

Posudek oponenta bakalářské práce

Autor/Autorka	Petr Trnka
Název práce	Porovnání vybraných databázových systémů s podporou prostorových dat
Studijní obor	Geomatika
Oponent práce	Ing. Michal Kepka, Ph.D.

Splnění cílů práce:

nadstandardně velmi dobře splněny s výhradami nebyly splněny

Odborný přínos práce:

nové výsledky netradiční postupy zpracování výsledků z různých zdrojů shrnutí výsledků z různých zdrojů bez přínosu

Odborná úroveň:

vynikající velmi dobrá průměrná podprůměrná nevyhovující

Věcné chyby:

téměř žádné vzhledem k rozsahu přiměřený počet méně podstatné, větší množství podstatnější, větší množství závažné

Grafická, jazyková a formální úroveň:

vynikající velmi dobrá průměrná podprůměrná nevyhovující

Slovní hodnocení a dotazy:

Vzhledem k neustálému rozvoji databázových systémů, ať už relačních, nebo i zástupců široké skupiny „NoSQL“ databází, kdy se rozšiřuje podpora pro práci s prostorovými daty v těchto systémech, je zvolené téma autora velice přínosné a aktuální.

Cílem práce bylo porovnat vybrané databázové systémy z hlediska výkonu při práci s prostorovými daty na základě autorem navržené a provedené sady testů.

Práce je vhodně členěna, v práci jsem ovšem nenašel seznam zkratk, který by byl vhodný, vzhledem k občasné absenci vysvětlení použitých zkratk v textu práce.

Po úvodní kapitole následuje obecný popis prostorového indexu a především řešerše různých druhů těchto prostorových indexů. Ke kapitole 2 bych měl drobnou připomínku, že časté odkazování se na podkapitoly předcházející a následující snižuje čtivost a přehlednost textu. V kapitole 3 je provedena řešerše tří databázových systémů s podporou pro prostorová data. V rozsahu, v jakém jsou uvedeny popisy zástupců NoSQL systémů, by mělo smysl uvést i další příklady systémů pracujících s prostorovými daty (relační i NoSQL). Zvláště v kontextu odkazování se na žebříček iniciativy DB-Engines. Kapitola 4 popisuje výběr testovací datové sady, přípravu a zpracování datové sady a distribuci mezi testovanými systémy. Zde se domnívám, že pokud by byl zvolen jiný nástroj pro import dat do PostgreSQL databáze (např. osm2pgsql), nemuselo by být řešeno spojování vrstev.

V kapitole 5 autor detailně popisuje jednotlivé navržené testy a strukturu výsledků. V této kapitole mě překvapuje volba stejné velikosti vkládaného vzorku pro všechny 3 typy geometrií, ačkoliv autor v kapitole 4 uvádí, že počty prvků podle typů geometrií jsou řádově rozdílné. Důležitým kritériem pro porovnání je určitě typ použitého prostorového indexu v konkrétním testovaném databázovém systému, ovšem v kapitole 5 ani jinde v práci tato informace není. V kapitole 6 jsou detailně popsány výsledky jednotlivých provedených testů, které jsou doprovázeny grafy v příloze B. V závěru autor shrnuje výsledky práce a krátce zdůvodňuje neprovedení většiny testů pro systém CouchDB.

V případě jazykové a formální úrovně práce se v práci nachází větší množství překlepů a evidentních přehlédnutých chyb. Např. na str. 8 jsou definovány pojmy minimální ohraničující obdélník resp. kostka, ovšem zavedeny jsou zkratky MOO/MOČ. Oproti tomu na str. 10 je najednou použito zkratky MOK. Na str. 13 je použit nesprávný termín „geodetický souřadný systém“, stejně tak pro GIS oblast netradiční termín „koordinované páry“.

Přínosy práce vidím především v systematickém provedení sady testů u 2 resp. 3 databázových systémů. A také v samotném návrhu sady testů, kdy sada pokrývá nejčastější operace s prostorovými daty.

Jako hlavní nedostatky práce vidím neuvedení důvodu pro vynechání části zásady pro vypracování číslo 2 – otestování funkcí pro správu dat, což bych chápal jako otestování nejen vkládání dat do DB, ale také jejich aktualizaci a odstranění. Jako další nedostatek vidím velice stručné zdůvodnění pro neprovedení většiny testů u systému CouchDB, kdy bych očekával, že u práce, která se zpracovává po dobu dvou let, dojde buď k překonání překážek, nebo k výměně testovaného databázového systému za jiný. V práci jsem také nenalezl informaci, zda autor práce provedl rešerši prací zabývajících se shodným nebo podobným tématem. V úplném závěru práce je zmíněna pouze jedna podobná práce. Znamená to, že nikdo jiný podobnou problematiku neřešil a výsledky nepublikoval?

Otázky:

1. Jaké konkrétní prostorové indexy byly použity v jednotlivých databázových systémech během testování?
2. Jak myslíte, že by se projevilo v testování, kdyby velikosti vzorků pro typy geometrií byly relativně shodné, nikoliv absolutně?
3. Kde konkrétně jsou publikovány zdrojové kódy testů? V práci je pouze zmíněno ve webové službě GitLab.

Práci doporučuji – ~~nedoporučuji~~ uznat jako kvalifikační (nehodící se škrtněte).

Navrhuji hodnocení známkou:

dobře

Datum, jméno a podpis:

10. 6. 2020

Ing. Michal Kepka, Ph.D.