



Posudek oponenta závěrečné práce

Student: Róbert Schönfeld
Oponent práce: Ing. Karel Klouda, Ph.D.
Název práce: Autocomplete in the system TA ČR Starfos
Obor: Znalostní inženýrství

Datum vytvoření: 28. 1. 2020

Hodnotící kritérium:	Způsob hodnocení – následující škálou 1 až 4:
1. Splnění zadání	1=zadání splněno, 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno
Popis kritéria: 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.	
Komentář: Cílem práce bylo vytvořit tzv. našeptávač při vyhledávání nad databází projektů a výsledků v TAČR, přesněji v jejich existujícím systému starfos.tacr.cz. Z pěti bodů zadání nebyly všechny splněny dostatečně. Bod první, tedy popis systému a problému našeptávání v případě systému starfos, byl splněn, a to navíc velmi dobře. Rešeršní část je nejsilnější složkou hodnocené práce. Bod druhý, týkající se extrakce klíčových slov, byl splněn částečně. Rešerše metod byla provedena, ale následně byl zvolen ten nejjednodušší způsob extrakce klíčových slov, tedy poměrně prosté využití klíčových slov zadaných autory publikací a projektů. To není nutně špatně, a v práci je uvedena a nějaká argumentace, proč se zůstalo u tohoto způsobu. Vše ale bylo uděláno jen pro anglický jazyk, nikoli i pro český, jak požaduje zadání. I to je v práci vysvětleno, ale už méně přesvědčivě. Bod čtvrtý byl splněn opět minimálně, v práci navíc není moc podrobně popsáno, jak probíhal ranking získaných návrhů. Rozhodně nebyla ale provedena požadovaná evaluace. Bod pět není v práci zmiňován a moc nerozumím tomu, co přesně požadoval.	
Hodnotící kritérium:	Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):
2. Písemná část práce	70 (C)
Popis kritéria: 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: Práce je psána poměrně dobrou angličtinou a první tři kapitoly se i dobře čtou. Napsané je podloženo velkým množstvím relevantních referencí a z práce se tak čtenář dozví hodně zajímavých pohledů na zdánlivě jednoduchý formulář pro fulltextové vyhledávání. Čtvrtá kapitola popisující implementaci a výsledky je ale velice odbytá, čtenář se nedozví skoro vůbec nic o tom, co bylo vytvořeno. I když nějaký software vznikl, o jeho fungování není v práci skoro nic. Lépe je popsána snad jen extrakce klíčových slov z řetězců obsahující klíčová slova zadaná autory. Typograficky je práce celkem v pořádku, ale dají se najít četné chyby a překlepy (formátování seznamů, odkazy na obrázky, atp.)	
Hodnotící kritérium:	Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):
3. Nepísemná část, přílohy	50 (E)
Popis kritéria: 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	

Komentář:

Přiložené kódy jsou neokomentované, ani z práce není jasné, kde lze najít jakou funkcionalitu, jak se kód spouští či instaluje, nebo jakou má architekturu. Je tak velmi obtížné výsledky hodnotit.

Hodnotící kritérium:

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

4. Hodnocení výsledků, jejich využitelnost

50 (E)

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:

Jak píše výše, v práci jsou výsledky víceméně utajeny, takže je těžké je hodnotit. Z toho, co lze pochopit, autor implementoval našeptávač pro fulltextové vyhledávání s velmi základními funkcemi. Oproti žádnému našeptávání se ale zřejmě jedná o pokrok, který může skončit nasazením do ostrého provozu.

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 odřádkami).

Otázky:

- 1) Proč jste použil zrovna Apache Solr? Nedalo by se například použít Elasticsearch?
- 2) Jak byste si poradil s extrakcí českých klíčových slov z anglicky psaných textů?

Hodnotící kritérium:

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

6. Celkové hodnocení

53 (E)

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:

Vzhledem k výše uvedenému navrhuji práci hodnotit známkou E. Zjevně se nepodařilo vše dokončit včas a práce je tím poznamenána. Z úvodních kapitol je přitom zřejmé, že autor je schopen dostat se k výrazně lepším výsledkům.

Podpis oponenta práce: