

Pressemeldung Januar 2018

HOBBIT

Holistic Benchmarking of
Big Linked Data

H2020 Forschung und Innovation
Fördernummer: 688227

Webseite

<http://project-hobbit.eu>

Kontakt

Prof. Dr. Axel-Cyrille Ngonga Ngomo
Institut für Angewandte Informatik
Hainstraße 11, 04109 Leipzig
Germany
Telefon.: +49 341 97 32362
ngonga@infai.org



HOBBIT

Holistic Benchmarking
of Big Linked Data

HOBBIT feiert Jahr 2!

Wir freuen uns sehr, mitteilen zu können, dass HOBBIT nun zwei Jahre alt ist! Die Mitglieder des Projekt-Konsortiums waren in diesem Jahr sehr aktiv und arbeiteten intensiv an den wichtigsten Ergebnissen des Projekts: *der Benchmark Entwicklung, der Bereitstellung der HOBBIT Plattform und der Challenge-Organisation!*

Wichtiger noch, HOBBIT lieferte die erste Version seiner Benchmarks für *Generierung & Akquisition, Analyse & Verarbeitung, Lagerung & Kuration sowie Visualisierung & Dienstleistungen*. Im Einzelnen:

- **Benchmarks für Generierung & Akquisition** messen die Leistungsfähigkeit von SPARQL Abfrageverarbeitungssystemen im Vergleich zu Datenströmen aus Industriemaschinen hinsichtlich Effizienz und Vollständigkeit. Um die realen Belastungen von Triple Stores, die in realen Anwendungen zum Einsatz kommen, abzubilden, wurden Public Transport, Twitter, Traffic und Sensordaten aus den Kunststoffspritzgussanlagen von Weidmüller gesammelt. In diesem Zusammenhang liefert HOBBIT Benchmarks zur Messung der Leistung von Extraktionssystemen für unstrukturierte Datenströme natürlichsprachlicher Daten. Bei den Benchmarks haben wir die realen, kuratierten unstrukturierten Datensätze von Experten und die unstrukturierten Datenströme von Bengal, einem generischen Datengenerator berücksichtigt.
- **Benchmarks für Analyse- & Verarbeitung** konzentrieren sich auf die Überprüfung der Leistungsfähigkeit von Link Discovery-Systemen und maschinellen Lernmethoden (überwacht und unbeaufsichtigt) zur Datenanalyse. Genauer gesagt, können die in HOBBIT entwickelten Link Discovery Benchmarks verwendet werden, um die Leistungsfähigkeit von (a) Instance Matching Tools zu testen, die string-basierte Ansätze zur Identifizierung von Matching Entities implementieren, und (b) Systemen, die sich mit topologischen Relationen befassen, die im hochmodernen DE-9IM-Modell (Dimensionally Extended Nine-Intersection Model) vorgeschlagen werden.
- **Benchmarks für Lagerung & Kuration** zielen darauf ab, die Leistungsfähigkeit von Datenspeicherungs- und Versionierungssystemen für Linked Data zu testen. Der Datenspeicher-Benchmark konzentriert sich auf die typischen Herausforderungen, denen sich Linked Data Storage Systeme gegenüberstehen, und basiert auf dem Social Network Benchmark (SNB), der im Rahmen des EU FP7 LDBC Projekts entwickelt wurde. Der Versionierungs-Benchmark von HOBBIT zielt darauf ab, die Fähigkeit von Versionierungssystemen zu testen, sich entwickelnde Linked Data-Datensätze und Abfragen, die über mehrere Versionen solcher Datensätze hinweg ausgewertet werden, effizient zu verwalten; er erweitert





den LDBC Semantic Publishing Benchmark (SPB), inspiriert von der Publishing-Domäne.

- Die **Benchmarks von Visualisierung & Dienstleistungen** zielen darauf ab, die Leistung von Abfragebeantwortungs- und Facetten-Browser-Systemen für verknüpfte Daten zu testen, ohne die Benutzer einzubeziehen. Die entwickelten Benchmarks sind nicht dazu gedacht, Benutzeroberflächen zu testen, sondern konzentrieren sich auf die Bereitstellung von Leistungs- und Genauigkeitsmessungen für Ansätze, die in solchen Schnittstellen verwendet werden. Für die Benchmarks wurden Browsing-Szenarien entwickelt, die einen authentischen Anwendungsfall widerspiegeln und die teilnehmenden Systeme in verschiedenen Schwierigkeitsgraden herausfordern.

Alle oben genannten Benchmarks sind auf der HOBBIT-Website CKAN: <https://ckan.project-hobbit.eu/dataset> verfügbar, zusammen mit dem Quellcode und den dazugehörigen Publikationen.



Die HOBBIT-Evaluierungsplattform wurde zu Beginn des zweiten Projektjahres veröffentlicht. IEs handelt sich um eine verteilte FAIR-Benchmarking-Plattform für den Linked Data Lifecycle, die Open Source ist und lokal heruntergeladen und ausgeführt werden kann. Auf die HOBBIT-Plattform kann über ihre Online-Instanz zugegriffen werden, die genutzt werden kann für a) die Durchführung öffentlicher Challenges und b) die Sicherstellung, dass auch Personen ohne die erforderliche Infrastruktur in der Lage sind, die Benchmarks, an denen sie interessiert sind, auszuführen.



Die Online-Instanz der HOBBIT-Benchmarking-Plattform ist unter master.project-hobbit.eu und ihr Code unter <https://github.com/hobbit-project> zugänglich.



