



HOBBIT

Newsletter
May 2017

HOBBIT
Holistic Benchmarking of
Big Linked Data

H2020 Research and
Innovation Grant
Agreement No. 688227

Contact
Dr. Axel-Cyrille Ngonga
Ngomo
Institute for Applied
Informatics
Hainstraße 11, 04109
Leipzig Germany
Phone: +49 341 97 32362
ngonga@informatik.uni-
leipzig.de



Dear Hobbit friends and Community members,

We are happy to announce that HOBBIT is actively participating in [ESWC 2017](#)! ESWC is a major venue for discussing the latest scientific results and technology innovations around semantic technologies. Building on its past success, ESWC is seeking to broaden its focus to span other relevant related research areas in which Web semantics plays an important role. It takes place from May 28th to June 1st, 2017 in Portoroz, Slovenia. Within ESWC 2017, members of the HOBBIT Consortium will present the Tutorial "Link Discovery – Algorithms, Approaches and Benchmarks", organize the [Workshop on Querying the Web of Data \(OuWeDa 2017\)](#) and the [Mighty Storage \(MOCHA\)](#), [Query Answering over Linked Data \(QALD\)](#) and [Open Knowledge Extraction \(OKE\) Challenges](#).

[Tutorial on Link Discovery – Algorithms, Approaches and Benchmarks](#) by *Axel-Cyrille Ngonga Ngomo (InfAI)*, *Irini Fundulaki (FORTH)* and *Mohamed Ahmed Sherif (InfAI)*.

Link Discovery is a task of central importance when creating Linked Datasets. Two challenges need to be addressed when carrying out link discovery: first, the quadratic a-priori runtime complexity of this task demands the development of time-efficient approaches for linking large datasets. Second, the need for accuracy demands the development of generic approaches that can detect correct and complete sets of links. Third, the development of benchmarks that test the ability of instance matching techniques and tools is crucial for identifying and addressing the technical difficulties and challenges in this domain. In this tutorial, we aim to help the audience when faced with all three challenges. First, we will provide an overview of existing solutions for link discovery. Then, we will look into some of the state-of-art algorithms for the rapid execution of link discovery tasks. In particular, we will focus on algorithms which guarantee result completeness. Then, we will present algorithms for the detection of complete and correct sets of links. Here, our focus will be on supervised, active and unsupervised machine learning algorithms. Last, we will discuss existing instance matching benchmarks for Linked Data and we will conclude the tutorial by providing hands-on experience with one of the most commonly used link discovery frameworks, [Limes](#). For more information see [here](#).

[Workshop on Querying the Web of Data \(OuWeDa 2017\)](#)

The constant growth of Linked Open Data (LOD) on the Web opens new challenges pertaining to querying such massive amounts of publicly available data. LOD datasets are available through various interfaces, such as data dumps, SPARQL endpoints and triple pattern fragments. In addition, various sources produce streaming data. Efficiently querying these sources is of central importance for the scalability of Linked Data and Semantic Web technologies. The trend of publicly available and interconnected data is shifting the focus of Web technologies towards new paradigms for Linked Data querying. To exploit the massive

amount of LOD data to its full potential, users should be able to query and combine this data easily and effectively. More information on QuWeDa 2017 Workshop can be found [here](#).

[Mighty Storage Challenge \(MOCHA\)](#)

Triple stores are the backbone of most applications based on Linked Data. Hence, devising systems that achieved an acceptable performance on real datasets and real loads is of central importance for the practical applicability of SemanticWeb technologies. So far, it is only partly known whether we have already passed this cap. With this challenge, we aim to (1) provide objective measures for how well current systems (including 3 commercial systems, which have already expressed their desire to participate) perform on real tasks of industrial relevance and (2) detect bottlenecks of existing systems to further their development towards practical usage.

The aim of the Mighty Storage Challenge is to test the performance of solutions for SPARQL processing in aspects that are relevant for modern applications. These include ingesting data, answering queries on large datasets and serving as backend for applications driven by Linked Data. The proposed challenge will test the systems against data derived from real applications and with realistic loads. An emphasis will be put on dealing with changing data in form of streams or updates. This version of the challenge (which we aim to run periodically for at least 3 years) will comprise the following tasks: (1) ingestion, (2) storage, (3) versioning and (4) browsing for RDF data. For more details on Mighty Storage Challenge, see [here](#).

[Question Answering over Linked Data \(QALD-7\) Challenge](#)

The past years have seen a growing amount of research on question answering (QA) over Semantic Web data, shaping an interaction paradigm that allows end users to profit from the expressive power of Semantic Web standards while at the same time hiding their complexity behind an intuitive and easy-to-use interface. At the same time the growing amount of data has led to an heterogeneous data landscape where QA systems struggle to keep up with the volume, variety and veracity of the underlying knowledge.

The Question Answering over Linked Data (QALD) challenge aims at providing an up-to-date benchmark for assessing and comparing state-of-the-art-systems that mediate between a user, expressing his or her information need in natural language, and RDF data. It thus targets all researchers and practitioners working on querying Linked Data, natural language processing for question answering, multilingual information retrieval and related topics. The main goal is to gain insights into the strengths and shortcomings of different approaches and into possible solutions for coping with the large, heterogeneous and distributed nature of Semantic Web data.

QALD has a 6-year history of developing a benchmark that is increasingly being used as standard evaluation benchmark for question answering over Linked Data. Since many of the topics relevant for QA over Linked Data lie at the core of ESWC (Multilinguality, Semantic Web, Human-Machine-Interfaces), we will run the 7th instantiation of QALD again at ESWC 2017. HOBBIT project guarantees a controlled setting involving rigorous evaluations via its platform. For more details on Question Answering over Linked Data (QALD-7) challenge, see [here](#).

[Open Knowledge Extraction \(OKE\) Challenge](#)

The Open Knowledge Extraction Challenge invites researchers and practitioners from academia as well as industry to compete with the aim of pushing further the state of the art of knowledge extraction for the Semantic Web. Most of the Web content consists of natural language text, e.g., websites, news, blogs, micro-posts, etc., hence a main challenge is to extract as much relevant knowledge

as possible from this content, and publish it in the form of Semantic Web triples. There is huge work on knowledge extraction (KE) and knowledge discovery contributing to address this problem. In fact, results of knowledge extraction systems are usually evaluated against tasks that do not focus on specific Semantic Web goals. For example, tasks such as named entity recognition, named entity disambiguation and relation extraction are certainly of importance for the SW, but in most cases such tasks are designed without considering the output design and formalization in the form of Linked Data and OWL ontologies. This makes the results of existing methods often not directly reusable for populating the SW, until a translation from linguistic semantics to formal semantics is performed.

The goal of the Open Knowledge Extraction Challenge is to test the performance of Knowledge Extraction Systems with respect to the Semantic Web. The OKE challenge has the ambition to provide a reference framework for research on "Knowledge Extraction from text for the Semantic Web" by redefining a number of tasks (typically from information and knowledge extraction) by taking into account specific Semantic Web requirements. For more details on Open Knowledge Extraction Challenge, see [here](#).

Join our community to remain informed at project-hobbit.eu/get-involved. If you have any questions, do not hesitate to contact us as explained at project-hobbit.eu/contacts. We are looking forward to benchmarking your system soon. Stay tuned!

Best,
HOBBIT team.



Share



Tweet



Forward