

DATA:SEARCH'18 – Searching Data on the Web

Preface

Paul Groth
Elsevier Labs
p.groth@elsevier.com

Laura Koesten
The Open Data Institute;
University of Southampton
laura.koesten@theodi.org

Philipp Mayr
GESIS – Leibniz
Institute for the Social Sciences
philipp.mayr@gesis.org

Maarten de Rijke
University of Amsterdam
derijke@uva.nl

Elena Simperl
University of Southampton
e.simperl@soton.ac.uk

This half day workshop explores challenges in data search, with a particular focus on data on the web. We want to stimulate an interdisciplinary discussion around how to improve the description, discovery, ranking and presentation of structured and semi-structured data, across data formats and domain applications. We welcome contributions describing algorithms and systems, as well as frameworks and studies in human data interaction. The workshop aims to bring together communities interested in making the web of data more discoverable, easier to search and more user friendly.

The aim of the workshop is to be a venue to present and exchange ideas and experiences for discovering and searching all types of structured or semi-structured datasets and to discuss how concepts and lessons learned from academic search, entity search, digital libraries, and web search could be transferred to data search scenarios.

The keynote will be given by Professor Krisztian Balog on "Table Retrieval and Generation", followed by a paper presentation titled "Recognizing Quantity Names for Tabular Data" by Yang Yi, Zhiyu Chen, Jeff Heflin and Brian Davison. The workshop will include lightning talks by participants and a round table discussion.

The opportunities to share and establish links between different perspectives on search and discovery for different kinds of data are significant and can inform the design of a wide range of information retrieval technologies, including search engines, recommender systems and conversational agents.

A broad range of methods and insights are important to enable the discovery of, and access to, data published on the web, including:

- analyzing contextual information for datasets, including mentions of datasets
- browsing and query support for structured and semi-structured data
- inference and data enrichment systems
- learning to match for datasets
- learning to rank datasets
- mining direct links between documents, datasets or data records
- summaries and descriptions of datasets targeting users or search engines
- concepts and methods to present data and entity-centric results.

We see a large space for discussion and future research in the development of federated data discovery and search technologies, which leverages the most recent advances in information retrieval, Semantic Web and databases, and is mindful of human factors. Workshop website: <https://datasearch-ws.github.io/2018/>.

Copyright © by the paper's authors. Copying permitted for private and academic purposes.

In: Joint Proceedings of the First International Workshop on Professional Search (ProfS2018); the Second Workshop on Knowledge Graphs and Semantics for Text Retrieval, Analysis, and Understanding (KG4IR); and the International Workshop on Data Search (DATA:SEARCH18). Co-located with SIGIR 2018, Ann Arbor, Michigan, USA – 12 July 2018, published at <http://ceur-ws.org>

PROGRAMME COMMITTEE

- Alexander Kotov (Wayne State University)
- Arjen de Vries (Radboud University Nijmegen)
- Arno Scharl (Modul University Vienna)
- Axel Polleres (Vienna University of Economics and Business)
- Eva Méndez (Open research data)
- Kuansan Wang (Microsoft)
- Laura Dietz (University of New Hampshire)
- Michael Gubanov (University of Texas, San Antonio)
- Peter Haase (Metaphacts)
- Steffen Lohmann (Fraunhofer IAIS)