Die entwickelten Benchmarks und Plattform wurden in den vom HOBBIT-Projekt organisierten Challenges umfassend genutzt: Mighty Storage (MOCHA), Query Answering over Linked Data (QALD), Open Knowledge Extraction (OKE) und DEBS Grand Challenge. Die MOCHA-, QALD- und OKE-Challenges wurden im Rahmen des ESWC 2017 organisiert, während die DEBS Grand Challenge im Rahmen der DEBS 2017-Konferenz stattfand.



Zusätzlich zu den oben Genannten organisierte HOBBIT die QALD-8 Challenge in Verbindung mit der ISWC 2017, die in den Workshop Natural Language Interfaces for Web of Data (NLIWoD) integriert wurde. HOBBIT hat auch einen neuen Track auf dem Workshop Ontology Matching (OM) 2017 vorgeschlagen und mitorganisiert, der unter der Schirmherrschaft von OAEI durchgeführt wurde; der Workshop fand in Verbindung mit der ISWC 2017 statt. Nach diesem erfolgreichen Unterfangen beschlossen die OM-Organisatoren und die Mitglieder des HOBBIT-Konsortiums, die SEALS-Plattform, die seit einigen Jahren für den Betrieb der OM-Benchmarks genutzt wird, durch die HOBBIT-Plattform zu ersetzen. Die Kampagne OAEI 2017.5, die in Verbindung mit ESWC 2018 durchgeführt wird, zielt auf Ontologie-Matcher ab, die ihre Systeme ausschließlich mit der HOBBIT-Plattform vergleichen. Weitere Informationen zur Kampagne finden Sie hier: <http://oaei.ontologymatching.org/2017.5>.

Ziel der Mighty Storage Challenge (MOCHA) war es, die Leistungsfähigkeit von Lösungen für die SPARQL-Verarbeitung in Aspekten zu testen, die für moderne Anwendungen relevant sind. Dazu gehören das Einlesen von Daten, das

Erfahren Sie mehr über unsere Challenges



und unsere Offenen Challenges



Werden Sie Teil der HOBBIT Community



Beantworten von Anfragen zu großen Datensätzen und die Funktion als Backend für Anwendungen, die von Linked Data gesteuert werden. Drei Systeme nahmen an den MOCHA-Aufgaben teil: (a) Virtuoso Open-Source Edition 7.2, entwickelt von OpenLink Software, das als Basissystem für alle MOCHA 2017-Aufgaben (MOCHA Baseline) diente, (b) QUAD, entwickelt von Ontos, und (c) Virtuoso Commercial Edition 8.0 (beta), entwickelt von OpenLink Software.

Die Challenge "Question Answering over Linked Data" (QALD) zielte darauf ab, einen aktuellen Benchmark für die Bewertung und den Vergleich von State-of-the-Art-Systemen zur Verfügung zu stellen, die zwischen einem Benutzer vermitteln und seinen Informationsbedarf in natürlicher Sprache und RDF-Daten ausdrücken. Die Challenge war in vier verschiedene Aufgaben gegliedert: Mehrsprachige Fragenbeantwortung über DBpedia, hybride Fragenbeantwortung, groß angelegte Fragenbeantwortung über RDF und schließlich Fragenbeantwortung über Wikidata. Drei Systeme waren an der Aufgabe beteiligt: WDAqua, AMAL und ganswer.

Die Open Knowledge Extraction (OKE) Challenge gliederte sich in vier verschiedene Aufgabenbereiche (a) Identifizieren und Verknüpfen gezielt benannter Entitäten, (b) Identifizieren und Verknüpfen von Entitäten mit breiter gefassten Namen (c) Erkennung und Verknüpfung von gezielten Musicals und schließlich (d) Wissensextraktion. Ziel der OKE Challenge ist es, die Leistungsfähigkeit von Knowledge Extraction Systemen im Hinblick auf das Semantic Web zu testen. Adel und Fox Systeme nahmen an der Evaluierung der OKE Challenge teil.

Schließlich lag der Schwerpunkt der DEBS 2017 Grand Challenge auf der Analyse der RDF-Streaming-Daten, die von digitalen und analogen Sensoren erzeugt werden, die in Fertigungsanlagen eingebettet sind. Ziel der Challenge war es, die Erkennung von Anomalien im Verhalten solcher Fertigungsanlagen zu implementieren. Die Herausforderung wurde gemeinsam mit AGT International im Auftrag des HOBBIT-Projekts umgesetzt.

Last but not least lancierte HOBBIT eine Reihe von Open Challenges, nämlich die OKE, MOCHA, SQA (Scalable Question Answering) und STREAML (Stream Machine Learning) Challenges. OKE und MOCHA orientieren sich an den Herausforderungen, die im Rahmen des ESWC 2017 gestartet wurden und derzeit laufen. Die Hauptaufgabe von SQA Open Challenge besteht darin, die Systeme herauszufordern, die Informationsanfrage eines Benutzers in eine solche Form zu übersetzen, dass sie mit Hilfe von Standard-Semantic Web-Abfrageverarbeitungs- und Inferenzierungstechniken effizient ausgewertet werden kann. Die StreamML Open Challenge konzentriert sich auf die Aufgabenstellung im Zusammenhang mit dem Problem der automatischen Erkennung von Anomalien in Fertigungsanlagen.

Zusätzlich zu diesen Herausforderungen, bei denen es vor allem darum geht, Systeme an ihre Grenzen zu bringen, organisiert HOBBIT eine offene Ausschreibung für Benchmarks, die in die Plattform integriert werden, so dass eine große Anzahl von Menschen ihre Systeme auf einer standardisierten Hardware mit vergleichbaren Ergebnissen vergleichen kann.

Informationen zum HOBBIT-Projekt, zur Plattform, zu den Benchmarks und den Challenges finden Sie hier: <https://project-hobbit.eu> .