
ANTAL JÁNOS BENJAMIN

mérnök informatikus
MSc, 4. félév

Budapesti Műszaki és Gazdaságtudományi
Egyetem
Villamosmérnöki és Informatikai Kar

ELEKES MÁRTON

mérnök informatikus
MSc, 1. félév

Budapesti Műszaki és Gazdaságtudományi
Egyetem
Villamosmérnöki és Informatikai Kar

Témavezetők:

Szárnyas Gábor

tudományos segédmunkatárs, BME VIK

Marton József Ernő

tudományos segédmunkatárs, BME VIK

Gráf információs rendszerek összehasonlító teljesítménymérése

Az elmúlt évtizedben sokféle különböző NoSQL technikát használó adatbázis-kezelő készült. Ezek egyik csoportja a gráfadatbázisoké, melyek lehetővé teszik az adatok gráf formában történő tárolását és lekérdezését. Ez az adatmodell gyakran jobban illeszkedik a sok összefüggést tartalmazó adatok tárolására, mint a relációs modell, és a tömörsége miatt gyakran képes jobb teljesítményt nyújtani. Mindezek ellenére, mivel a relációs adatbázisokat majdnem 50 éve fejlesztik és optimalizálják, jelenleg is nyitott kérdés, hogy szükség van-e specializált gráfadatbázisokra a gráf adatok hatékony feldolgozásához.

Gráflekérdezések megfogalmazására új lekérdező nyelvek jelentek meg, mint például az openCypher. Ezek a nyelveken gyakran kényelmesebben fogalmazhatunk meg gráflekérdezéseket, mint az SQL-alapú nyelveken. Az üzleti adatok jelentős része azonban jelenleg is hagyományos relációs adatbázisban van tárolva, emiatt ezen adatokat át kell tölteni gráfadatbázisokba, amely éles adatbázisok esetén általában nem megoldható. Célunk, hogy egy olyan megoldást készítsünk, amelyben lehetséges a magas kifejezőerővel rendelkező gráflekérdező nyelveken megfogalmazott lekérdezéseket hatékony relációs lekérdezőmotorokon futtatni anélkül, hogy szükség lenne egyik rendszerből a másikba áttölteni az adatokat. Ennek érdekében olyan fordítót (transpilert) készítettünk, ami képes openCypher lekérdezéseket SQL-re fordítani.

Különböző adatbázis-kezelő rendszerek összehasonlításához elengedhetetlenek a teljesítménymérési specifikációk (benchmarkok). Relációs adatbázisok esetében ezt a szerepet a Transaction Processing Performance Council benchmarkjai töltik be. A gráfadatbázisok relatív kiforratlansága miatt jelenleg kevés benchmark létezik a gráflekérdezések teljesítménymérésére. Mi az LDDB (Linked Data Benchmark Council) Social Network Benchmark fejlesztésébe kapcsolódunk be, amelynek keretében frissítettük és jelentősen fejlesztettük a meglévő implementációkat, továbbá elkészítettük a SPARQL nyelvű implementációt. Ezek felhasználásával alaposan megvizsgáltuk és részletesen elemeztük az adatbázis kezelőket különböző adatmodellek (relációs, gráf és szemantikus) felhasználásával.
