

# Posudek oponenta na diplomovou práci Ondřeje Vošty

## *Automatická konstrukce schématu pro množinu XML dokumentů*

Cílem práce bylo navrhnout algoritmus, který z množiny daných XML dokumentů bude schopen odvodit jejich společné schéma a to podrobně zapsat v jazyce XML Schema. V první části práce se autor věnuje analýze problému, analyzuje vlastnosti, které bychom od výsledku očekávali, a diskutuje již existující řešení. Ta rozčleňuje do dvou základních skupin – na heuristická řešení problému a na řešení založená na odvozování gramatiky. V obou dvou skupinách se věnuje především těm algoritmům a částem řešení, které pak ve své vlastní práci dále rozvíjí, či využívá pro další srovnání. Z textu je patrné, že se autor v problematice dobře orientuje, v celé práci cituje a rozebírá myšlenkově skutečně různorodé přístupy k řešení problému.

Ve druhé části pak pan Vošta popisuje vlastní algoritmus vycházející zejména z metod sk-Ant a Xtract. Popis je veden metodou shora dolů, kdy autor nejprve nastíní celkový algoritmus a vztahy mezi jednotlivými částmi a teprve později se jim věnuje podrobně. Výsledný algoritmus je spíše evolučního typu, vzniklý vhodnou kombinací a modifikacemi již existujících postupů. Jeho hlavním přínosem je především využití pokročilejších možností jazyka XML Schema oproti předchozím postupům uvažujícím pouze DTD.

Text je psán poměrně jasně a čitelně s dostatečnou mírou formalizmů, jeho strukturování by nicméně mohlo být podstatně lepší – i při členění do dokonce šesti vrstev působí logická struktura textu matoucím dojmem. Stejně tak přílišné zestručnění v některých pasážích činí text „velmi hutným“ a obtížně pochopitelným bez nahlédnutí do citovaných pramenů. Výhrady lze mít i k občasným překlepům a hrubkám. Naopak dobrým dojmem působí zápis algoritmů, které jsou vedle vysvětlujícího textu také přehledně popsány v programovacím pseudo-jazyce.

Popis algoritmu je veden dostatečně podrobně včetně vysvětlujících diagramů, jako hlavní výhradu bych uvedl to, že se autor téměř nevěnuje časovým a prostorovým nárokům jeho jednotlivých částí a to, že u některých modifikací neukazuje ani správnost či (u heuristických algoritmů) alespoň zaručenou konečnost procedury při prohledávání stavového prostoru.

Bohužel ukázková implementace je zaměřena také spíše na názornost a přehlednost implementace, takže kvůli některým přílišným zjednodušením si nelze ani z vlastního programu udělat podrobnější představu o praktické použitelnosti této metody na větších datech. Určitá diskuze, nebo alespoň srovnání s jiným programem řešícím podobnou úlohu, by jistě byla na místě. Pro méně komplexní reálné XML dokumenty nicméně ukázkový program dává dobré výsledky.

**Závěr:** Autor ukazuje možnosti při generování schémat dle vzorových XML dokumentů a představuje i dobré vlastní řešení, jak se s tímto problémem vypořádat. Je implementován i ukázkový program, který je schopen zvládat reálné i uměle vytvořené XML dokumenty. Celkově práce splňuje zadání v dostatečné míře a proto ji doporučuji k obhajobě.

V Praze dne 22.1.2006

Mgr. Kamil Toman