

Optimization and Refinement of XML Schema Inference Approaches

Předkládaná diplomová práce se zabývá problematikou odvozování XML schémat z množiny dostupných XML dokumentů pomocí pokročilých metod MDL. Autorem navržený postup umožňuje využít k odvození již existující, ale zastaralé či neúplné schéma.

V první části práce autor vytváří rozsáhlý teoretický aparát, který v dalších částech využívá k přesné definici problému i jeho řešení. V dalších částech práce jsou popsány existující přístupy. Autor tyto přístupy kombinuje a spolu s obhájeným softwarovým projektem popisuje komplexní nástroj pro odvozování XML schémat. V rámci tohoto řešení autor implementoval několik algoritmů pro odvozování a zjednodušování schématu a tyto přístupy experimentálně ověřuje na uměle generovaných datech a výsledky detailně komentuje.

K práci nemám výhrady. Jazykově i strukturou je na velmi dobré úrovni, psána dobrou angličtinou s minimem typografických chyb. Hlavními klady práce jsou:

- Množství příkladů, které dobře vysvětlují základní prvky navrhovaného řešení
- Obsáhlá experimentální část včetně důkladného popisu naměřených výsledků a diskuze nad nimi
- Univerzální řešení umožňující snadné rozšíření implementovaných metod o další a jejich snadné porovnání

K práci mám následující dotaz:

- Používané metody způsobují zjednodušení schématu, která nemusí být vždy žádaná. Bylo by možné upravit algoritmy tak, aby zachovávali některé uživatelem definované prvky schématu a jaké by to mělo důsledky na výpočetní náročnost jednotlivých metod?

Práce splňuje původní zadání i cíl. Student pronikl do zadané problematiky a prokázal schopnost detailně analyzovat a řešit zadaný problém. Navržené přístupy lze ihned publikovat.

Práce Michala Klempy dle mého názoru splňuje podmínky kladené na diplomovou práci, a proto ji doporučuji k obhajobě.

V Praze, dne 20.8.2011

Mgr. Jakub Stárka
KSI MFF UK