



Inkrementális tranzitív lezárt számítás gráfokban

Szabó Tamás (KCTZMJ), V. évf, MSc mérnök inf. szakos hallgató

Konzulens: Bergmann Gábor Phd hallgató, MIT

Szolgáltatásbiztos rendszertervezés szakirány

Önálló laboratórium II beszámoló

2011/12. I. félév

Egy homogén kétváltozós relációt akkor nevezünk tranzitívnak, ha az elempárok azon viszonya, hogy egymással relációban állnak, „láncszerűen” tovább adódik. Egy A halmazon értelmezett \sim reláció tehát tranzitív, ha $a \sim b$ és $b \sim c$ teljesülése esetén $a \sim c$ is teljesül (a, b és c az A halmaz elemei). Gráfok esetén a tranzitív lezárt az a reláció, amely megadja, hogy az a csomópontból elérhető-e a b csomópont irányított úton. Ha a gráféleket egy e reláció adja meg, a tranzitív lezárt például az alábbi Datalog [1] programmal definiálható:

(1) $tc(a,b) :- e(a,b)$

(2) $tc(a,b) :- tc(a,c), e(c,b)$

Az Önálló laboratórium II. tárgy keretében az előző félévben megkezdett munkát folytattam;

- a korábban elkészített tranzitív lezártat számító algoritmusok továbbfejlesztésével foglalkoztam
- elkészült az EMF-IncQuery Base projekt, amely egy EMF modelleken gyakran alkalmazott műveletek hatékony implementációját tartalmazó könyvtár [2]
- a félév során Bergmann Gábor ötlete alapján elkészítettem egy új inkrementális tranzitív lezártat számító algoritmust (Incremental SCC – IncSCC)
- az elkészült algoritmusok hatékonyságát méréseken keresztül hasonlítottam össze.

Az IncSCC algoritmus elkészítését egy hosszabb ideig tartó irodalomkutatás előzte meg, ugyanis számos más algoritmus, adatszerkezet implementációjára is szükség volt. Az algoritmus alapötlete, hogy a korábbi félévben elkészült Counting algoritmust alkalmazza az alapgráf inkrementálisan karbantartott redukált gráfján. Ehhez szükség van az erősen összefüggő komponensek meghatározására, amelyhez a Tarjan [3] algoritmust használtam fel. Ezen kívül az erősen összefüggő komponensek UNIÓ-HOLVAN adatszerkezetben [4] vannak eltárolva.

Az elkészített algoritmus hatékonyságát számos mérésen keresztül vizsgáltam meg, illetve hasonlítottam össze az előző félévben elkészített algoritmusokkal. A mérések során vizsgáltam a tranzitív lezárt inicializálásának az idejét, illetve él beszúrás és törlés esetén a reláció frissítésének idejét. Az eredmények alapján elmondható, hogy az algoritmus futási ideje nagyobb modellek esetén is jól skálázódik. A munka folytatásaként az algoritmus lépésszámát szeretnénk részletesebben megvizsgálni és az elért eredményeket publikálni.

Hivatkozások

- [1] S. Ceri *et al.* 1989. What You Always Wanted to Know About Datalog (And Never Dared to Ask). *IEEE Trans. on Knowl. and Data Eng.* 1, 1 (March 1989), 146-166.
- [2] EMF IncQuery Base: <http://viatra.inf.mit.bme.hu/incquery/base>
- [3] Tarjan, R. E. (1972), "Depth-first search and linear graph algorithms", *SIAM Journal on Computing* 1 (2): 146–160, doi:10.1137/0201010
- [4] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*, Second Edition. MIT Press and McGraw-Hill, 2001. ISBN 0-262-03293-7. Chapter 21: Data structures for Disjoint Sets, pp. 498–524.