

RDF modellek feletti inkrementális gráfmenta kiértékelés vizsgálata

Év végi egyoldalas összefoglaló.

Pusztai Péter – U30VWN

Konzulens: Izsó Benedek

Az absztrakt ábrázolás segítségével lehetővé válik a számítógépes adatfeldolgozás. Ez az ábrázolás gyakran gráf jellegű. Azt, hogy a gráfban az egyes csomópontok és élek mit jelentenek, a metamodell írja le. A metamodellek megadására modellező nyelvek, a gráf adatainak kinyerésére lekérdezések használatosak. Ha sok a gráfra vonatkozó lekérdezés, akkor – a hatékonyság növelése érdekében – célszerű őket inkrementális kiértékelő motorral ellátni.

Az önálló labor keretén belül megismertem modellező nyelveket, mint az EMF (Eclipse Modeling Framework), az RDF (Resource Description Framework). A modellező nyelvek összehasonlításából kiderült, hogy az EMF és RDF adatleíró nyelvek különböző előnyökkel rendelkeznek. Megismertem az ezekhez kapcsolódó EMF-IncQuery és SPARQL lekérdező nyelveket. Ezen gráfmenta alapú lekérdező nyelveken megfogalmazott lekérdezéseket az EMF esetén EMF-IncQuery-vel, míg RDF esetén például Sesame-val lehet kiértékelni. Az EMF-IncQuery egy inkrementális gráfmenta illesztő, míg RDF esetén nincs olyan kiforrott lekérdező motor, ami inkrementálisan működne.

A félév során implementáltam egy RDF feletti inkrementális kiértékelő motort egy konkrét lekérdezéshez. „Kézzel” tetszőleges RETE háló építhető, amiben jelenleg három fajta csomópont lehet. Ezek a csomópontok rendre a StatementFilter (amik alapreláció kényszerekre szűrnék), a Join (amik változók alapján illesztik össze az egyes tuple-eket (n-eseket)) és a Production csomópont, ahol létrejön az eredmény. A csomópontok a részeredményeket tárolják, az aktuális eredmény mindig kiolvasható. Ha új állítást vesz fel valaki egy Sesame tárba, akkor azt az implementáció automatikusan leköveti, és ez alapján frissíti a cache-t. Így az előre meghatározott lekérdezésre azonnal megkapjuk az illeszkedő részgráfokat.

Tervem, hogy a témával tovább foglalkozom szakdolgozat formájában. A lekérdezőre számos továbbfejlesztés vár, mint például az automatikus RETE háló építés tetszőleges lekérdezéshez. Emellett célom még a triple törlésének támogatása, a negálás megvalósítása (egy új csomóponttípus bevezetésével), és az adatkövetkeztetések használata. Feladatom még, hogy a megoldást összevessem az ezen a téren elért legújabb kutatási eredményekkel.