

PROFILES14: 1st International Workshop on Dataset PROFiling & fEderated Search for Linked Data

Preface



While the Web of Data, and in particular Linked Data, has seen tremendous growth over the past years, take-up, usage and reuse of data is still limited and is often focused on well-known reference datasets. The main obstacles preventing users from obtaining relevant, correct and up-to-date information from distributed LOD datasets is the lack of scalable and usable methods for formulating and distributing semantic and keyword queries across the Web of Data. This problem is further alleviated by the lack of trust in the quality of search results retrieved using federated search over distributed third party data. Hence, dataset and endpoint selection and discovery are inherent challenges for query distribution. These are currently hindered by the lack of trust-worthy and up-to-date information about the nature, characteristics, currentness and suitability of particular datasets for a given task. Given the heterogeneous and large-scale context of Linked Open Data, state-of-the-art semantic and keyword search techniques for structured data face increased query ambiguity and scalability problems already in single-source search scenarios. In the federated search scenarios, dataset selection and adoption of queries to the respective schemas used poses even further challenges.

PROFILES14: the 1st International Workshop on Dataset PROFiling & fEderated Search for Linked Data¹ had been organised on the 26 May 2014 during the 11th ESWC (ESWC2014) held in Anissaras / Hersonissou, Crete, Greece². The PROFILES14 workshop aimed to gather innovative search approaches for large-scale, distributed and heterogeneous linked datasets inline with dedicated approaches to analyse, describe and discover endpoints, as an inherent task of query distribution. PROFILES14 equally considered both novel scientific methods and techniques for querying, assessment, profiling, and curation distributed datasets as well as the application perspective, such as the innovative use of tools and methods for providing structured knowledge about distributed datasets, their evolution and fundamentally, means to search and query the Web of Data.

PROFILES14 received a good amount of very relevant and original submissions, six of which have been accepted as full papers and three as poster papers. This volume covers a range of topics related to data source contextualization for search and exploration in Linked Data, profiling of linked datasets as well as measuring and modelling dynamics and evolution of Linked Data. In particular, the contributions in this volume include novel approaches towards exploring Linked Data sources by providing contextually relevant data source recommendations, methods for optimization of SPARQL queries for Linked Data profiling, sampling methods to estimate distributions of Linked Data and vocabulary extensions for expressing connectivity metrics. With

¹ <http://www.keystone-cost.eu/profiles2014/>

² <http://2014.eswc-conferences.org/>

respect to the dynamics and evolution, methods for measuring dynamics of linked datasets, as well as models for preservation of evolving Linked Data are proposed. Finally, it includes practical cases of using Linked Data in collaborative education-related scenarios. The paper "Entity-based Data Source Contextualization for Searching the Web of Data" by Andreas Wagner, Peter Haase, Achim Rettinger and Holger Lamm has been selected for the 'Best of Workshops' session in the main program track of the ESWC 2014.

We would like to take this opportunity to sincerely thank the the authors for their invaluable and inspiring contributions to the workshop. Our sincere thanks are due to the program committee members for reviewing the submissions and assuring the good quality of the workshop program. We are also very grateful to the organisers of the ESWC 2014 conference and especially to the ESWC 2014 Workshop Chair Dr. Harald Sack for his support in the workshop organisation. Great appreciation of his time and expertise goes to our keynote speaker Dr. Thanassis Tiropanis for his talk entitled "Linked Data Affordances and Challenges for Web Observatories".

May 2014

Elena Demidova
Stefan Dietze
Julian Szymański
John Breslin

The PROFILES14 workshop was co-organised by the ICT COST Action KEYSTONE Semantic keyword-based search on structured data sources (IC1302).

Organisation

Organising Committee

Elena Demidova, L3S Research Center, Leibniz Universität Hannover Germany
Stefan Dietze, L3S Research Center, Leibniz Universität Hannover, Germany
Julian Szymański, Gdańsk University of Technology, Poland
John Breslin, National University of Ireland, Galway, Ireland

Program Committee

Charlie Abela, University of Malta, Malta
Dan Brickley, W3C and Google, UK
Marco Antonio Casanova, Pontifical Catholic University of Rio de Janeiro, Brazil
Philippe Cudré-Mauroux, University of Fribourg, Switzerland
Maciej Dabrowski, Insight Centre for Data Analytics, NUI Galway, Ireland
Mathieu d'Aquin, The Open University, UK
Christophe Gueret, Data Archiving Network Services (DANS), NL
Tom Heath, The Open Data Institute (ODI), UK
Laura Hollink, Vrije Universiteit Amsterdam, NL
Markus Luczak-Rösch, University of Southampton, UK
Abdulhussain E. Mahdi, University of Limerick, Ireland
Andreas Nürnberger, Otto-von-Guericke-Universität Magdeburg, Germany
Heiko Paulheim, Universität Mannheim, Germany
Bernardo Pereira Nunes, PUC-Rio, Brazil
Carlos Pedrinaci, The Open University, UK
Andreas Rauber, Vienna University, Austria
Thanassis Tiropanis, University of Southampton, UK
Raquel Trillo Lado, Universidad Zaragoza, Spain
Pierre-Yves Vandenbussche, Fujitsu Limited, Ireland

List of Contributions

Full papers

Andreas Wagner, Peter Haase, Achim Rettinger and Holger Lamm. Entity-based Data Source Contextualization for Searching the Web of Data.

Sergej Zerr, Mathieu D'Aquin, Ivana Marenzi, Davide Taibi, Adamou Alessandro and Stefan Dietze. Towards Analytics and Collaborative Exploration of Social and linked Media for Technology-Enhanced Learning Scenarios.

Benedikt Forchhammer, Anja Jentzsch and Felix Naumann. LODOP - Multi-Query Optimization for Linked Data Profiling Queries.

Renata Dividino, Thomas Gottron, Ansgar Scherp and Gerd Gröner. From Changes to Dynamics: Dynamics Analysis of Linked Open Data Sources.

Thomas Gottron. Of Sampling and Smoothing: Approximating Distributions over Linked Open Data.

Michalis Mountantonakis, Carlo Allocca, Pavlos Fafalios, Nikos Minadakis, Yannis Marketakis, Cristina Lantzaki and Yannis Tzitzikas. Extending VoID for Expressing Connectivity Metrics of a Semantic Warehouse.

Poster papers

Mohamed Ben Ellefi, Zohra Bellahsene, François Scharffe and Konstantin Todorov. Towards Semantic Dataset Profiling.

Marios Meimaris, George Papastefanatos, Christos Pateritsas, Theodora Galani and Yannis Stavrakas. Towards a Framework for Managing Evolving Information Resources on the Data Web.