

Optimization and Refinement of XML Schema Inference Approaches

Předkládaná diplomová práce se zabývá oblastí automatického odvozování XML schématu pro dané vstupní XML dokumenty. Vzhledem k tomu, že pro množinu XML dat je obecně možné odvodit velké množství různě vhodných schémat, autor se zaměřuje na možnosti zpřesnění tohoto výběru. Pro tyto účely zkoumá různé varianty vyhodnocení kvality schématu pomocí MDL principu a možnost využití původního, již nekorektního schématu.

K práci samotné nemám žádné výhrady. Kvalita zpracování tématu odpovídá požadavkům na diplomovou práci MFF UK. V textu práce nalezneme veškeré očekávané součásti. Hlavní přínosy vidím v následujících bodech:

- Autor využívá úspěšně obhájený studentský SW projekt *jInfer* – framework umožňující odvozování XML schémat pro danou množinu XML dat. Projekt je založený na principu plug-inů, které reprezentují jednotlivé kroky odvozování, tudíž mohl autor snadným způsobem doplnit nové přístupy a tyto srovnat se stávajícími.
- Pro účely návrhu možných vylepšení autor nastudoval poměrně netriviální problematiku principu MDL (minimum description length) a kódování obecně.
- V práci je navrženo několik možných vylepšení a tyto jsou experimentálně srovnány na různých typech uměle vygenerovaných dat. Autor hodnotí navržená vylepšení objektivně a nesnaží se za každou cenu experimentálně dokázat teoretické hypotézy a očekávaná zefektivnění.
- Z obecného hlediska mají výsledky práce potenciál publikace na mezinárodní konferenci. Celý SW projekt *jInfer*, na jehož vývoji se autor podílel a který je dále rozvíjen v dalších diplomových pracích, také ve vhodném odborném časopise.

Práce má kvalitní úpravu, je napsána anglicky na velmi dobré úrovni a obsahuje minimum typografických chyb. Pro popis algoritmů zavedl autor netriviální formální aparát, který precizně využívá. Velmi pěkně jsou zasazeny citace – čtenář vždy přesně ví, které algoritmy, tvrzení apod. byly převzaty.

Z obecného hlediska práce původní zadání i cíl splňuje. Student dostatečně pronikl do zvolené problematiky a prokázal schopnost detailně zanalyzovat a řešit zadaný problém. Řešení je popsáno na teoretické úrovni a současně doplněno kvalitní implementací zasazenou do kontextu většího projektu.

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

V Praze, 6.8. 2011



RNDr. Irena Mlýnková, Ph.D.
KSI MFF UK