



Fundusze Europejskie
Wiedza Edukacja Rozwój

**Politechnika
Warszawska**

Unia Europejska
Europejski Fundusz Społeczny



MATERIALY DYDAKTYCZNE **zajęć praktycznych** **„Semantyczne grafowe bazy danych”**

realizowanego w zadaniu 41 „Rozszerzenie i certyfikacji kwalifikacji studentów
Wydziału Geodezji i Kartografii Politechniki Warszawskiej”

w ramach projektu „NERW PW. Nauka – Edukacja – Rozwój – Współpraca”

Wprowadzenie

Teoria grafów to dział matematyki zajmujący się badaniem własności grafów. Graf jest strukturą matematyczną służącą do przedstawiania i badania relacji między obiektami. Graf składa się z węzłów (zwanymi czasem wierzchołkami) i krawędzi. Zarówno węzły jak i krawędzie mogą posiadać przypisane sobie własności. Węzły grafu stanowią reprezentację obiektów, krawędzie obrazują relacje między obiektami. Krawędzie mogą mieć wyznaczony kierunek, a graf zawierający takie krawędzie nazywany jest grafem skierowanym. Krawędź grafu może posiadać wagę. W bazie grafowej każdy węzeł zawiera bezpośredni wskaźnik na sąsiadujące węzły i nie jest konieczne wyszukiwanie indeksowe do określenia związków pomiędzy obiektami. Relacje umożliwiają bezpośrednie łączenie danych w bazie i zazwyczaj pozwalają na pobieranie związanych ze sobą obiektów za pomocą jednej operacji. Grafowa baza danych wykorzystuje struktury grafów do przedstawiania i przechowywania danych oraz do obsługi zapytań semantycznych. Zapytania semantyczne pozwalają na przeprowadzenie analiz kontekstowych. Umożliwiają pobieranie danych w oparciu o informacje syntaktyczne, semantyczne i strukturalne w nich zawarte. Zapytania semantyczne są wykorzystywane w celu dostarczania precyzyjnych odpowiedzi na bardziej rozmyte zapytania poprzez dopasowywanie wzorców i rozumowanie cyfrowe.

Obecnie na rynku dostępnych jest około 30 baz grafowych. Portal db-engines.com prowadzi comiesięczne badania dotyczące popularności poszczególnych systemów zarządzania bazami danych. Ranking obejmuje zarówno systemy zarządzania relacyjnymi bazami danych jak i bazy typu NoSQL w tym bazy grafowe. Metoda pomiaru obejmuje analizę liczby stron internetowych dotyczących poszczególnych systemów baz danych znajdujących przez wyszukiwarki Google, Bing i Yandex, częstotliwość wyszukiwania nazw poszczególnych systemów podawaną przez Google Trends, liczbę dyskusji technicznych na portalach Stack Overflow i DBA Stack Exchange, liczbę ofert pracy dotyczących poszczególnych systemów na portalach Indeed i Simply Hired, liczbę profili na portalach LinkedIn i Upwork, w których są odwołania do poszczególnych systemów zarządzania bazami danych oraz liczbę wiadomości na portalu Twitter odnoszących się do poszczególnych systemów. W czerwcu 2018 wśród baz grafowych najpopularniejszą bazą był Neo4j. W zestawieniu popularności baz danych wszystkich typów Neo4j zajmował 22 pozycję. Natomiast wyniki ankiety przeprowadzonej przez portal Stack Overflow wskazują, iż Neo4j jest dwunastą najbardziej lubianą przez programistów bazą danych na świecie (Stack Overflow, 2018).

Neo4j został opracowany przez Neo Technology, Inc. w 2007 roku. Kod źródłowy został napisany w języku Java. Wersja bezpłatna jest dostępna na licencji GPL, natomiast wersja komercyjna na licencji AGPL. Językiem zapytań stosowanym w Neo4j jest język Cypher, którego składnia jest wzorowana na języku SQL. W bazie danych występują dwa typy elementów: węzły, będące odpowiednikami obiektów oraz krawędzie łączące węzły, odzwierciedlające relacje zachodzące pomiędzy węzłami. Węzły mogą być proste, składające się jedynie z nazw lub złożone, zawierające atrybuty obiektów. Krawędzie posiadają nazwę identyfikującą typ połączenia oraz mogą posiadać kierunek. Przemieszczanie się po krawędziach pomiędzy węzłami nazywane jest trawersowaniem i pozwala na szybki dostęp do obiektów powiązanych ze sobą. Neo4j nie posiada mechanizmów operacji na danych geoprzestrzennych, ale dostępne są one w Neo4j Spatial – bibliotece narzędzi dla Neo4j, która ułatwia udostępnianie danych przestrzennych. W szczególności można dodawać indeksy przestrzenne do już zlokalizowanych danych i wykonywać operacje przestrzenne na danych, takie jak wyszukiwanie danych w określonych regionach lub w określonej odległości od wskazanego punktu.

