

Слухай Яна Олександрівна

магістр комп'ютерних наук

Національного технічного університету України

«Київський політехнічний інститут імені Ігоря Сікорського»

Слухай Яна Александровна

магістр компьютерных наук

Национального технического университета Украины

«Киевский политехнический институт имени Игоря Сикорского»

Slukhai Yana

Master of Computer Science of the

National Technical University of Ukraine

«Igor Sikorsky Kyiv Polytechnic Institute»

**ПОРІВНЯЛЬНИЙ АНАЛІЗ ПРОГРАМНИХ ЗАСОБІВ
ДЛЯ ПОБУДОВИ БАЗИ ЗНАНЬ
СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПРОГРАММНЫХ СРЕДСТВ
ДЛЯ ПОСТРОЕНИЯ БАЗЫ ЗНАНИЙ
THE COMPARATIVE ANALYSIS OF SOFTWARE
FOR BUILDING A KNOWLEDGE BASE**

Анотація. Проведено порівняльний аналіз архітектурних особливостей таких програмних засобів як Apache Jena – TDB, Dydra, Sesame, GraphDB, RDFLib.

Ключові слова: семантичний веб, Apache Jena – TDB, Dydra, Sesame, GraphDB, RDFLib, сховище триплетів.

Аннотация. Проведен сравнительный анализ архитектурных особенностей таких программных средств как Apache Jena – TDB, Dydra, Sesame, GraphDB, RDFLib.

Ключевые слова: семантический веб, Apache Jena – TDB, Dydra, Sesame, GraphDB, RDFLib, хранилище триплетов.

Summary. The comparative analysis of the architectural features of such software tools has been performed: Apache Jena – TDB, Dydra, Sesame, GraphDB, RDFLib

Key words: semantic web, Apache Jena – TDB, Dydra, Sesame, GraphDB, RDFLib, triplestore.

Сховище триплетів або RDF-сховище є базою знань для зберігання та вилучення триплетів через семантичні запити. Триплет — це об'єкт даних, що складається з суб'єкта, об'єкта і предиката, наприклад «Боб любить спагеті» або «Боб знає Фреда».

На відміну від реляційної бази даних, сховище триплетів оптимізоване для зберігання та вилучення триплетів. Окрім запитів, зазвичай можна імпортувати / експортувати триплети, використовуючи структуру опису ресурсів (RDF) та інші формати [2, с. 35–37].

Деякі сховища триплетів були побудовані як бази знань з нуля, в той час як інші були побудовані на основі існуючих комерційних реляційних баз даних (наприклад, на базі SQL) або на основі

баз даних, орієнтованих на документи (NoSQL). Подібно ранньому розвитку баз даних аналітичної онлайн обробки (OLAP), цей проміжний підхід дозволив побудувати великі та потужні бази знань, що потребували невеликих зусиль для програмування на початкових етапах розробки сховищ триплетів. Найбільш ймовірно, що нативні сховища триплетів матимуть перевагу для роботи протягом більш тривалого періоду часу.

Труднощі з реалізацією сховищ триплетів за допомогою SQL полягають в тому, що, хоча триплети таким чином, можуть бути «збережені», але реалізація ефективного запиту до графічної RDF-моделі (наприклад, відображення з SPARQL) на SQL-запити є досить складною [2, с. 36].

Apache Jena — TDB

Apache Jena — це Java-фреймворк для побудови семантичних веб-додатків. Він забезпечує програмне середовище для RDF, RDFS і OWL, SPARQL, і включає в себе механізм побудови висновків на основі правил (підтримка роботи ризонерів).

Apache Jena TDB надає легке, масштабовне, транзакційне сховище та можливість виконання SPARQL запитів.

Сховище TDB може бути доступним та керованим за допомогою скриптів командного рядка та через Jena API. При доступі за допомогою транзакцій TDB набір даних захищений від пошкоджень, несподіваних завершень процесу та системних збоїв [1].

Якщо є необхідність поділитися набором даних TDB між кількома додатками, можна використовувати компонент Fuseki, який надає сервер SPARQL. Fuseki може використовувати TDB для постійного зберігання, і надає протоколи SPARQL для запитів та оновлення згідно REST через HTTP.

Dydra

Завдяки Dydra можна отримати доступ до своїх даних і оновлювати їх за допомогою стандартної мови запитів, спеціально розробленої для обробки графів SPARQL.

Користувачі можуть виконувати більшість функцій облікового запису на Dydra.com за допомогою стандартного REST API [4]. Деякі з функцій включають: читання та оновлення інформації про обліковий запис; створення сховища; завантаження в сховище в різних форматах RDF. Відповіді надаються у форматі XML або JSON.

