

WOWS 2024

Proceedings of the first International Workshop
on Open Web Search



CEUR Workshop Proceedings
ISSN 1613-0073

Preface

This volume contains the papers of Wows 2024, the first International Workshop on Open Web Search, co-located with the 46th European Conference on Information Retrieval (ECIR) in Glasgow, UK, on 28 March 2024.

Wows 2024 had two calls for contributions. The first call targets scientific contributions on cooperative search engine development. This includes cooperative crawling of the web and cooperative deployment and evaluation of search engines. We specifically highlight the potential of enabling public and commercial organizations to use an indexed web crawl as a resource to create innovative search engines tailored to specific user groups, instead of relying on one search engine provider. The second call aims at gaining practical experience with joint, cooperative evaluation of search engine prototypes and their components using the Information Retrieval Experiment Platform TIREx.

The workshop involved a keynote by Negar Arabzadeh from the University of Waterloo, 8 paper presentations (5 full papers and 3 short papers accepted out of 13 submissions), and a breakout session with participant discussions. Wows 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 Christine Plote, Managing Director of the Open Search Foundation for the Wows 2024 website at: <https://opensearchfoundation.org/wows2024/>

28 March 2024

Sheikh Mastura Farzana, Maik Fröbe, Gijs Hendriksen, Michael Granitzer,
Djoerd Hiemstra, Martin Potthast, Saber Zerhoudi

Program Committee

Jan van Acken (Utrecht University)

Hui Fang (University of Delaware)

Noor Afshan Fathima (CERN)

Christian Guetl (TU Graz)

Claudia Hauff (Spotify)

Katja Mankinen (IT Center for Science, CSC)

Craig Macdonald (University of Glasgow)

Jan Heinrich Reimer (Friedrich-Schiller-Universität Jena)

Benno Stein (Bauhaus-Universität Weimar)

Nicola Tonello (University of Pisa)

Andrew Trotman (University of Otago)

Table of Contents

Preface	i
Program committee	ii
Keynote: Evaluation of Information Access Systems in the Generative Era	1
<i>Negar Arabzadeh</i>	
A Survey of Web Content Control for Generative AI	2
<i>Michael Dinzinger, Florian Heß, Michael Granitzer</i>	
Efficiently Scoring the Health-relatedness of Web Pages.....	14
<i>Ferdinand Schlatt</i>	
ORCAS-I query intent predictor as component of TIRA	23
<i>Daria Alexander, Wojciech Kusa, Arjen P. de Vries</i>	
Embedding-based Query Spelling Correction	30
<i>Ines Zelch, Gustav Lahmann, Matthias Hagen</i>	
A Mastodon Corpus to Evaluate Federated Microblog Search	37
<i>Matti Wiegmann, Jan Heinrich Reimer, Maximilian Ernst, Martin Potthast, Matthias Hagen, Benno Stein</i>	
QPPTK@TIREx: Simplified Query Performance Prediction for Ad-Hoc Retrieval	50
Experiments <i>Oleg Zendel, Maik Fröbe, Guglielmo Faggioli</i>	
Integrating Query Interpretation Components into the Information Retrieval	63
Experiment Platform <i>Marcel Gohsen, Benno Stein</i>	
Assembling four Open Web Search Components: TU Dresden at WOWS 2024	73
<i>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, Anja Reusch</i>	