Plan zajęć

1. Wprowadzenie do baz grafowych
 - Wprowadzenie do teorii grafów
 - Podstawy modelowania danych w bazach grafowych
 - Instalacja Neo4j w środowisku Windows i Linux.
 - Konfiguracja Neo4j w środowisku Windows i Linux.
 - Interfejs użytkownika systemu Neo4j.
2. Podstawy języka Cypher
 - Składnia języka



- Operacje CRUD dla węzłów i krawędzi
- 3. Zaawansowane wykorzystanie języka Cypher
 - Filtrowanie danych
 - Sortowanie
 - Agregacja danych
 - Tworzenie unii
- 4. Migracja z baz relacyjnych do baz grafowych
 - Migracja schematu bazy
 - Migracja danych
 - Migracja zapytań
- 5. Algorytmy zaawansowane
 - Wyszukiwanie węzłów oddalonych o zadaną liczbę krawędzi
 - Wyszukiwanie najkrótszej drogi w grafie.

Materiały przydatne do wykonania ćwiczeń projektowych.

Oprogramowanie do pobrania oraz niezbędne instrukcje znajdują się na stronie:
<https://neo4j.com/download/other-releases/#releases>

Literatura

Amirian, P., Winstanley, A., & Basiri, A. (2013). NoSQL storage and management of geospatial data with emphasis on serving geospatial data using standard geospatial web services.

Baton, J., & Van Bruggen, R. (2017). Learning Neo4j 3.x - Second Edition. Birmingham: Packt Publishing Limited.

Brahim, M. B., Drira, W., Filali, F., & Hamdi, N. (2016). Spatial data extension for Cassandra NoSQL database. Journal of Big Data.

Celko, J. (2014). Joe Celko's Complete Guide to NoSQL. Waltham, MA: Elsevier Inc.

Edlich, P. D. (2016, 09 08). NoSQL . Retrieved from NoSQL Your Ultimate Guide to the Non-Relational Universe!: <http://nosql-database.org/>

Fauerbach, C. (2017). Learning Neo4j Graphs and Cypher [Video]. Birmingham: Packt Publishing Limited.

Goel, A. (2015). Neo4j Cookbook. Birmingham: Packt Publishing Limited.

Gupta, S. (2015). Building Web Applications with Python and Neo4j. Birmingham: Packt Publishing Limited.

Gupta, S. (2015). Neo4j Essentials. Birmingham: Packt Publishing Limited.

Hills, T. (2016). NoSQL and SQL Data Modelling. Basking Ridge, NJ: Technics Publications.

Jordan, G. (2014). Practical Neo4j. New York, NY: Apress Media LLC.

Kemper, C. (2015). Beginning Neo4j. New York, NY: Apress Media LLC .

Lal, M. (2015). Neo4j Graph Data Modeling. Birmingham: Packt Publishing Limited.

McCreary, D., & Kelly, A. (2014). Making Sense of NoSQL. Shelter Island, New York: Manning Publications Co.

Neo Technology. (2017). The Neo4j Developer Manual v3.1. Neo Technology.



Neo4j Spatial. (2017, 08 31). Neo4j Spatial. Pobrano z lokalizacji Neo4j Spatial: <http://neo4j-contrib.github.io/spatial/>

Panzarino, O. (2014). Learning Cypher. Birmingham: Packt Publishing Limited.

Raj, S. (2015). Neo4j High Performance. Birmingham: Packt Publishing Limited.

Sadalage, P. J. i Fowler, M. (2014). NoSQL Kompendium wiedzy. Helion.

Sadalage, P. J., & Fowler, M. (2012). NoSQL Distilled. Boston: Addison-Wesley Professional.

Sasaki, B. M., Chao, J., & Howard, R. (2018). Graph Databases for Beginners. Neo4j, Inc.

Stack Overflow. (2018, 06 03). Stack Overflow Developer Survey Results 2018. Retrieved from Stack Overflow: <https://insights.stackoverflow.com/survey/2018#technology-most-loved-dreaded-and-wanted-databases>

Strauch, C. (n.d.). NoSQL Databases. Stuttgart: Stuttgart Media University.

Tiwari, S. (2011). Professional NoSQL. Indianapolis: John Wiley & Sons, Inc.

Vaish, G. (2013). Getting Started with NoSQL. Birmingham: Packt Publishing Limited.

Van Bruggen, R. (2014). Learning Neo4j. Birmingham: Packt Publishing Limited.

Vohra, D. (2015). NoSQL Web Development with Apache Cassandra. Cengage Learning PTR.

Vukotic, A., Watt, N., Abedrabbo, T., Fox, D., & Partner, J. (2015). Neo4j in Action. Shelter Island, NY: Manning Publications Co.

Wyszomirski, M. (2018). Przegląd możliwości zastosowania wybranych baz danych NoSQL do zarządzania danymi przestrzennymi. Roczniki Geomatyki Tom 16, Zeszyt 1(80), 55–69.