

Posudek oponenta diplomové práce

Bc. Milan PLACHÝ

Fuzzy databáze založená na E-R modelu

Cílem práce bylo prostudovat možnosti rozšíření databáze o fuzzy logiku a její sémantiku na konceptuální úrovni, a navrhnout rozšíření relační databáze, pocházející z konceptuálního návrhu.

Text práce je po formální stránce poměrně dobře zpracovaný a seznamuje čtenáře s různými možnostmi fuzzyfikaci E-R modelů. Značení je přes fakt, že bylo přebíráno z mnoha citovaných článků, v celé práci konzistentní, což zvyšuje čitelnost a přehlednost práce. Nicméně bych uvítal nějaké konkrétní příklady u jednotlivých přístupů, které by více osvětlily rozdíly v přístupech jednotlivých autorů.

Příklad M:N vztahu s příbuzností osob (obr. 4 na str. 5) nebyl zcela vhodně zvolen. Vztah by měl být zřejmě tranzitivní (příbuzní mých příbuzných jsou mými příbuznými), čemuž příklad neodpovídá. Autor navíc explicitně zmiňuje to, že je vztah reflexivní. Tedy, že je každá osoba příbuzná sama se sebou. Pokud by to takto bylo chápáno, obrázek opět tyto vztahy neobsahuje.

Vzhledem k zadání bych očekával, že implementovaná aplikace dostane na vstupu fuzzyfikovaný E-R model vyjádřený v nějakém libovolném formalismu (např. v podobě XML, popisujícím entity, vztahy, atributy, a jejich fuzzyfikaci). Popis takového převodu je přitom v práci popsán. Aplikace ve stávající podobě je naproti tomu spíše prototypem, implementujícím jedno konkrétní E-R schéma, sestávající ze dvou entit, propojených M:N vztahem (viz str. 45). Domnívám se přitom, že samotná aplikace by zvládla zobrazovat i jiná, složitější, schémata, pokud by byla metadata o nich uložena v odpovídajících tabulkách. Chybí proto, zdá se, pouze nějaký preprocesor, který by z popisu E-R modelu vygeneroval odpovídající SQL schéma, a naplnil tabulky potřebnými metadaty.

Stávající prototypová implementace přenáší většinu aplikační logiky, spojené s výpočtem příslušnosti entit vzhledem k dotazu, na stranu klienta. Server eliminuje pouze řádky s nulovou podobností. V případě, že je nastavena nenulová minimální požadovaná podobnost záznamu vůči dotazu, na klienta se přenáší některé řádky zbytečně. Pro praktické účely by bylo vhodné implementovat logiku na straně serveru a realizovat fuzzy funkce a operátory jako vestavěné funkce a operátory databáze. Vzhledem k implementaci v C# a použití MS SQL serveru by možná šlo některé části kódu na server pouze přenést, případně provést jen minimální úpravy. Výhodu implementace na klientovi vidím pouze v tom, že je možné jí snadněji přenést na jinou cílovou databázi.

Pokud měla implementace pouze ukázat možnosti, nabízené fuzzyfikaci databáze, je toto řešení odpovídající. Hodilo by se však nabídnout více variant sémantiky. Například možnost zvolit, jakou sémantiku mají logické spojky, zatímco nyní se používá výhradně produktová sémantika. Stejně tak by mohlo být přínosné mít možnost zvolit, jakým způsobem bude vyhodnocena podobnost dvou zhodnocovacích funkcí.

Dále práce obsahuje testy rychlosti odezvy databáze vzhledem k několika dotazům různé složitosti. Ukazuje se, že odezva je úměrná velikosti dat, ale také velikosti odpovědi, protože data byla vygenerována s rovnoměrným rozdělením. Lineární závislost vzhledem k velikosti dat je celkem odpovídající, ale přesto bych uvítal kromě zobrazeného SQL dotazu i plán jeho provedení, aby bylo možné posoudit, zda se ve výkonu neskrývají ještě nějaké rezervy.

Celkově se domnívám, práce přes výše uvedené připomínky splňuje podmínky kladené na diplomovou práci. Doporučuji proto práci k obhajobě.

V Praze dne 27. 8. 2012

RNDr. Michal Kopecký, Ph.D.
KSI MFF UK