

Název: Stromové indexační metody pro podobnostní vyhledávání
v metrických a nemetrických prostorech

Autor: Mgr. Jakub Lokoč

Katedra: Katedra softwarového inženýrství

Matematicko-fyzikální fakulta

Univerzita Karlova v Praze

Školitel: Doc. RNDr. Tomáš Skopal, Ph.D.

Email autora: lokoc@ksi.mff.cuni.cz

Email školitele: skopal@ksi.mff.cuni.cz

Abstrakt: M-strom je dnes již klasická indexační metoda používaná pro efektivní podobnostní vyhledávání v metrických prostorech. Ačkoliv M-strom již nepatří mezi nejnovější metody, věříme, že stále nabízí zatím neobjevený potenciál. V této práci se proto zaměřujeme na způsoby, jak vylepšit jeho původní algoritmy a strukturu. Abychom umožnili rychlejší zpracování dotazů pomocí M-stromu, navrhli jsme několik nových metod jeho konstrukce (i paralelních), které vedou k vytváření kompaktnějších metrických hierarchií a přitom nejsou extrémně drahé. Dále jsme ukázali snadný způsob, jak rozšířit M-strom na novou indexační metodu NM-strom, která slouží k efektivnímu nemetrickému podobnostnímu vyhledávání za pomoci algoritmu TriGen. Všechna tato experimentálně ověřená vylepšení prokazují, že můžeme M-strom stále ještě považovat za důležitou dynamickou metrickou přístupovou metodu vhodnou pro správu rozsáhlých kolekcí nestrukturovaných dat. Všechna prezentovaná vylepšení mohou být navíc implementována do následníků M-stromu (např. do PM-stromu), což otevírá dveře pro další výzkum v této oblasti.

Klíčová slova: podobnostní vyhledávání, metrické přístupové metody, indexování, M-strom, TriGen