

Iteratívny výpočet relačných algebraických výrazov

Tomáš Plachetka

Výrazy relačnej algebry môžu generovať obrovské výsledky, ktoré nie je možné uložiť operačnej pamäte ani na disk. Predstavíme iteratívny spôsob výpočtu, ktorý sa používa v implementáciách moderných databázových systémoch, a umožňuje výpočet takýchto výrazov aj s malou pamäťou.

V tomto kontexte sa objavuje niekoľko zaujímavých problémov. Jedným je automatické odstraňovanie duplikovaných záznamov z výsledkov. V matematike je relácia štandardne definovaná ako množina usporiadaných n -tíc, avšak v SQL sa s reláciami pracuje ako s multimnožinami. Dôvodom je snaha o úsporu zložitosti spojenej s odstraňovaním duplikátov pri výpočte dotazov na úrovni relačnej algebry. Na druhej strane, časová zložitosť závisí predovšetkým od veľkosti medzivýsledkov, obzvlášť ak sa počítajú opakovane. Tak to je v prípade rekurzívnych dotazov. Pri návrhu databázového systému je dôležité vyhýbať sa predčasným optimalizáciám. Preferovať treba funkčnosť cieľového systému, ktorá prirodzene rekurziu zahŕňa. Duplikáty je preto rozumné dôsledne odstraňovať, aj za cenu počiatočnej réžie. Ukážeme algoritmické riešenie, ktoré funguje aj s extrémne malou pamäťou.

Kontakt: *doc. Dr. Tomáš Plachetka*
Katedra informatiky, Fakulta matematiky, fyziky a informatiky,
Univerzita Komenského, 84248 Mlynská dolina, Bratislava
e-mail: plachetka@fmph.uniba.sk