# SPARQL Querying Benchmarks

Muhammad Saleem, Ricardo Usbeck, Michael Roder, and Axel-Cyrille Ngonga Ngomo

Universität Leipzig, IFI/AKSW, PO 100920, D-04009 Leipzig
{lastname}@informatik.uni-leipzig.de

**Abstract.** Benchmarking is indispensable when aiming to assess technologies with respect to their suitability for given tasks. The importance of SPARQL queries for Linked Data management has led to the development of several benchmarks that allow assessing the performance of SPARQL query processing systems. This tutorial will explore the different benchmarks used for SPARQL query processing over Linked Data. In particular, we will focus on the design, key features, evaluation setup, and pros and cons of the benchmarks used to evaluate the performance of triple stores and federated SPARQL query engines. State-of-the-art triple stores and federated SPARQL benchmarks will be practically demonstrated with examples along with hands-on experience and exercises to be carried out by the participants. By the end of the tutorial, participants will obtain hands-on knowledge of the triple stores and federated SPARQL querying benchmarks, understand the main differences between state-of-the-art systems and benchmarks, and be able to pick the right benchmark based on their pros and cons and the use-case scenario.

## 1 Motivation

Triple stores (also called RDF storage engines) are the backbone of many Linked Data applications [12]. Over the last years, different triple stores (e.g., Virtuoso, Sesame, Fuseki, GraphDB etc.) have been proposed for storing and querying RDF data. Furthermore, due to the decentralized architecture of Linked Data Web, many applications rely on more than one data sources hence need to be able to execute queries over multiple sources at the same time. We call such queries *federated queries* and the systems that execute such queries are called *federated engines*. Consequently, various federated engines and applications have been developed to enable the execution of federated queries on these different data infrastructures [1,24,28,8]. Hence, the performance of triple stores and federated SPARQL querying engines is of central importance for many Linked Data-based software ranging from real-time applications [11,19] to on-the-fly data integration frameworks [1,24,28,8]. The importance of SPARQL queries for Linked Data management has led to the development of several triple stores (e.g.,[2,4,10,12,27,32,22]) and federated SPARQL (e.g., [26,18,9]) benchmarks that allow assessing the performance of SPARQL query processing systems.

The aim of this tutorial is two-fold. First, we aim to provide the participants with an overview of the state of the art in SPARQL querying benchmarks. In particular, we will focus on the design, key features, and will highlight the main shortcomings of these benchmarks. In addition, we aim to provide the participants with practical insights and

hands-on experience that will allow them to configure and run these benchmarks as well as to select the right system for their purposes or even improve upon existing solutions in their future research.

### 1.1 Related Events

Fundulaki and Kementsietsidis are going to present (at ESWC2016) a tutorial[1] assessing the performance of RDF Engines using RDF benchmarks. They focus more on the principles of RDF benchmarks, the dimensions (i.e., dataset, queries, performance metrics, and benchmark evaluation rules) of RDF benchmarks, and a discussion and comparison of the state-of-the-art triple stores and social network benchmarks according to the dimensions of RDF benchmarks. However, one of the key issue is how to properly setup these benchmarks and run the performance test. To this end, complimentary to the previous work, in this tutorial we will be focusing on the practical demonstrations by using running examples and hand-on exercises. In particular, we will introduce IGUANA[2], a generic SPARQL benchmark execution framework. IGUANA provides a generic execution framework to run triple stores and federated benchmarks. It is a configurable framework where one can simulate different stress tests, e.g., how a triple store will respond to concurrent user requests or dynamic data updates. In addition, we will cover SPARQL query federation benchmarks. Note, we will also provide a detailed comparison of the SPARQL querying benchmarks and real query logs as given in Saleem et al. [22].

## 2 Detailed Description

In this section we describe the contents of tutorial, the aims and learning objectives, presentation style and tutorial format, and the prior knowledge required by the attendees.

### 2.1 Content Overview

Our tutorial will consist of three sessions:

– **Triple Store Benchmarks:** In this session, we will briefly describe the SPARQL querying benchmark components (also called dimensions). These include: the dataset(s), the SPARQL queries, the performance metrics, and the rules that should be followed when executing a benchmark. We will then introduce the two main categories of triple store benchmarks namely the synthetic and real-life benchmarks. The synthetic benchmarks [2,4,10,27] either rely on synthetic data or synthetic queries while real benchmarks [22,12] make use of the real data and queries. The key advantages and disadvantages of both of these categories will be discussed. For example, synthetic benchmarks are good to test the scalability of the systems, since they allow generating datasets of virtually any size and queries with varying SPARQL query features, e.g., number of triple patterns, result size, use of SPARQL

[1] *http://www.ics.forth.gr/isl/RDF-Benchmarks-Tutorial/index.html*
[2] *https://github.com/AKSW/IGUANA*

clauses, number of joins, selectivities of triple patterns, joins and so on. However, they often fail to reflect reality [7]. In particular, previous works [7] point out that artificial benchmarks are typically highly structured while real Linked Data sources are less structured. Moreover, synthetic queries most commonly fail to reflect the characteristics of the real queries (i.e., they should show typical requests on the underlying datasets [22,3,13]). We will then compare these benchmarks based on the aforementioned dimensions. In particular, we will show a comparison of the structuredness or coherence [7] of the different benchmark datasets, structural (e.g., number of triple patterns, number of join vertices, number of BGPs, mean join vertices degree etc.) and data-driven (e.g., resultset size, execution time, mean triple pattern selectivities etc.) features of the benchmark queries, and the set of performance metrics, as shown in Saleem et al. [22].
The aim of this session will be to familiarize the audience with the topic and enable them to understand the main differences between the state-of-the-art triple store benchmarks. The insights gained in this session will also be central for the hands-on exercises in the last session.

- **Federated SPARQL Querying Benchmarks:** In this session, we will talk about the federated SPARQL querying benchmarks. They key differences between the triple store benchmarks and federated SPARQL querying benchmarks will be discussed. The variables that may impact the behaviour of federated SPARQL query engines will be discussed [21]. We will compare FedBench [26], a well-known SPARQL federation benchmark with LargeRDFBench [18]. We will also introduce SPLODGE [9], a federated SPARQL benchmark generation framework. In the end, we will give a brief introduction to the state-of-art SPARQL endpoint federation engines [1,24,28,8,31,6].
- **Hands-on Experience and Discussion:** In the last session, we will have a hands-on session where the audience will be provided with a set of practical exercises related to each of the above three sessions. In particular, we show how to configure, load data, and run the different triple stores, e.g., Fuseki, GraphDB, Sesame, Virtuoso etc. Moreover, we will show how to setup the IGUANA benchmark execution framework. We will practically demonstrate running the selected benchmarks on this framework. Audience will be asked to generate customized triple stores benchmarks using the FEASIBLE [22] benchmark generation framework[3]. In the last, we will discuss some of the open problems in SPARQL querying benchmarks. Furthermore, we will introduce the HOBBIT project[4], a Holistic Benchmarking of Big Linked Data.

### 2.2 Aims and Learning Objectives

Our learning objectives are the following:

- Provide basic knowledge about the SPARQL querying benchmarks.
- Elaborate on the main differences between state-of-the-art SPARQL querying benchmarks by using examples and hand-on exercises.

[3] A demo is available at *http://feasible.aksw.org/*.
[4] *http://project-hobbit.eu/*

- Position triple stores, federated SPARQL engines, and the corresponding benchmarks based on their pros and cons.
- Present the setup of triple stores, federated environment for experiments and evaluation in the context of the performance evaluation of the triple stores and federated SPARQL query engines.

### 2.3 Material, Presentation Style and Format

Our presentations will be mostly based on slides, running examples, hands-on exercises and visualization. Basic questions will be asked during the session to keep the audience alert and ensure that the core message is understood. Questions will be allowed throughout the presentations. The first session will be one hour, the second session be will be 30 minutes, and the last session will be one and a half hours. All of the tutorial materials will be available (at least a month before) on the tutorial website as per GPL licence.

