

HOBBIT: Holistic Benchmarking for Big Linked Data

Axel-Cyrille Ngonga Ngomo and Michael Röder

Universität Leipzig, Institut für Informatik, AKSW,
Postfach 100920, D-04009 Leipzig, Germany,
ngonga|roeder@informatik.uni-leipzig.de

Project Overview. Linked Data has gained significant momentum over the last years and is now used at industrial scale in many sectors including bio-medicine, plant manufacturing, smart cities, content creation and management, data journalism, asset management and storage, banking and many more. In these sectors, an increasingly large amount (volume) of rapidly changing (velocity) data in different formats (i.e., unstructured to structured data, variety) needs to be integrated (vocabulary), refined, curated (veracity) or even bundled (value) to serve as data basis for large-scale applications. However, deciding on the right framework to build a given pipeline is still a demanding endeavour.

Works such as GERBIL [3] and BAT [2] point to the comparability of Linked Data solutions in the area of Named Entity Recognition and Linking being undermined by (1) frameworks being evaluated on different datasets (2) based on different implementations of homonymous measures as well as on (3) non-standardized hardware. This makes the comparability of the results presented across different publications rather difficult to establish. This reality however holds across the whole of the Linked Data lifecycle.

HOBBIT aims to address these drawbacks by providing a family of industry-relevant benchmarks for the Big Linked Data value chain through a generic evaluation platform deployed on standardized hardware. In more detail, we aim to:

- Make open deterministic data generators available to test the performance of existing systems. The data will mimic real industrial data assembled during the course of the project. We will organize or join challenges that aim to measure the performance of technologies for the different steps of the lifecycle. In contrast to existing benchmarks, we will also provide benchmarks that allow to benchmark whole suites of software that cover more than one step of the Big Data lifecycle (e.g., tools for extraction and curation).
- Specify key performance indicators (KPIs) for technologies that address the different steps of the lifecycle of Big Linked Data [1]. We will reuse well-known KPIs (e.g., F-measure, Precision, Recall, Runtime) for steps for which such exist (e.g., knowledge extraction) and will define KPIs as required by referring to the community.
- We will assemble the infrastructure (hardware and software) necessary to run the evaluation campaigns and publish periodic public reports on the performance of the participants. In particular, we will provide the experimental results in both human-readable and machine-readable formats. We will also provide static URLs for all the experimental results and settings to ensure the reproducibility of the experiments.

- We will provide periodic reports on the performance of state-of-the-art technology that will include automatic diagnostics to suggest possible areas of improvements to vendors.

Our architecture will rely on web interfaces and cloud infrastructures to ensure that we do not restrict or limit the number of participants while scaling up during the tests. The project will aim to create an association after the second project year. The aim of the association will be to be an independent body for the provision of benchmarking results over the linked data lifecycle. The clear portfolio of added value for the members will be defined in the early project stages through collaborations with other European projects and the early inclusion of external experts into the project.

With this project, we aim to provide vendors and academic software developers with a holistic and open benchmarking platform that will allow them to (1) measure their performance against existing frameworks, (2) detect the weaknesses of their tools (automatic diagnostics) on data similar to real data and (3) contribute to a community-driven effort towards self-sustained benchmarking.

Session input and expectations. Within the EU networking session we focus on several points. First, we want to share our knowledge about state-of-the-art benchmarking of linked data systems. Thus, we will discuss with other participants the benchmarks they are using and the features—especially the KPIs—they are interested in but are still missing at the moment. This will support the development of the HOBBIT benchmarks towards useful and widely accepted tools. Moreover, we want to invite other researchers and practitioners to evaluate the architecture of the HOBBIT platform. Finally, we want to identify and collaborate with complementary activities close to the HOBBIT research and development area.

References

1. S. Auer, J. Lehmann, A.-C. Ngonga Ngomo, and A. Zaveri. Introduction to linked data and its lifecycle on the web. In *Reasoning Web. Semantic Technologies for Intelligent Data Access*, pages 1–90. Springer, 2013.
2. M. Cornolti, P. Ferragina, and M. Ciaramita. A framework for benchmarking entity-annotation systems. In *Proceedings of the 22nd international conference on World Wide Web*, pages 249–260. International World Wide Web Conferences Steering Committee, 2013.
3. R. Usbeck, M. Röder, A.-C. Ngonga Ngomo, C. Baron, A. Both, M. Brümmer, D. Ceccarelli, M. Cornolti, D. Cherix, B. Eickmann, et al. Gerbil: general entity annotator benchmarking framework. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1133–1143. International World Wide Web Conferences Steering Committee, 2015.