

Oponentský posudek Tezí disertační práce

Indexing structure for Similarity Search in Metric Spaces

Mgr. Vlastimila Dohnala

Obsah práce

Předložené teze se zabývají aktuální problematikou vyhledávání vzhledem k podobnostnímu kritériu. Podobnost je modelována pomocí metriky.

Práce podává přehled aktuálního stavu výzkumu v této oblasti. V úvodní části jsou uvedeny potřebné pojmy. V další části je shrnut současný stav studované problematiky. V této části je uveden poměrně podrobný přehled indexovacích struktur pro indexování metrických prostorů.

Práce je zaměřena na navržení nové datové struktury, která by kombinovala techniku shlukování a pivotovací strategii s cílem zrychlit provádění podobnostních dotazů. Dále autor uvažuje o možnosti zpracovávat navrženou strukturu v paralelních nebo distribuovaných prostředích.

Techniku shlukování a pivotovací strategii lze neformálně chapat jako dvě ortogonální metody a proto lze očekávat, že jejich kombinace přinese zajímavé výsledky.

Dotazy a připomínky

K předloženým tezím mám následující dotazy a připomínky:

V práci jsou uváděny pouze metriky, které je možno získat z archimedovských norem. Metriky odpovídající nearchimedovským normám splňují vlastnost ultrametriky. Ultrametriky mají některé vlastnosti neodpovídající běžným představám o vzdálenosti. Pro příklad uvádím alespoň dvě tyto vlastnosti.

- Každý bod koule je jejím středem.
- Dvě koule jsou buď disjunktní nebo je jedna obsažena v druhé.

Vzhledem k tomu, že se v práci mluví o metrice obecně mohou být některé úvahy zavádějící viz Figure 4.1.

Připomínám, že ultrametriky hrají velmi důležitou úlohu v oborech jako je biologie, ekonomie.

Dále na základě Ostrovského věty z p-adické analýzy je zřejmé, že ultrametriky tvoří významnou třídu metrik na racionálních číslech.

Zajímalo by mne o jakých metrikách autor uvažuje a jaké metriky by byly použitelné pro navrhovanou metodu?

Dále by mne zajímalo jakým způsobem reaguje navrhovaná struktura na problém dimenzionality. Se vzrůstající dimenzí jsou téměř všechny body koule u jejího povrchu. Pro příklad uvádím, že "slupka" silná 1cm koule o poloměru 20cm ve 20ti rozměrném prostoru obsahuje $\frac{2}{3}$ objemu celé koule.

Závěr

Teze jsou napsány srozumitelně a jasně. Přehledová část práce svědčí o velmi dobré orientaci v dané problematice.

Doktorand je spoluautorem tří zahraničních publikací. Tyto publikace byly prezentovány na konferencích s přísným recenzním řízením.

Z výše uvedeného je zřejmé, že autor má jasnou představu o cíli disertační práce a prokázal také schopnost vědecké práce.

Předložené teze disertační práce mohou sloužit jako východisko k disertační práci.

V Ostravě dne 12. prosince 2002
Doc. RNDr. Václav Snášel CSc.,
Katedra informatiky
Fakulta elektrotechniky a informatiky
VŠB-TU Ostrava

