



# Posudek oponenta závěrečné práce

**Student:** Maria Karzhenkova  
**Oponent práce:** Ing. Tomáš Pecka  
**Název práce:** Automata Approach to XML Data Indexing: Selecting Unknown Nodes  
**Obor:** Teoretická informatika

**Datum vytvoření:** 11. 6. 2018

<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení - následující škálou 1 až 5:</b>
<b>1. Náročnost a další komentář k zadání</b>	<b>1=mimořádně náročné zadání, 2=náročnější zadání, 3=průměrně náročné zadání, 4=lehčí, ale ještě dostatečně náročné zadání, 5=nedostatečně náročné zadání</b>
<b>Popis kritéria:</b> Podrobněji charakterizujte diplomovou (bakalářskou) práci a její případné návaznosti na předchozí nebo běžící projekty. Dále posuďte, čím je zadání této ZP náročné. (U obtížnější ZP lze dále tolerovat některé nedostatky, které by u ZP standardní obtížnosti tolerovány nebyly; a naopak u jednoduché ZP mohou být zjištěné nedostatky hodnoceny přísněji.)	
<b>Komentář:</b> Studentka musela nastudovat metody indexování XML pro XPath dotazy za pomoci konečných automatů, stringologické základy a z těchto poznatků navrhnout vlastní vylepšení.	
<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení - následující škálou 1 až 4:</b>
<b>2. Splnění zadání</b>	<b>1=zadání splněno, 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno</b>
<b>Popis kritéria:</b> Posuďte, zda předložená ZP splňuje zadání. V komentáři uveďte body zadání, které nebyly zcela splněny, případně rozšíření ZP oproti původnímu zadání. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.	
<b>Komentář:</b> Studentka nastudovala a vytvořila rozšíření nejen indexovacích metod TSPA a TSPSA, ale i TPA, což nebylo požadováno. Dále rozšíření všech tří indexů naimplementovala v jazyce Java.	
<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení - následující škálou 1 až 4:</b>
<b>3. Rozsah písemné zprávy</b>	<b>1=splňuje požadavky, 2=splňuje požadavky s menšími výhradami, 3=splňuje požadavky s většími výhradami, 4=nesplňuje požadavky</b>
<b>Popis kritéria:</b> Zhodnotte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části.	
<b>Komentář:</b> Rozsah je oproti jiným bakalářským pracím lehce nadprůměrný. Všechny kapitoly jsou relevantní, jen kapitola 3 by mohla být trochu obsáhlejší.	
<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):</b>
<b>4. Věcná a logická úroveň práce</b>	<b>90 (A)</b>
<b>Popis kritéria:</b> Posuďte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti. Zhodnotte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře.	
<b>Komentář:</b> Po věcné stránce je práce v pořádku, chybí mi však definice subsequence automatu, prefix automatu a backbone, na které se odvolává v kapitolách 3 a 4. V definici 2.24 má být množina koncových stavů automatu M2 značena F2, nikoli F1. V druhém odstavci kapitoly 6.1 a v prvním odstavci kapitoly 6.2 je zmíněno 6 datasetů, v tabulkách a grafech kapitoly 6 je jich však znázorněno jen 5, stejně jako je jich jen 5 přiloženo k práci.	
<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):</b>

## 5. Formální úroveň práce

95 (A)

### Popis kritéria:

Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3.

### Komentář:

Práce je psána anglickým jazykem. V textu se vyskytuje několik překlepů, občas chybějí členy. Nic z toho však nebrání pochopení práce.

### Hodnotící kritérium:

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

## 6. Práce se zdroji

100 (A)

### Popis kritéria:

Vyjádřete se k úrovni dosažených hlavních výsledků ZP, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, apod. Případně také zhodnoťte, zda software nebo zdrojové texty, které nevytvořil sám student, byly v ZP použity v souladu s licenčními podmínkami a autorským právem. a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

### Komentář:

V pořádku.

### Hodnotící kritérium:

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

## 7. Hodnocení výsledků, publikační výstupy a ocenění

85 (B)

### Popis kritéria:

Vyjádřete se k úrovni dosažených hlavních výsledků ZP, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, apod. Případně také zhodnoťte, zda software nebo zdrojové texty, které nevytvořil sám student, byly v ZP použity v souladu s licenčními podmínkami a autorským právem. Popište případnou publikační činnost a získaná ocenění související s řešením této ZP.

### Komentář:

Dosažené teoretické výsledky jsou v pořádku.

V experimentálním testování v kapitole 6 by bylo vhodné použít i datasey s jinou strukturou.

Autorka se, bohužel, rozhodla rozšířit implementaci z [6], která, přinejmenším, není dokonalá. Je škoda, že si autorka nenapsala vlastní implementaci, která by např. využila lepších možností reprezentace automatu. Takto je to jen proof-of-concept, což ale nevádí.

V dodaném zdrojovém kódu se počítá se stejnou XML strukturou jakou mají zvolené datasey, neboť XPath dotazy jsou hardcoded, stejně jako zvolený typ indexu (TPA\*).

V příložených zdrojových kódech jsem nenašel žádné testy, které by se daly jednoduše znovu spustit a ověřit tak, že XPath dotazy nad automatovým indexem vrací správné výsledky.

### Hodnotící kritérium:

Způsob hodnocení - nehodnotí se

## 8. Komentář o využitelnosti výsledků

### Popis kritéria:

Uveďte, zda hlavní výsledky ZP rozšiřují již publikované známé výsledky a/nebo přinášející zcela nové poznatky. Uveďte možnosti využití výsledků ZP v praxi.

### Komentář:

Práce je jistě zajímavým teoretickým rozšířením projektu indexování. Z hlediska asymptotické velikosti indexu vzhledem k velikosti XML stromu je otázkou, jestli bude moct být využíváno v praxi.

### Hodnotící kritérium:

Způsob hodnocení - nehodnotí se

## 9. Otázky k obhajobě

### Popis kritéria:

Uveďte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odrážkami).

### Otázky:

- Jak moc pomohl času konstrukce TPA\* paralelní přístup? Kolik mělo testovací prostředí k dispozici vláken procesoru?
- Jak jste implementovala testování, zda vaše implementace vrací správné výsledky?

### Hodnotící kritérium:

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

## 10. Celkové hodnocení

92 (A)

### Popis kritéria:

Shrňte stránku ZP studenta, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení **nesmí** být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích 1 až 9.

### Text hodnocení:

Největší váhu přikládám teoretické části práce, která je i přes drobné formální chyby splněna výborně. Studentka vylepšila i další index TPA a to nad rámec zadání.

Dodaná proof-of-concept implementace je v pořádku.

V souladu s předchozími body hodnocení navrhuji známku A.

Podpis oponenta práce: