

Komunikat prasowy Styczeń 2018

HOBBIT
Holistic Benchmarking of
Big Linked Data

H2020 Badania naukowe i innowacja
Grant 688227

Strona internetowa
<http://project-hobbit.eu>

Kontakt

Dr. Axel-Cyrille Ngonga Ngomo
Institute for Applied Informatics
Hainstraße 11, 04109 Lipsk
Niemcy
Nr telefonu: +49 341 97 32362
ngonga@informatik.uni-leipzig.de



HOBBIT
Holistic Benchmarking
of Big Linked Data

HOBBIT świętuje swój drugi rok!

Możemy z radością ogłosić, że HOBBIT ma już dwa lata! Członkowie konsorcjum byli bardzo aktywni w tym roku i pracowali nad głównymi składowymi projektu: stworzeniem benchmarku, wdrożeniem platformy Hobbit i organizacją wyzwań!

Co ważniejsze, HOBBIT dostarczył pierwszą wersję benchmarków: *Generation & Acquisition, Analysis & Processing, Storage & Curation Visualization & Services*. Dokładny opis dostarczonych benchmarków:

- **Generation & Acquisition benchmark** mierzy wydajność systemów procesujących zapytania *SPARQL* podczas przetwarzania strumieni danych pochodzących z urzędzeń przemysłowych. Badane wskaźniki to skuteczność i kompletność. Aby odwzorować prawdziwe obciążenie danymi używanymi w prawdziwych aplikacjach, zebrano i wykorzystano dane o transporcie zbiorowym, dane Twittera, dane o ruchu drogowym i dane o wtryskiwaniu tworzyw sztucznych w przemyśle. HOBBIT dostarcza benchmarku do mierzenia wydajności systemów ekstrakcyjnych dla nieustrukturyzowanych strumieni danych opisanych językiem naturalnym. Wzięto pod uwagę zarówno dane prawdziwe (poddane analizie przez ekspertów) jak i strumienie danych pochodzące z generatorów danych generycznych takich jak Bengal.
- **Analysis & Processing benchmark** skupia się na testowaniu wydajności *link discovery systems* i metod nauczania maszynowego (*nadzorowanego i nienadzorowanego*) dla analiz danych. Dokładniej, benchmark utworzony w HOBBIT może zostać użyty do przetestowania wydajności: (a) narzędzi służących do dopasowywania instancji, które implementują metody łańcuchowe (string) w celu identyfikacji pasujących encji i (b) systemów, które mają do czynienia z relacjami topologicznymi zaproponowanymi w modelu *DE-9IM* (Dimensionally Extended nine-Intersection Model). Benchmark analizujący jest używany do oceny wydajności i efektywności nauczania maszynowego na danych ustrukturyzowanych.
- **Storage & Curation benchmark** służy do przetestowania wydajności przechowywania danych i systemów wersjonowania dla danych powiązanych (Linked Data). Benchmark skupia się na typowych wyzwaniach napotykanym przy przechowywaniu danych powiązanych i jest oparty o Social Network Benchmark (SNB) stworzony w kontekście projektu EU FP7 LDDB. Benchmark wersjonujący HOBBIT ma na celu sprawdzenie zdolności systemów wersjonujących do skutecznego zarządzania ewoluującymi danymi

 **Fraunhofer**
IAIS

 **InfAI**[®]
Institut für Angewandte Informatik

 **imec**

 **FORTH**



połączonymi i zapytaniami uruchamianymi na wielu wersjach takich zbiorów danych; rozszerza on LDBC Semantic Publishing Benchmark (SPB), który wzięł inspirację z domeny wydawniczej.

- **Visualization & Services benchmarks** ma na celu sprawdzenie wydajności odpowiedzi na zapytania i systemów do przeglądania danych powiązanych (Linked data) bez angażowania użytkowników. Stworzony benchmark nie ma na celu przetestowania interfejsu użytkownika, ale skupia się na dostarczeniu miar związanych z wydajnością i trafnością w tych interfejsach. Dla benchmarku, scenariusze przeglądania, które odwzorowują prawdziwe użycie systemów, które zostały zbadane.

Wszystkie powyższe benchmarki są dostępne na HOBBIT CKAN: <https://ckan.project-hobbit.eu/dataset>, wraz z ich kodem źródłowym i powiązаныmi publikacjami.

Platforma ewaluacyjna HOBBIT została uruchomiona na początku drugiego roku trwania projektu. Jest to rozdystrybuowana platforma benchmarkująca FAIR dla danych powiązanych (linked data), której źródła są open source i która może zostać pobrana i uruchomiona lokalnie. Platforma HOBBIT może być również użyta online do a) uruchomienia publicznych wyzwań i b) upewnienia się, że nawet ludzie nieposiadający odpowiedniej infrastruktury mogą uruchomić interesujący dla nich benchmark.

Wersja online platformy HOBBIT jest dostępna pod adresem master.project-hobbit.eu a jej kod źródłowy znajduje się na GitHubie: <https://github.com/hobbit-project>.

