

Posudek oponenta k diplomové práci

Adaptive Similarity of XML Data

Bc. Eva Jílková

Cílem hodnocené práce bylo studium různých existujících přístupů v oblasti (polo-)automatického vyhodnocování podobnosti schémat a dokumentů z rodiny XML technologií. Na základě jejich vzájemného porovnání měla být navržena, implementována a také i experimentálně ověřena nová metoda, která by řešila nedostatky těchto existujících přístupů. Zadání práce bylo v tomto smyslu splněno.

Autorka se ve své práci zaměřila na možnosti praktické aplikace těchto podobnostních metod v oblasti konceptuálního modelování schémat XML dokumentů, a to v rámci dvouúrovňového systému platformě-specifických (PSM) a platformě-nezávislých schémat (PIM). Cílem bylo diskutovat a navrhnout vlastní metody, jak hledat mapování prvků (tříd, atributů) namodelovaných v jednotlivých PSM schématech (odpovídajících DTD nebo XSD schémátům) na prvky jejich jednoho společného PIM schématu.

Základem tohoto mapování je využití myšlenky rozhodovacích stromů, jejichž průchodem jsme schopni pro každou konkrétní dvojici prvků z daného PSM a PIM schématu ne/doporučit vhodnost příslušného mapování. Tyto stromy jsou vybudovány pomocí široké sady podporovaných elementárních porovnávacích metod, jako např. Levenshteinova editační vzdálenost a další. Konstrukce rozhodovacích stromů je však kritickým místem. Proto autorka navrhuje použít algoritmus C5.0, tedy natrénovat rozhodovací strom na základě ručně anotovaných dat.

Vlastní text práce je napsán kvalitní angličtinou s minimem chyb a překlepů, stejně jako rozvržení a typografie jsou převážně dobré. Celkový rozsah textu naplňuje očekávání a práce také obsahuje všechny obvyklé části. Více prostoru však mělo být věnováno ucelenému a důkladnému popisu navržené metody, a to nejenom ve vztahu k rozsahu ostatních částí, ale také proto, aby čtenář nebyl nucen poskládat si konkrétní obrázek o fungování navrženého řešení sám.

Jednotlivé části textu navíc často kvůli rozdílnému způsobu vyjadřování i dosažené úrovni jeho přesnosti a formálnosti budí pocit nekonzistence. To je nejspíše způsobeno tím, že informace o existujících přístupech jsou podávány stylem velmi blízkým jejich příslušným publikacím, aniž by jejich výklad byl dostatečně integrován, přizpůsoben potřebám a pohledu této práce a hlavně doplněn potřebnými vysvětleními, vlastními názory, kvalitní argumentací a systematičtějším přístupem.

Přestože je práce z teoretického pohledu spíše méně složitá a je založena pouze na nalezení vhodné aplikace a kombinace různých přístupů, silnou stránkou je nepochybně široký rozhled autorky v existujících technologiích a práce s literaturou obecně. Pozitivem je také zpracování celé řady doprovodných příkladů a ilustrací, které na vhodných místech obecnou rovinu textu doplňují a dobře rozvádějí.

Navržené řešení na mapování schémat bylo implementováno jako komponenta existujícího systému eXolutio. Díky tomu mají uživatelé k dispozici základní prototypové grafické rozhraní, v rámci kterého jsou schopni realizovat téměř celý proces mapování. Aby však byl systém v praxi použitelný a dostatečně uživatelsky přívětivý (což je v oblasti modelování více než nezbytné), bude nutné tomuto rozhraní i jeho interaktivitě potřeba věnovat další zásadní pozornost. V každém případě rozsáhlou práci autorky také potvrzuje provedení ruční anotace testovacích dat a samotných experimentů, byť jejich výstupy asi nemohou přinášet žádné obecně platné závěry.

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

V Brisbane dne 29. prosince 2013

RNDr. Martin Svoboda
KSI MFF UK