### 2.4 Required Prior Knowledge

The audience are required to have basic knowledge of the SPARQL query, Linked Data, and SPARQL endpoints.

## 3 Audience

Given the centrality of triple stores and federated queries for all complex Linked Data applications, we expect that most of the Semantic Web community will be interested in our tutorial and are thus expecting between 20 and 30 attendees (conservative estimate).

## 4 Length

This will be a half day tutorial.

## 5 Technical Requirements

We will only need standard projection equipment. Participants should bring their own laptop.

## 6 Presenters

**Muhammad Saleem**
University of Leipzig, Germany,
saleem@informatik.uni-leipzig.de
Homepage: *https://sites.google.com/site/saleemsweb/*

Muhammad Saleem obtained his Bachelor in Computer Software Engineering from N-W.F.P University of Engineering and Technology and a Master in Computer Science

and Engineering from Hanyang University, South Korea. His research interests include SPARQL benchmarking [22,18,17], federated SPARQL query processing and optimization, Linked Data summaries, personalized query execution, and top-k query processing [20,23,24,25,5]. He already presented a tutorial on federated SPARQL query processing at ISWC 2015[5]. Previously, he was working as research assistant at Digital Enterprise Research Institute (DERI), National University of Ireland, Ireland. He is now a PhD student at the University of Leipzig (Agile Knowledge Engineering and Semantic Web Group) in Germany.

**Ricardo Usbeck**
University of Leipzig, Germany,
usbeck@informatik.uni-leipzig.de
Homepage: *http://aksw.org/RicardoUsbeck* Ricardo first studied at the Martin-Luther-University Halle-Wittenberg where he received his Bachelor and Master degree in computer science. Afterwards, he joined the Unister GmbH, a large e-Commerce vendor, where he studied as an industrial PhD student and later joined the AKSW research group at the Leipzig University. He specialised in information extraction, question answering and benchmarking and now co-leads the DIESEL and QAMEL Eurostars projects. He was organizer of the 2014 SEMANTiCS conference and is an active member of more than a dozen programme committees as well as Chair of the W3C community group for Natural Language Interfaces for the Web of Data.

**Michael Röder**
University of Leipzig, Germany,
roeder@informatik.uni-leipzig.de
Homepage: *http://aksw.org/MichaelRoeder*

Michael Röder obtained his Bachelor of Science as well as his Master of Science from an integrated degree program at the Hochschule Darmstadt in Germany. The industrial partners of these programs were the Deutsche Telekom Training and the Controlware GmbH at which he conducted several projects. He is a PhD student at the University of Leipzig in the Agile Knowledge Engineering and Semantic Web group. His research interests involve Text Mining and Semantic Web technologies with a focus on the evaluation of systems [14,15,16,29,30]. In 2014, he was PC member of the FoRESEE workshop at the Informatik 2014.

**Dr. Axel-Cyrille Ngonga Ngomo**
University of Leipzig, Germany,
ngonga@informatik.uni-leipzig.de
Homepage: *http://aksw.org/AxelNgonga*

Axel Ngonga leads the Agile Knowledge Engineering and Semantic Web research group at the University of Leipzig. His research interests revolve around Semantic Web technologies, especially link discovery, federated queries, machine learning and natural-language processing. Axel (co-)authored more than 100 reviewed publications, has developed/led the development of several widely used frameworks such as LIMES, FOX and GERBIL. In addition, he has received manifold awards including best (student)

[5] tutorial homepage: *https://sites.google.com/site/qfedld/*

research paper awards at CiCLING 2008, ISWC 2011 and 2014 as well as ESWC 2013 and 2014. He has also won several challenges such as the I'Challenge 2013 and the Big Data Challenge at ISWC 2013. Axel has given university lectures on Semantic Web and Information Retrieval/Text Mining.

