



Hodnocení vedoucího závěrečné práce

Student: Lukáš Renc
Vedoucí práce: Ing. Eliška Šestáková
Název práce: Automata Approach to XML Data Indexing: Implementation and Experimental Evaluation
Obor: Teoretická informatika

Datum vytvoření: 7. 6. 2018

<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – následující škálou 1 až 4:</i>
1. Splnění zadání	1=zadání splněno, 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno
<i>Popis kritéria:</i> Posuďte, zda předložená ZP dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posuďte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.	
<i>Komentář:</i> Hlavním cílem práce byla implementace a experimentální vyhodnocení tří existujících metod pro indexování XML dokumentů (TSPA, TSPSA, TPA). Protože součástí zadání byla instrukce soustředit se na efektivitu těchto implementací, očekávala bych, že rešeršní kapitola 3 bude podrobnější. Oceňuji ovšem, že se student rozhodl promyslet celou konstrukci jednotlivých metod od začátku a věnovat se návrhu nového (efektivnějšího) algoritmu pro jejich budování.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
2. Písemná část práce	60 (D)
<i>Popis kritéria:</i> Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posuďte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti. Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře. Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3. Posuďte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.	

Komentář:

Rozsah odpovídá bakalářské práci. Text však není místy úplně šťastně strukturován - například sekci 1.2.1 zabývající se typy XPath dotazů bych nezařadila do kapitoly o XML (1.2) ale pod kapitolu o XPath (1.3). Obecně mi přijde, že je práce obtížně čitelná - zvláště pro někoho, kdo se v dané oblasti vůbec nepohybuje. Části na sebe občas úplně nenavazují a obrázky se občas v textu vyskytují bez na první pohled jasného kontextu.

Oceňuji, že je text psán v anglickém jazyce - bohužel však obsahuje větší množství gramatických a stylistických chyb (chybějící nebo nevhodně použité členy, stavba vět, chybějící podmět). Též volba některých výrazů není do odborné práce příliš vhodná (např. API is not super friendliest). Po formální stránce obsahuje práce též několik nesrovnalostí, které zřejmě vznikly z důvodu spěšného psaní textu - např. v abstraktu se mluví o / a // přechodech (správně by se ovšem mělo mluvit o XPath osách), v sekci 1.2.1 jsou nadefinovány zkratky pro množiny XPath dotazů, které jsou však dále v práci použity v jiném formátu (např. XP^/,nametest vs. XP^{/,nametest}). V nadpisu sekce 1.2.1.1 je použita zkratka pro označení XPath dotazů, která nebyla definována. Dále je v práci několikrát odkazováno na metodu Tree Path Automaton, která se však správně jmenuje Tree Paths Automaton.

Bohužel ani z typografického hlediska není práce dokonalá - části textu přechnívají okraje, výčet ukončen čárkou namísto tečky či latexové příkazy vyskytující se v tištěném textu (např. str. 11, algoritmus 1), proměnné jsou občas psány kurzívou a občas ne, text obsahuje velká bílá místa a nevhodně umístěné obrázky.

Rešeršní kapitola 3, která se zabývá efektivní implementací automatů by mohla být podrobnější. Oceňuji však, že student navrhl a popsal jiný způsob konstrukce jednotlivých indexovacích metod - popis těchto nových algoritmů by ovšem mohl být názornější.

V části o implementaci a popisu knihovny mi chybí názornější informace o provázání tříd formou class diagramu.

Z celkového pohledu práce obsahuje informace, které má. Bohužel se spěch na formě textu podepsal. Pokud by finalizace textu získala větší prostor, jistě by to práci pomohlo a zvýšilo i mé hodnocení.

Hodnotící kritérium:

Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):

3. Nepsímná část, přílohy

70 (C)

Popis kritéria:

Dle charakteru práce se případně vyjádřete k nepsímné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů

Komentář:

Oceňuji, že knihovna využívá knihovny SAX namísto JDOM a efektivně tak z XML dokumentu extrahuje informace o string paths. Chybí mi však dokumentace knihovny (javadoc). Kód by bylo třeba ještě zrevidovat a začistit (obsahuje například zakomentované části kódu). Nestandardně na mě působí také fakt, že třídy implementace nejsou součástí žádného balíčku - což momentálně znesnadňuje použití knihovny. Spuštění Examples vyžaduje balíček saxon, který není zmíněn v readme ani přiložen na disku.

Hodnotící kritérium:

Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):

4. Hodnocení výsledků, jejich využitelnost

80 (B)

Popis kritéria:

Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

Komentář:

Výsledkem práce je knihovna implementující metody pro indexování XML dokumentů za využití konečných automatů. Za hlavní výsledek bych považovala návrh nového efektivnějšího algoritmu pro budování indexů. Ačkoli by experimentální část mohla být bohatší (například pracovat s XML s l-property a bez), považuji výsledky za přínosné.

Hodnotící kritérium:

Způsob hodnocení – následující škálou 1 až 5:

5. Aktivita a samostatnost studenta

5a:

1=výborná aktivita,

2=velmi dobrá aktivita,

3=průměrná aktivita,

4=slabší, ale ještě dostatečná aktivita,

5=nedostatečná aktivita

5b:

1=výborná samostatnost,

2=velmi dobrá samostatnost,

3=průměrná samostatnost,

4=slabší, ale ještě dostatečná samostatnost,

5=nedostatečná samostatnost

Popis kritéria:

V souvislosti s průběhem a výsledkem práce posudte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven (5a). Posudte schopnost studenta samostatně tvůrčí práce (5b).

Komentář:

Student pravidelně navštěvoval schůzky, ne vždy však zvládl v termínu plnit zadané úkoly.

Hodnotící kritérium:

Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):

6. Celkové hodnocení

69 (D)

Popis kritéria:

Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.

Text hodnocení:

Z celkového pohledu práce obsahuje informace, které má. Bohužel se spěch na formě textu podepsal. Práce bohužel není perfektní po stránce formální, jazykové ani typografické. Pokud by finalizace textu získala větší prostor, jistě by to práci pomohlo a zvýšilo i mé hodnocení. Stejně tak by revizi zasloužila i implementace, kde například chybí komentáře - to činí knihovnu obtížně použitelnou v praxi. Oceňuji však návrh nového efektivnějšího algoritmu pro budování indexů.

Vzhledem k těmto skutečnostem hodnotím práci 69 body, tedy stupněm D.

Podpis vedoucího práce: