



Budapesti Műszaki és Gazdaságtudományi Egyetem  
Méréstechnika és Információs Rendszerek Tanszék

## Gráflekérdezések leképzése relációs adatbázisokra

**Elekes Márton I. évf, MSc mérnökinformatikus szakos hallgató**  
**Konzulens: Szárnyas Gábor, MIT**  
**Kritikus rendszerek specializáció**  
**Önálló laboratórium 2. összefoglaló**  
**2018/19. I. félév**

Az adatbázis-kezelés területén belül növekvő jelentőséggel bírnak a nem-relációs adatmodellt használó NoSQL rendszerek. Ezen rendszerek egyik kategóriája, az ún. gráfadatbázisok képesek hálózat jellegű adatokat hatékonyan tárolni és azokon lekérdezéseket futtatni. A gráfadatbázisokat jelenleg elterjedten használják is több specifikus használati esetben (pl. családetekció pénzügyi rendszerekben, ajánlórendszerek stb.), azonban széles körben még nem elterjedtek. Ennek elősegítésére a 2015-ben alapított openCypher kezdeményezés egy olyan nyílt szabvány készítésén dolgozik a Cypher lekérdezőnyelvhez, ami alapján különböző adatbáziskezelők egy közös nyelvet tudnak támogatni – hasonlóan a relációs adatbázis-kezelőkben alkalmazott SQL nyelvhez.

Az önálló laboratórium során a célom az volt, hogy a gráflekérdezések relációs adatbázisokon való alkalmazhatóságát vizsgáljam, továbbá openCypher nyelvű lekérdezések SQL nyelvre való leképzésével demonstráljam a megközelítést.

A hagyományos relációs adatbázisok tábla alapú szemléletével szemben a gráfadatbázisok mind a tárolt elemek tulajdonságainak, mind a struktúrájuk leírására alkalmasak. A relációs adatbázisok gyakorta tartalmaznak ehhez hasonló adatokat, azonban ehhez a tárolás és a lekérdezés során is különböző kényeszmegoldásokat (pl. kapcsolótáblák) kell alkalmazni. Felmerül az igény, hogy a meglévő relációs adatbázisokban tárolt adataink elérése során használhassunk gráflekérdezéseket.

A relációs adatbázisok gráfadatbázisokra való leképzéséhez egy az Object-Relational Mappinghez hasonló problémát kell megoldani. A kezdeti prototípus során egy rögzített, generikus relációs adatmodellen valósítottam meg a gráflekérdezések SQL-re való fordítását. Ez a séma a gráf csúcsait, éleit, a címkéket és a tulajdonságokat is egy-egy táblában tárolta. A megoldás lehetővé teszi, hogy meglévő relációs séma felett gráfnézetet biztosítsunk. Ehhez a Cytosm eszköz gTop sémaleíró nyelvét használok fel. A fordítást egy köztes, gráfokra kiterjesztett relációs algebrai fa felhasználásával végeztem, amelyet a tanszéken fejlesztett ingraph keretrendszer valósított meg. A megvalósítás során a relációs algebra egyes elemeit képeztem le SQL lekérdezésekre, továbbá lehetővé tettem ezek komponálását. Az elkészült implementáció PostgreSQL relációs adatbázist használ, mivel ez az adatbázis támogatja a tranzitív lezártat tartalmazó gráflekérdezésekhez szükséges rekurzív lekérdezéseket, valamint a JSON típusú oszlopok használatát.

Az elkészült Cypher-to-SQL rendszer vizsgálatához az openCypher TCK tesztjeit használtam fel. Az ebben található lekérdezések referenciaeredményét hasonlítottam össze a saját implementáció által szolgáltatott eredményekkel. Emellett az LDDB szervezet SNB gráfbenchmarkjának segítségével is mértük a rendszer teljesítményét. Ezek alapján a megoldás teljesítményén további fejlesztések szükségesek, azonban jelenlegi állapotában is képes volt egy kereskedelmi forgalomban kapható szemantikus adatbázissal összemérhető, több esetben jobb teljesítményt nyújtani a Cypher-to-SQL rendszer.

A jövőbeli terveim közé tartozik, hogy növeljem a nyelvi lefedettséget. Továbbá szeretném a meglévő, kézzel írott teszteseteket kiterjeszteni úgy, hogy automatikus tesztgenerálással a gráflekérdező motorok vizsgálatának fedését javítsam.