



HOBBIT

Newsletter
November 2016

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 Community Member,

We are proudly presenting the results of HOBBIT's first Survey. The Survey was developed and subsequently distributed to the HOBBIT community in order to gather the requirements for the HOBBIT platform and benchmarks. The aim of the fully anonymized Survey was for the HOBBIT consortium to determine how a participant's company or organization could benefit from the HOBBIT platform. The Survey ended end of June 2016, while some [Preliminary Survey Results](#) were presented earlier in mid June.

HOBBIT 1st Survey description and results

The participants were classified in the following categories:

- Solution Providers
- Technology Users
- Scientific Community

Solution Providers were asked whether their software under development could have its performance measured using a benchmark, which parts of the Linked Data lifecycle the provided solution belongs to and if they could provide a brief description of the provided services. We asked Technology Users whether they use a third party software to solve a problem, which parts of the Linked Data lifecycle they outsource and finally to provide a brief description of the use cases of interest. Regarding the Scientific Community we wanted to know the part of the Linked Data lifecycle they are working on, the Key Performance Indicators they are concerned for their solution and whether they could provide a brief description thereof.

61 people (representing their organization) participated in the Survey: 48 solution providers, 46 technology users and finally 47 people from the scientific community (note that participants may be associated with multiple profiles).

The majority of Solution Providers (80.3%) are developing software that could have its performance measured using a benchmark, while 80% of Technology Users are using 3rd party software for solving a problem. Finally, members of the Scientific Community by a very large majority (84.5%) are working on a solution that is part of the Linked Data Lifecycle.

The table below shows the answers of the participants regarding the type of Linked Data Solutions they are interested in when asked one of the following questions:

- "which parts of the Linked Data lifecycle the provided solution belongs to" (Solutions Provider),
- "which parts of the Linked Data lifecycle they outsource" (Technology User) and
- "which part of the Linked Data lifecycle they are working on" (Scientific Community).

	Solution Provider	Technology User	Scientific Community
Storage/Querying	36	32	27
Interlinking	24	25	19
Classification/Enrichment	23	27	24
Discovery	21	21	23
Extraction	18	20	17
Reasoning	18	15	15
Other	2	3	4

From the results of the Survey we can note that

- Solution Providers are very interested in Storage and Querying solutions, equally but less interested in Interlinking, Classification/Enrichment and Discovery. Reasoning and Extraction seem to interest them significantly but less than the aforementioned solutions,
- Technology Users are mostly interested in Storage/Querying, followed by Classification/Enrichment and then Interlinking. They are equally but less interested in Discovery, Extraction and Reasoning and
- members of the Scientific Community are very interested in Storage/Querying, followed by Classification/Enrichment and Discovery. Interlinking, Extraction, and Reasoning follow.

Participants were also asked the Key Performance Indicators of interest and provided the following answers in order of importance: Correctness, Accuracy (precision, recall, F-measure, mean reciprocal rank) followed by Runtime / Speed, Total triple pattern-wise sources selected and Number of intermediate results. Scalability, Memory and CPU usage are also important KPIs for the audience.

More importantly, participants were asked whether they are using an existing benchmark to evaluate their solutions. 50.8% of the participants responded positively to this question. The majority of the participants benchmark their storage/querying solutions. Benchmarks for Classification/Enrichment and Extraction tasks are also broadly used, followed by the remaining benchmark categories.

An interesting point here is that 69% of the datasets used in benchmarks are later made available to the public. Furthermore, the majority of the datasets used are fairly large with close to 40% containing 1M to 100M triples and 35% contains more than 100M.

What comes next ?

A new Survey is now under preparation and will be announced in the following month. Stay tuned and take part in it !

Join our community to remain informed at <http://project-hobbit.eu/get-involved/>. If you have any questions, do not hesitate to contact us (see contact details at <http://project-hobbit.eu/contacts/>). We are looking forward to benchmarking your system soon.

Best,
HOBBIT team.



Share



Tweet



Forward