

Posudek oponenta k diplomové práci

Mining XML Integrity Constraints

Stanislav Fajt

Cílem hodnocené diplomové práce měl být návrh nové metody na odvozování integritních omezení u množin XML dokumentů, který bude řešit nevýhody existujících přístupů.

Autor v řešení zadaného problému nejprve provedl detailní analýzu několika konkrétních existujících metod na odvozování klíčů a inkluzních závislostí nejenom nad XML dokumenty, ale také nad standardními relačními databázemi. Stejně tak byly popsány existující poznatky ze souvisejících oblastí, jako například odvozování XML schémat, teorie klíčů v XML dokumentech, role jiných typů integritních omezení či odvozování referenčních omezení pomocí XQuery dotazů.

Přínosem práce je návrh vylepšení dvou základních existujících metod, s cílem efektivního odvozování primárních a cizích klíčů nad jednotlivými XML dokumenty. Kardinalita těchto klíčů je omezena na základě zjištěných průměrných charakteristik reálných XML dokumentů. Složitost samotného řešení je netriviální a potřebnost práce je motivačně dobře odůvodněna, stejně jako jsou popsána možná využití v praxi.

Práce obsahuje dvě základní části, z nichž každá je tématicky rozdělena do několika kapitol. První část předkládá základní a částečně formální aparát řešené oblasti a dále uvádí popis řady existujících řešení i souvisejících prací. Druhá část je věnována nově navrženému řešení problému včetně popisu provedených experimentů. Struktura práce a účel jednotlivých kapitol jsou čitelné, prostor vyhrazený popisu existujících přístupů je však zbytečně velký. Celkový rozsah práce je nadprůměrný.

Positivním prvkem práce je snaha o formální zpracování na základě čitelně oddělených definic, příkladů a vysvětlujícího textu. Některé pasáže však díky své nepřesnosti nejsou pro čtenáře dobře čitelné, teprve díky dobře voleným příkladům je možné navržené myšlenky správně interpretovat. Formální definice celkově obsahují poměrně velké množství věcných chyb, překlepů i nejasných formulací. Práci by dále prospělo výraznější odůvodňování učiněných rozhodnutí, stejně jako větší promyšlenost představeného formálního modelu a jeho důsledné a koncepční používání v průběhu celé práce.

V souladu se zadáním byly provedeny experimenty, popisu jejich výchozích podmínek a zjištěných závěrů však mohl být věnován větší prostor. Charakterizace dat použitých pro experimenty nepřináší dostatečné údaje z hlediska činnosti navrženého algoritmu, také není diskutováno vyhodnocení časové složitosti algoritmu v závislosti na velikosti vstupních dat.

Přestože přínosy práce jsou zřejmé, více prostoru by mohlo být věnováno jasnému vymezení se vůči existujícím řešením, nově dodaným výhodám či kritické diskuzi slabších stránek nově navrženého řešení. Rozsah citovaných zdrojů je nadprůměrný, což svědčí o dobré teoretické přípravě. Zdroje se týkají tématu, stejně jako popisují osvědčené výsledky a aktuální přístupy v řešené oblasti.

Po stylistické a gramatické stránce je práce napsána dobrou angličtinou. Množství chyb je přijatelné, některé občasné se opakující systematické chyby jsou však v textu ponechány zbytečně. Z hlediska typografie obsahuje práce minimum chyb, grafická úprava a rozvržení jsou převážně dobré.

I přes uvedené nedostatky práce jako celek splňuje požadavky na diplomovou práci kladené, a proto ji doporučuji k obhajobě.

V Praze dne 19. 1. 2011

Mgr. Martin Svoboda
KSI MFF UK