

Oponentský posudek diplomové práce Mateje Vitáska

### **Inference of XML Integrity Constraints**

Diplomová práce si klade za cíl rozšíření zavedených praktik automatického odvozování schémat pro XML dokumenty o odvození integritních omezení. Autor se zaměřuje na ID/IDREF atributy z DTD. Práce je členěna do několika logických celků. V první části (Kapitoly 1, 2) autor nastiňuje cíle práce, zavádí terminologii a prezentuje související práce. V Kapitole 3 je pak popsán vlastní autorův algoritmus (přístup) k odvozování daných integritních omezení, který je pak otestován v rozsáhlé sadě experimentů (Kapitola 4).

Hlavním přínosem práce je algoritmus hledající optimální sadu ID atributů na základě vstupního dokumentu. Algoritmus je založený na řešení úlohy lineárního programování – tzv. mixed integer problemu (využití GLPK solver simplexové metody). Autor také prezentuje sadu heuristik pro urychlení běhu tohoto algoritmu na úkor optimality řešení. Volba algoritmu je vzhledem k povaze problému (NP-těžký) rozumná.

Vydvihnul bych určitě kvalitní formalismus problému a rozsáhlou sadu experimentů, které mají za cíl porovnat autorův přístup se současnými jinými přístupy a také porovnat heuristiky navzájem.

Jedinou mojí výtkou je nenaplnění všech bodů zadání (key, keyref constructs, XPath constraints), nicméně vzhledem ke kvalitě předkládané práce lze prominout.

Práce je napsána v anglickém jazyce. Jazyková úroveň práce je nadprůměrná.

K práci nemám jiné zásadnější výtky. Práce Mateje Vitáska splňuje nároky na diplomovou práci kladené, proto ji **doporučuji uznat** jako práci diplomovou.

V Praze, 18. 1. 2012

RNDr. Tomáš Knap

Oponent diplomové práce