

The 1st Workshop on Benchmarking RDF Systems

BeRSys2013 @ ESWC-2013

On behalf of all the organizers, the BerSys 2013 chairs are happy to welcome you to the 1st International Workshop on Benchmarking RDF Systems!

Our goal when proposing the workshop has been to create a discussion forum where researchers and industrials can meet to discuss topics related to the performance of RDF systems, and more widely, to discuss best practices, different application needs and scenarios related to RDF data management.

Processing efficiently large amounts of RDF data is currently a hot problem in many applications and domains. From the Semantic Web perspective, one of the most important ones is the W3C linking open data initiative that has boosted the publication and interlinking of datasets on the Semantic Web, resulting in the Linked Open Data Cloud. These datasets with billions of RDF triples such as Wikipedia, DBPedia, and government sites have been created and published online. Moreover, numerous datasets and vocabularies are published nowadays as RDF graphs most notably in life and earth sciences in order to facilitate community annotation and interlinking of both scientific and scholarly data of interest. A significant number of commercial semantic repositories exist by now (RDF databases endowed with both reasoning and query evaluation capabilities) that are cornerstone tools for the Semantic Web vision.

Despite the widespread use of RDF data, there is no *comprehensive suite of benchmarks* to facilitate the advancement of technology by providing both academia and industry with clear targets for performance and functionality. Existing benchmarks have been mainly developed to test the performance of research prototypes and do not intend to capture the full spectrum of RDF data management.

In this context, the aim of BerSys has been to provide a forum where topics specific to RDF performance analysis and benchmarking can be discussed and elaborated.

The keynote presentation of the workshop is by Professor Manolis Koubarakis from the University of Crete. Prof. Koubarakis has worked extensively on Semantic Web data management topics, having recently led a successful research effort to define spatial extensions to RDF. In his keynote Prof. Koubarakis presents a functional and performance benchmark which is used to compare the functionality and performance of most of the geospatial RDF stores that are available today

The BerSys workshop features three accepted papers.

In their work titled "Towards a SPARQL 1.1 Feature Benchmark on Real-World Social Network Data", Martin Przyjaciel-Zablocki, Alexander Schätzle, Thomas Hornung and Io Taxidou consider the very hot area of social network data management, and revisit it from the perspective of RDF data manipulations by means of SPARQL 1.1, the latest W3C standard language for querying RDF; the work is validated by preliminary experiments on data from the last.fm social site.

Daniele Dell’Aglia, Marco Balduini and Emanuele Della Valle contribute the work “On the need to include functional testing in RDF stream engine benchmarks”. Given the concomitant increase of interest in both RDF data and streams, numerous RDF streaming engines have been developed, as well as some performance benchmarks. However, the authors of the present paper point out that correctness and the operational semantics of each engine are all but straightforward, given the different interpretations assigned by each system to queries over stream data. The authors provide guidelines on how semantics and correctness can be taken into account within existing RDF stream engine performance benchmarks.

Tatiana Tarasova and Maarten Marx present “ParlBench: a SPARQL-benchmark for electronic publishing applications”. The work is based on a real open data set, consisting of proceedings of the Dutch parliament, a richly structured text (or, equivalently, text-rich graph data), a valuable real source to test RDF data management techniques against. As pointed out in the paper, such data sets contribute to answering the need of real data sources on which to test “Big Data” tools and techniques, a need recently discussed by Gerhard Weikum on the ACM SIGMOD blog. The authors propose a benchmark consisting of 28 queries, some of which are inspired from operations currently performed on the data as part of the Parliament’s functioning. The work concludes with experimental benchmark results gathered on the Virtuoso RDF store.

We are looking forward to the workshop’s presentation and discussions. Given the richness and the many facets of RDF usage, we expect that interest in RDF system evaluation for completeness, correctness and performance will remain very strong in the next few years. We are glad BerSys will give a venue for discussing and building connections among such efforts.

We would like to thank the ESWC conference organizers for their valuable input and support, and the BerSys program committee for their thoughtful and constructive work while reading the submitted papers.

May, 2013

Irini Fundulaki, Ioana Manolescu and Ioan Toma

Workshop Chairs:

Irini Fundulaki, FORTH, Greece

Ioana Manolescu, Inria Saclay, France

Ioan Toma, STI Innsbruck, Austria