WOWS 2025

Proceedings of the second International Workshop on Open Web Search



CEUR Workshop Proceedings ISSN 1613-0073

Preface

This volume contains the papers of WOWS 2025, the second International Workshop on Open Web Search, co-located with the 47th European Conference on Information Retrieval (ECIR) in Lucca, Italy, on 10 April 2025.

WOWS 2025 had two calls for contributions. The first call targets scientific contributions on collaborative search engine building, including crawling, deployment, evaluation, and use of the web as a resource by researchers and innovators. The second call introduces the shared task called WOWS-Eval, which aims to gain practical experience with collaborative, cooperative search engine evaluation by focusing on enriching the Open Web Index (OWI) with relevance judgments cooperatively transferred from existing TREC-like test collections.

The workshop involved a keynote by Charles Clarke from the University of Waterloo, 6 paper presentations, and a breakout session with participant discussions. The Workshop on Open Web Search received funding from the European Union's Horizon Europe research and innovation program under grant agreement No 101070014 (OpenWebSearch.EU, https://doi.org/10.3030/101070014). We would like to thank the Program Committee members for helpful reviews and suggestions to improve the contributions to the workshop. Special thanks go to the Open Search Foundation for managing the WOWS 2025 website at: https://opensearchfoundation.org/en/events-osf/wows2025/

10 April 2025

Sheikh Mastura Farzana, Maik Fröbe, Michael Granitzer, Gijs Hendriksen, Djoerd Hiemstra, Martin Potthast, Arjen P. de Vries, Saber Zerhoudi

Program Committee

Daria Alexander (Radboud University)
Timo Breuer (TH Cologne)
Claudia Hauff (Spotify)
Jan Heinrich Merker (Friedrich-Schiller-University Jena)
Philipp Schaer (TH Cologne)
Nicola Tonellotto (University of Pisa)

Table of Contents

Preface
Program committee
Keynote: Annotative Indexing
Investigating the Robustness of Embedding Models on Noisy Input Texts
Michael Granitzer
Document Quality Scoring for Web Crawling
Francesca Pezzuti, Ariane Mueller, Sean MacAvaney, Nicola Tonellotto
On Precomputation and Caching in Information Retrieval Experiments with Pipeline Architectures
Sean MacAvaney, Craig Macdonald
Combining Embedding Models for RAG: A Similarity-Based Approach
Kanishka Ghosh Dastidar, Michael Dinzinger, Laura Caspari, Jelena Mitrović, Michael Granitzer
Supporting Vertical Web Search and Customized Search Applications with the
Modular and Open Framework MOSAIC
Lexical Baselines for Relevance Label Transfer
Daria Alexander, Maik Fröbe, Gijs Hendriksen