



# Posudek oponenta závěrečné práce

**Student:** Yuliia Syzon  
**Oponent práce:** Mgr. Eva Pernecká, Ph.D.  
**Název práce:** Deduplikace dat o veřejných zakázkách  
**Obor:** Znalostní inženýrství

**Datum vytvoření:** 15. 6. 2020

<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – následující škálou 1 až 4:</i>
<b>1. Splnění zadání</b>	<b><u>1=zadání splněno,</u> 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno</b>
<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> Závěrečná práce jasně formuluje a zcela naplňuje cíle vymezené v zadání. Byla provedena adekvátně podrobná studie existujících metod propojení záznamů a důsledná analýza zpracovávaného problému, na jejichž základě byl navržen, implementován a testován odpovídající algoritmus.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
<b>2. Písemná část práce</b>	<b>94 (A)</b>
<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 vyskytl-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.	
<i>Komentář:</i> Závěrečná práce je přehledně a logicky členěna, úhledně sepsána a obsahuje všechny náležitosti, včetně řádných citací.  Ráda bych zejména vyzdvihla vysokou úroveň psaného textu, jeho srozumitelnost, čitelnost a logickou strukturu. Narazila jsem pouze na občasné překlepy (např. str. 31, řádek 14) a chyby ve slovosledu (např. str. 75, řádek 5).  Uvítala bych však podrobnější a přesnější vysvětlení některých metod a algoritmů (např. v Jaro-Winkler algoritmu chybí řádná definice shodných znaků a transpozice, nebo v metodě pravděpodobnostní klasifikace nejsou uvedené vzorce a v nich se vyskytující parametry jasně zavedeny). Naopak, některé popisné části mohly být sestručněny.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
<b>3. Nepísemná část, přílohy</b>	<b>100 (A)</b>
<i>Popis kritéria:</i> Dle charakteru práce se případně vyjádřete k nepísemné 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	
<i>Komentář:</i> Autorka mi osobně prezentovala implementaci algoritmu a běh aplikace. Všechno fungovalo správně a v souladu s písemnou částí.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
<b>4. Hodnocení výsledků, jejich využitelnost</b>	<b>100 (A)</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:**

Výstupem této práce je algoritmus pro propojování záznamů o veřejných zakázkách, který vyhovuje specifikám zpracovávaných dat, a jeho funkční implementace. Vylepšuje již existující částečné řešení tohoto problému - umožňuje práci s několika páry zdrojů na rozdíl od původního pouze jednoho páru, zvyšuje úspěšnost procesu z původních 0,77 na 0,93, a je přizpůsoben pro práci s velkými objemy dat prostřednictvím postupného zpracování záznamů. K nalezení vhodného parametrického nastavení byl použit genetický algoritmus a bylo provedeno několik experimentů k ověření vhodnosti výběru poměřujících atributů. Navržený model a jeho případné adaptace tedy umožňují využití v praxi.

*Hodnotící kritérium:*

*Způsob hodnocení – nehodnotí se*

## 5. Otázky k obhajobě

**Popis kritéria:**

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

**Otázky:**

Zohledňuje některý z provedených experimentů (sekce 6.3) také nějaké změny v kroku indexování? Pokud ne, uvažovali jste o experimentu zaměřeném na ladění výběru blokovacích klíčů a jejich metrik porovnání?

*Hodnotící kritérium:*

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

## 6. Celkové hodnocení

97 (A)

**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:**

Práce byla vypracována se zřejmým zájmem a důkladností. Výsledkem je použitelný algoritmus, který plně splňuje stanovené požadavky, a kvalitní písemná dokumentace.

Podpis oponenta práce: