ACM, 2021b. doi: 10.1145/3404835.3463098. URL https://doi.org/10.1145/3404835.3463098.

Maik Fröbe, Jan Heinrich Reimer, Sean MacAvaney, Niklas Deckers, Simon Reich, Janek Bevendorff, Benno Stein, Matthias Hagen, and Martin Potthast. The information retrieval experiment platform. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2023a.

Maik Fröbe, Matti Wiegmann, Nikolay Kolyada, Bastian Grahm, Theresa Elstner, Frank Loebe, Matthias Hagen, Benno Stein, and Martin Potthast. Continuous Integration for Reproducible Shared Tasks with TIRA.io. In *Advances in Information Retrieval. 45th European Conference on IR Research (ECIR 2023)*, Lecture Notes in Computer Science. Springer, 2023b.

Maik Fröbe, Harrisen Scells, Theresa Elstner, Christopher Akiki, Lukas Gienapp, Jan Heinrich Reimer, Sean MacAvaney, Benno Stein, Matthias Hagen, and Martin Potthast. Resources for Combining Teaching and Research in Information Retrieval Courses. In *47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2024)*. ACM, July 2024.

Marcel Gohsen and Benno Stein. Integrating Query Interpretation Components into the Information Retrieval Experiment Platform. In Farzana et al. [2024b], pages 63–72. URL https://ceur-ws.org/Vol-3689/WOWS_2024_paper_7.pdf.

Michael Granitzer, Stefan Voigt, et al. Impact and development of an open web index for open web search. *Journal of the Association for Information Science and Technology*, 2023.

Gijs Hendriksen, Michael Dinzinger, Sheikh Mastura Farzana, Noor Afshan Fathima, Maik Fröbe, Sebastian Schmidt, Saber Zerhoudi, Michael Granitzer, Matthias Hagen, Djoerd Hiemstra, Martin Potthast, and Benno Stein. The open web index - crawling and indexing the web for public use. In Nazli Goharian, Nicola Tonellotto, Yulan He, Aldo Lipani, Graham McDonald, Craig Macdonald, and Iadh Ounis, editors, *Advances in Information Retrieval - 46th European Conference on Information Retrieval, ECIR 2024, Glasgow, UK, March 24-28, 2024, Proceedings, Part V*, volume 14612 of *Lecture Notes in Computer Science*, pages 130–143. Springer, 2024. doi: 10.1007/978-3-031-56069-9_10. URL https://doi.org/10.1007/978-3-031-56069-9_10.

Vaibhav Kasturia, Marcel Gohsen, and Matthias Hagen. Query Interpretations from Entity-Linked Segmentations. In *15th ACM International Conference on Web Search and Data Mining (WSDM 2022)*. ACM, February 2022. doi: 10.1145/3488560.3498532.

Dirk Lewandowski and Sebastian Schultheiß. Public awareness and attitudes towards search engine optimization. *Behaviour & Information Technology*, 42:1025–1044, 2022.

Sean MacAvaney, Andrew Yates, Sergey Feldman, Doug Downey, Arman Cohan, and Nazli Goharian. Simplified data wrangling with ir_datasets. In Fernando Diaz, Chirag Shah, Torsten

Suel, Pablo Castells, Rosie Jones, and Tetsuya Sakai, editors, *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, pages 2429–2436. ACM, 2021. doi: 10.1145/3404835.3463254. URL https://doi.org/10.1145/3404835.3463254.

Sean MacAvaney, Nicola Tonellotto, and Craig Macdonald. Adaptive re-ranking with a corpus graph. In Mohammad Al Hasan and Li Xiong, editors, *Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, October 17-21, 2022*, pages 1491–1500. ACM, 2022. doi: 10.1145/3511808.3557231. URL https://doi.org/10.1145/3511808.3557231.

Craig Macdonald, Nicola Tonellotto, Sean MacAvaney, and Iadh Ounis. Pyterrier: Declarative experimentation in Python from BM25 to dense retrieval. In *Proceedings of the 30th acm international conference on information & knowledge management*, pages 4526–4533, 2021.

Astrid Mager, Cristian Norocel, and Richard Rogers, editors. *The State of Google Critique and Intervention*, 2023. Sage Journals.

Rodrigo Nogueira and Jimmy Lin. From doc2query to doctttttquery. *Online preprint*, 2019.

Alexander Ratner, Stephen H. Bach, Henry R. Ehrenberg, Jason Alan Fries, Sen Wu, and Christopher Ré. Snorkel: Rapid training data creation with weak supervision. *Proc. VLDB Endow.*, 11(3):269–282, 2017. doi: 10.14778/3157794.3157797. URL http://www.vldb.org/pvldb/vol11/p269-ratner.pdf.

Harrisen Scells, Shengyao Zhuang, and Guido Zuccon. Reduce, reuse, recycle: Green information retrieval research. In Enrique Amigó, Pablo Castells, Julio Gonzalo, Ben Carterette, J. Shane Culpepper, and Gabriella Kazai, editors, *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022*, pages 2825–2837. ACM, 2022.

Ferdinand Schlatt. Efficiently Scoring the Health-relatedness of Web Pages. In Farzana et al. [2024b], pages 14–22. URL https://ceur-ws.org/Vol-3689/WOWS_2024_paper_2.pdf.

Ferdinand Schlatt, Dieter Bettin, Matthias Hagen, Benno Stein, and Martin Potthast. Mining Health-related Cause-Effect Statements with High Precision at Large Scale. In Nicoletta Calzolari, Chu-Ren Huang, Hansaem Kim, James Pustejovsky, Leo Wanner, Key-Sun Choi, Pum-Mo Ryu, Hsin-Hsi Chen, Lucia Donatelli, Heng Ji, Sadao Kurohashi, Patrizia Paggio, Nianwen Xue, Seokhwan Kim, Younggyun Hahm, Zhong He, Tony Kyungil Lee, Enrico Santus, Francis Bond, and Seung-Hoon Na, editors, *29th International Conference on Computational Linguistics (COLING 2022)*, pages 1925–1936. International Committee on Computational Linguistics, October 2022. URL https://aclanthology.org/2022.coling-1.167.

Andrew Trotman, Charles L.A. Clarke, Iadh Ounis, Shane Culpepper, Marc-Allen Cartright, and Shlomo Geva. Open source information retrieval: A report on the SIGIR 2012 workshop. *SIGIR Forum*, 46(2):95–101, dec 2012. ISSN 0163-5840.

Matti Wiegmann, Jan Heinrich Reimer, Maximilian Ernst, Martin Potthast, Matthias Hagen, and Benno Stein. A Mastodon Corpus to Evaluate Federated Microblog Search. In Farzana et al. [2024b], pages 37–49. URL https://ceur-ws.org/Vol-3689/WOWS_2024_paper_5.pdf.

Wai Gen Yee, Michel Beigbeder, and Wray Buntine. SIGIR06 workshop report: Open source information retrieval systems (OSIR06). *ACM SIGIR Forum*, 40(2):61–65, 2006.

Ines Zelch, Gustav Lahmann, and Matthias Hagen. Embedding-based Query Spelling Correction. In Farzana et al. [2024b], pages 30–36. URL https://ceur-ws.org/Vol-3689/WOWS_2024_paper_4.pdf.

Oleg Zendel, Maik Fröbe, and Guglielmo Faggioli. QPPTK@TIREx: Simplified Query Performance Prediction for Ad-Hoc Retrieval Experiments. In Farzana et al. [2024b], pages 50–62. URL https://ceur-ws.org/Vol-3689/WOWS_2024_paper_6.pdf.