

R-tree

Pavel Jahoda

April 24, 2018

Popis Projektu

Cílem projektu je vytvoření vlastní perzistentní implementace R-stromu, což je stromová struktura pro vyhledávání n-dimenzionálních objektů často využívaná v geografických informačních systémech. Řešení překonává požadavky zadání a kromě dotazů nad databází 2D nebo 3D objektů, zvládá i požadavky na n-dimenzionální objekty. Dotazovat se lze jak na nejbližšího souseda, tak na rozsahových dotazem (tj. které objekty jsou K jednotek daleko od bodu X). Aplikace obsahuje testovací modul, který si dokáže generovat náhodná data tak i pracovat s ručně zadanými daty. Testovací modul kontroluje vyváženost stromu, korektnost velikosti bounding boxů, také testuje dotazování na nejbližšího sousedu a rozsahové dotazy

Navíc oproti zadání, které umožňuje generování náhodných dat (které generuji v testovacím modulu), stahuji data z aplikace flightradar od které s 5 minutovým spožděním získávám data o letištích a také letadlech, která jsou v daný moment ve vzduchu. Kromě polohy získávám rychlost, typ letadla, leteckou společnost ke které letadlo patří a další zajímavé údaje. Kromě náhodných dat je tedy možné se dotazovat i na letecká data.

Způsob řešení

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum

Implementace

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum

Příklad výstupu

Experimentální Sekce

R-strom nabízí hned několik věcí na testování. Z hlediska rychlosti vkládání prvků a následných dotazů můžeme porovnávat vliv různých metod dělení uzlů, počet potomků uzlu, počet dimenzí nebo vzdálenost u rozsahového dotazu.

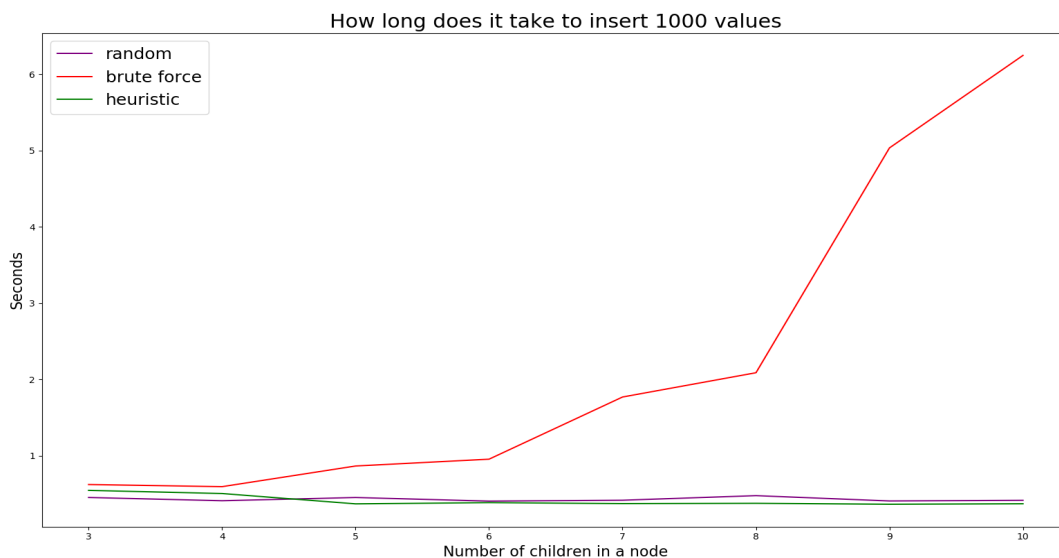
Tabulka níže (Table 1) ukazuje vliv volby metody dělení uzlů (když uzel dosáhl maximálního množství potomků) na rychlost vkládání prvků a rychlost rozsahových dotazů. Data v tabulce jsou zprůměrovaná z pěti různých testování na stromu s maximálním množstvím 8 potomků na uzel, na trojdimenzionálních datech na 2 000 datech a 20 000 dotazech. Data jsou v sekundách.

Table 1: Rychlost vkládání a dotazů

| | Insert | Query |
|-------------|---------|----------|
| Random | 1.0669 | 19.45292 |
| Brute force | 4.73896 | 5.33882 |
| Heuristic | 0.91115 | 7.51662 |

Výsledek je očekávaný. Metoda hrubou silou (brute force) je jasně nejpomalejší při vkládání, zatímco na dotazech vyhrává. Metoda dělení heuristikou nejenže nejrychleji vkládá prvky, ale vede si obstojně i v rozsahových dotazech, což je důvodem proč jsou heuristiky na R-stromech používány. Na druhou stranu metoda náhodného rozdělení uzlů je suveréně nejhorší na rozsahové dotazy.

Na grafu níže můžeme pozorovat vliv počtu potomků uzlu na dobu vkládání prvků. Zatímco vkládání s dělením uzlu heuristikou nebo náhodným dělením je s vzrůstajícím počtem potomků dokonce nepatrně rychlejší tak u dělení hrubou silou můžeme pozorovat velký pokles rychlosti přidávání. Tento fakt je způsoben tím, že řešení hrubou silou skouší všechny možné permutace rozdělení uzlu, což je asymptoticky exponenciální.



Na závěr experimentální sekce bych zde uvedl, že počet dimenzí hraje naprosto minimální vliv na rychlost vkládání nebo na rychlost rozsahových dotazů. Toto zjištění potvrzuji silné stránky R-stromu.

Diskuze

Závěr