## References


1. M. Acosta, M.-E. Vidal, T. Lampo, J. Castillo, and E. Ruckhaus. ANAPSID: an adaptive query processing engine for SPARQL endpoints. In *ISWC*, 2011.
2. G. Aluç, O. Hartig, M. T. Özsu, and K. Daudjee. Diversified stress testing of rdf data management systems. In *The Semantic Web–ISWC 2014*, pages 197–212. Springer, 2014.
3. M. Arias, J. D. Fernández, M. A. Martínez-Prieto, and P. de la Fuente. An empirical study of real-world SPARQL queries. *CoRR*, 2011.
4. C. Bizer and A. Schultz. The berlin sparql benchmark. *IJSWIS*, 5(2):1–24, 2009.
5. E. Cem Ozkan, M. Saleem, E. Dogdu, and A.-C. Ngonga Ngomo. Upsp: Unique predicate-based source selection for sparql endpoint federation. In *PROFILES at Extended Semantic Web Conference (ESWC)*, 2016.
6. A. Charalambidis, A. Troumpoukis, and S. Konstantopoulos. Semagrow: Optimizing federated sparql queries. In *Proceedings of the 11th International Conference on Semantic Systems*, SEMANTICS '15, pages 121–128, New York, NY, USA, 2015. ACM.
7. S. Duan, A. Kementsietsidis, K. Srinivas, and O. Udrea. Apples and oranges: a comparison of rdf benchmarks and real rdf datasets. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, pages 145–156. ACM, 2011.
8. O. Görlitz and S. Staab. SPLENDID: SPARQL Endpoint Federation Exploiting VoID Descriptions. In *O. Hartig, A. Harth, and J. F. Sequeda, editors, 2nd International Workshop on Consuming Linked Data (COLD 2011) in CEUR Workshop Proceedings*, volume 782, October 2011.
9. O. Görlitz, M. Thimm, and S. Staab. Splodge: systematic generation of sparql benchmark queries for linked open data. In *The Semantic Web–ISWC 2012*, pages 116–132. Springer, 2012.
10. Y. Guo, Z. Pan, and J. Heflin. LUBM: A benchmark for OWL knowledge base systems. In *Web Semantics: Science, Services and Agents on the World Wide Web*, volume 3, pages 158–182. Elsevier, 2005.
11. M. Kamdar, A. Iqbal, M. Saleem, H. Deus, and S. Decker. Genomesnip: Fragmenting the genomic wheel to augment discovery in cancer research. In *CSHALS*, 2014.
12. M. Morsey, J. Lehmann, S. Auer, and A.-C. Ngonga Ngomo. Dbpedia sparql benchmark - performance assessment with real queries on real data. In *International Semantic Web Conference*, pages 454–469, 2011.
13. F. Picalausa and S. Vansummeren. What are real sparql queries like? In *Proceedings of the International Workshop on Semantic Web Information Management*, page 7. ACM, 2011.
14. M. Röder, A. Both, and A. Hinneburg. Exploring the space of topic coherence measures. In *Proceedings of the eight International Conference on Web Search and Data Mining, Shanghai, February 2-6*, 2015.
15. M. Röder, M. Speicher, and R. Usbeck. Investigating quality raters' performance using interface evaluation methods. In M. Horbach, editor, *Informatik 2013, 43. Jahrestagung der Gesellschaft für Informatik e.V. (GI), Informatik angepasst an Mensch, Organisation und Umwelt, 16.-20. September 2013, Koblenz*, pages 137–139. GI, 2013.
16. M. Röder, R. Usbeck, and A.-C. Ngonga Ngomo. Developing a sustainable platform for entity annotation benchmarks. In *ESWC Developers Workshop 2015*, 2015.