Utworzone benchmarki i platforma zostały obszernie użyte podczas wyzwań zorganizowanych przez projekt Hobbit: the Mighty Storage (MOCHA), Query Answering over Linked Data (QALD), Open Knowledge Extraction (OKE) i DEBS Grand Challenge. Wyzwania MOCHA, QALD i OKE zostały zorganizowane w ramach ESWC 2017, podczas gdy DEBS Grand Challenge miał miejsce podczas konferencji DEBS 2017.

Dodatkowo, HOBBIT zorganizował QALD-8 Challenge w połączeniu z ISWC 2017, które było częścią warsztatu Natural Language Interfaces for Web of Data (NLIWoD). HOBBIT zaproponował również i współorganizował nową ścieżkę na warsztacie Ontology Matching (OM) 2017, który był prowadzony pod patronatem OAEI; warsztat był prowadzony w połączeniu z ISWC 2017. Po tym pomyślnym wysiłku, organizatorzy OM i członkowie konsorcjum HOBBIT zdecydowali aby dążyć ku zastąpieniu platformy SEALS, która była używana przez długie lata do uruchamiania benchmarków OM, przez platformę HOBBIT. Kampania OAEI 2017.5, która będzie miała miejsce w połączeniu z ESWC 2018, skupia się na wykorzystaniu benchmarków wyłącznie używając platformy HOBBIT przy łączeniu ontologii. Więcej informacji o kampanii można znaleźć tutaj: <http://oaei.ontologymatching.org/2017.5>

Celem Mighty Storage Challenge (MOCHA) było sprawdzenie rozwiązań dla procesowania SPARQL w aspektach ważnych dla dzisiejszych aplikacji. Między innymi było to: dodawanie danych, odpowiedzi na zapytania na dużych zbiorach danych i służyć jako backend dla aplikacji pracujących z Linked data. Trzy systemy uczestniczyły w zadaniach MOCHA: (a) Virtuoso



Dowiedz się więcej o naszych wyzwaniach:



Open-Source Edition 7.2, stworzony przez OpenLink Software, który został użyty jako punkt odniesienia dla wszystkich zadań MOCHA 2017 (MOCHA Baseline), (b) QUAD, stworzony przez Ontos, i (c) Virtuoso Commercial Edition 8.0 (beta), stworzony przez OpenLink Software.

Wyzwanie Question Answering over Linked Data (QALD) skupiło się na dostarczeniu aktualnego benchmarku do oceny i porównania nowoczesnych systemów, które pośredniczą między użytkownikiem, który wyraża swoje potrzeby informacyjne w języku naturalnym, a danymi RDF. Wyzwanie było podzielone na 4 różne zadania: odpowiedzi na wielojęzyczne zapytanie na DBpedia, odpowiedzi hybrydowe na zapytanie, odpowiedzi na zapytania o wielkiej skali przy użyciu RDF i odpowiedzi na zapytania poprzez Wikidata. Trzy systemy partycypowały w zadaniu: WDAqua, AMAL i ganswer.

i o naszych otwartych wyzwaniach:



Wzywanie Open Knowledge Extraction (OKE) było zorganizowane w 4 zadania: (a) Focused Named Entity Identification and Linking, (b) Broader Named Entity Identification and Linking (c) Focused Musical Named Entity Recognition and Linking i (d) Knowledge Extraction. Celem OKE jest sprawdzenie wydajności systemów wydobywania wiedzy (punkt d) w powiązaniu z Semantic Web. Systemy Adel i Fox uczestniczyły w wyzwaniu OKE.

Celem DEBS 2017 Grand Challenge była analiza strumienia danych RDF wygenerowanego przez cyfrowe i analogowe czujniki będące częścią sprzętu wytwórczego. Celem tego wyzwania była implementacja wykrywania anomalii w zachowaniu takiego sprzętu. Wyzwanie było współorganizowane przez AGT International w imieniu projektu HOBBIT.

Dołącz do społeczności HOBBIT:



Nie mniej ważną informacją jest uruchomienie w ramach projektu HOBBIT zestawu otwartych wyzwań (*Open Challenges*), takich jak OKE, MOCHA, SQA (Scalable Question Answering) i STREAML (Stream Machine Learning). OKE i MOCHA zostały oparte o wyzwania, które zostały uruchomione w kontekście ESWC 2017 i obecnie trwają. Głównym zadaniem SQA Open Challenge jest sprawdzenie systemów podczas tłumaczenia żądania użytkownika o informacje w formę, która może zostać efektywnie oceniona używając standardowych zapytań używanych w Semantic Web. StreamML Open Challenge skupia się na zadaniach powiązanych z problemami automatycznej detekcji anomalii wśród maszyn wytwórczych.

Dodatkowo, poza uruchomieniem wyzwań, które głównie skupiają się na pchaniu systemów do granic ich możliwości przetwarzania, HOBBIT wystosował otwarte zaproszenie dla benchmarków, które będą częścią platformy, dzięki czemu duża liczba ludzi może użyć benchmarków na swoich systemach na zestandaryzowanym sprzęcie, otrzymując w ten sposób porównywalne rezultaty.

Informacje o projekcie HOBBIT, platformie, benchmarkach i wyzwaniach można znaleźć tutaj: <https://project-hobbit.eu>