Sesame

Sesame — це сховище RDF із відкритим вихідним кодом, що підтримує логічне виведення та запити RDFS. Він пропонує великий набір інструментів для розробників, щоб використовувати можливості RDF і RDF Schema. Це Java-фреймворк, що також має Python-інтерфейс [5, с. 247].

Окрім того, що він пропонує використовувати нативну базу даних, фреймворк також надає розширений API для створення інших сховищ. Є багато сторонніх сховищ, які можна використовувати за допомогою Sesame Sail API. Прикладами є Bigdata і OWLIM (яка пропонує додаткову підтримку OWL виведення) [4].

GraphDB

GraphDB (колишній OWLIM) — це масштабовний семантичний репозиторій. Він включає в себе сховище триплетів, механізм висновків та обробник SPARQL запитів. GraphDB використовує основу TRREE (Triple Reasoning and Rule Entailment Engine) для виконання RDFS, OWL DLP, OWL Horst виведення та OWL 2 RL [3, с. 125–128]. Мови, що підтримуються включають OWL 2 RL, RDFS. GraphDB пропонує масштабовну підтримку логічного виведення. Доступні як звичайна версія продукту так і хмарна.

RDFLib

RDFLib — бібліотека на мові Python для роботи з RDF, потужною мовою для представлення інформації. Бібліотека містить аналізатор та серіалізатор RDF / XML, який відповідає специфікації синтаксису RDF / XML. У бібліотеці також наявні

Таблиця 1

Порівняльний аналіз розглянутих програмних засобів

| Назва програмного засобу | Apache Jena — TDB | Dydra | Sesame | GraphDB | RDFLib |
|---|--|-----------------------------------|-------------------|---|---------------|
| Нативні програмні API | Java | REST API | Java | Java (Jena and RDF4J (Sesame)) | Python |
| Нативна підтримка SPARQL | + | + | + | + | + |
| Нативна підтримка SPARQL оновлення | + | + | + | + | + |
| Мова імплементації | Java | Common Lisp, C++ | Java | Java | Python |
| Можливість розгортання у хмарі | + | – (доступ як до сервісу) | + | + | + |
| Нативна підтримка постійного збереження даних | Сховище триплетів | Графова база даних у хмарі (SaaS) | Сховище триплетів | Сховище триплетів | Berkeley DB |
| Підтримка роботи з OWL | + | – | – | + | + |
| Логічне виведення | Pellet, Transitive, RDFS, OWL, OWL Mini, OWL Micro, Generic rule reasoners | – | RDFS reasoning | GraphDB reasoner, OWL DLP, OWL Horst, OWL 2 R reasoning | FuXi reasoner |

як in-memory, так і постійна графові бази знань [5, с. 245–247].

Щоб вибір програмного забезпечення максимально відповідав необхідним критеріям до продукту з використанням сховища триплетів, необхідно проаналізувати їх за критеріями, що проілюстровані у Таблиці 1:

Звісно, вибір програмного забезпечення залежить від вимог до конкретного сховища триплетів: при необхідності широко використовувати можливості логічного виведення та працювати з OWL-онтологіями як з семантичним доповненням до RDF-да-

них, найкраще з розглянутих засобів підійде Apache Jena — TDB, оскільки ця система підтримує велику кількість різнорівнів та здатна працювати з багатьма діалектами OWL. Також для таких цілей, але з меншими можливостями логічного виведення може підійти GraphDB або ж RDFLib (вибір залежить від необхідності розробниками використовувати Java або Python). При необхідності якнайшвидше розгорнути додаток для ефективної роботи зі SPARQL, але без необхідності використовувати семантичні OWL різнорівні підійде Dydra (за використання REST API) або ж Sesame (за використання Java).

Література

1. Apache Jena TDB [Електронний ресурс]. — Режим доступу: <https://jena.apache.org/documentation/tdb/> — 01.02.2018
2. Berners-Lee, T.; Hendler, J.; Lassila, O. The Semantic Web. Sci. Am. / T. Berners-Lee, J. Hendler, O. Lassila // 2001, С. 35–37.
3. Guo Y., Pan Z. A Benchmark for OWL Knowledge Base Systems / Y. Guo, Z. Pan // Journal of Web Semantics 3, 2005, С. 125–128.
4. Pluskiewicz T. Introduction to Dydra, the Cloud-based RDF Store / T. Pluskiewicz [Електронний ресурс]. — Режим доступу: <http://t-code.pl/blog/2016/03/getting-started-with-dydra/> — 01.02.2018.
5. Thakker, D., Osman, T., Gohil, S., Lakin, P. A Pragmatic Approach to Semantic Repositories Benchmarking / D. Thakker, T. Osman, S. Gohil, P. Lakin // In Proceedings of the 7th Extended Semantic Web Conference (ESWC2010), Heraklion, Greece, 30 May–3 June 2010; Volume 6088, С. 245–247.