17. M. Saleem, M. I. Ali, A. Hogan, Q. Mehmood, and A.-C. N. Ngomo. Lsq: The linked sparql queries dataset. In *The Semantic Web-ISWC 2015*, pages 261–269. Springer, 2015.
18. M. Saleem, A. Hasnain, and A.-C. Ngonga Ngomo. LargeRDFBench: a billion triples benchmark for sparql endpoint federation. In *Submitted to Web Semantics: Science, Services and Agents on the World Wide Web*. Elsevier, 2016.
19. M. Saleem, M. R. Kamdar, A. Iqbal, S. Sampath, H. F. Deus, and A.-C. Ngonga Ngomo. Big linked cancer data: Integrating linked tcga and pubmed. *JWS*, 2014.
20. M. Saleem, Y. Khan, A. Hasnain, I. Ermilov, and A.-C. N. Ngomo. A fine-grained evaluation of sparql endpoint federation systems. *SWJ*, 2014.
21. M. Saleem, Y. Khan, A. Hasnain, I. Ermilov, and A.-C. Ngonga Ngomo. A fine-grained evaluation of sparql endpoint federation systems. *Semantic Web*, pages 1–26, 2015.
22. M. Saleem, Q. Mehmood, and A.-C. N. Ngomo. Feasible: A feature-based sparql benchmark generation framework. In *The Semantic Web-ISWC 2015*, pages 52–69. Springer, 2015.
23. M. Saleem and A.-C. Ngonga Ngomo. HiBISCuS: Hypergraph-Based Source Selection for SPARQL Endpoint Federation. In V. Presutti, C. d'Amato, F. Gandon, M. d'Aquin, S. Staab, and A. Tordai, editors, *The Semantic Web: Trends and Challenges*, volume 8465 of *Lecture Notes in Computer Science*, pages 176–191. Springer International Publishing, 2014.
24. M. Saleem, A.-C. Ngonga Ngomo, J. Xavier Parreira, H. Deus, and M. Hauswirth. DAW: Duplicate-AWare Federated Query Processing over the Web of Data. In H. Alani, L. Kagal, A. Fokoue, P. Groth, C. Biemann, J. Parreira, L. Aroyo, N. Noy, C. Welty, and K. Janowicz, editors, *The Semantic Web – ISWC 2013*, volume 8218 of *Lecture Notes in Computer Science*, pages 574–590. Springer Berlin Heidelberg, 2013.
25. M. Saleem, S. S. Padmanabhuni, A.-C. N. Ngomo, A. Iqbal, J. S. Almeida, S. Decker, and H. F. Deus. Topfed: Tcga tailored federated query processing and linking to lod. *Journal of biomedical semantics*, 5(1):1, 2014.
26. M. Schmidt, O. Görlitz, P. Haase, G. Ladwig, A. Schwarte, and T. Tran. Fedbench: a benchmark suite for federated semantic data query processing. In *ISWC*, 2011.
27. M. Schmidt, T. Hornung, G. Lausen, and C. Pinkel. Spˆ 2bench: a sparql performance benchmark. In *ICDE*, pages 222–233, 2009.
28. A. Schwarte, P. Haase, K. Hose, R. Schenkel, and M. Schmidt. Fedx: Optimization techniques for federated query processing on linked data. In *ISWC*, 2011.
29. R. Usbeck, M. Röder, and A.-C. Ngonga Ngomo. Evaluating entity annotators using gerbil. In *Proceedings of 12th European Semantic Web Conference (ESWC 2015)*, 2015.
30. R. Usbeck, M. Röder, A.-C. Ngonga Ngomo, C. Baron, A. Both, M. Brümmer, D. Ceccarelli, M. Cornolti, D. Cherix, B. Eickmann, P. Ferragina, C. Lemke, A. Moro, R. Navigli, F. Piccinno, G. Rizzo, H. Sack, R. Speck, R. Troncy, J. Waitelonis, and L. Wesemann. GERBIL – general entity annotation benchmark framework. In *24th WWW conference*, 2015.
31. X. Wang, T. Tiropanis, and H. C. Davis. Lhd: Optimising linked data query processing using parallelisation. In *LDOW at WWW*, 2013.
32. H. Wu, T. Fujiwara, Y. Yamamoto, J. Bolleman, and A. Yamaguchi. Biobenchmark toyama 2012: an evaluation of the performance of triple stores on biological data. *Journal of biomedical semantics*, 5(1):1, 2014.