



POSUDEK OPONENTA BAKALÁŘSKÉ PRÁCE

Jméno studenta: Lukáš Kratochvíl
Název práce: Optimalizace fulltextového vyhledávání
Autor posudku: Ondřej Klapka
Cíl práce: Prozkoumat a otestovat možnosti optimalizace technologie Elasticsearch, pracující nad knihovnou Apache Lucene.

Povinná kritéria hodnocení práce	Stupeň hodnocení (známka)					
	A	B	C	D	E	F
Práce svým zaměřením odpovídá studovanému oboru	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vymezení cíle a jeho naplnění	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zpracování teoretických aspektů tématu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zpracování praktických aspektů tématu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Adekvátnost použitých metod, způsob jejich použití	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Hloubka a správnost provedené analýzy	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Práce s literaturou	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Logická stavba a členění práce	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jazyková a terminologická úroveň	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Formální úprava a náležitosti práce	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vlastní přínos studenta	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Využitelnost výsledků práce v teorii (v praxi)	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Díličí připomínky a náměty:

- V první části práce jsou popsány některé základní principy fulltextového vyhledávání, avšak konkrétní principy jednotlivých algoritmů nejsou v textu příliš rozvedeny – zejména fragmentační algoritmy, které, jak autor sám v textu uvádí, jsou moderními vyhledávacími nástroji hojně využívány.
- V rámci testování navrženého řešení byla dle textu práce „zaindexována reálná webová aplikace obsahující softwarovou dokumentaci čítající 50 stránek“. V textu ovšem není naznačen alespoň přibližný rozsah jednotlivých stránek nebo struktura dokumentace. Získat konkrétní představu o rozsahu indexovaných dat je tak poněkud problematické.

Celkové posouzení práce a zdůvodnění výsledné známky:

Práce je praktického zaměření a vztahuje se ke studovanému oboru studenta. Formální náležitosti práce jsou dodrženy, práce se zabývá netriviální a velmi aktuální problematikou, množství odkazů na literaturu lze považovat za dostačující. V první části práce jsou popsány užívané technologie pro fulltextové vyhledávání. Dále jsou představeny technologie Apache Lucene a Elasticsearch. Druhá část práce začíná poněkud nestandardně definováním cíle práce (mělo by být již v úvodu práce) a návrhem vlastních optimalizací pro technologii

Elasticsearch. Navržené optimalizace jsou následně testovány na reálné dokumentaci webové aplikace s využitím vlastní metodiky založené na několika kategoriích dotazů. Výsledná data jsou statisticky zpracována a z nich vytvořeny závěry, že navržené optimalizace lze považovat za přínosné. Vzhledem k tomu, že nejsou dostatečně představena indexovaná data a rovněž není zcela zřejmé, jakým způsobem byla hodnocena správnost vrácených výsledků (zejména jak byl sestaven rejstřík očekávaných vrácených výsledků pro jednotlivé vyhledávací fráze), nelze v práci uvedené závěry ověřit.

Otázky k obhajobě:

- Představte konkrétněji indexovanou dokumentaci – zejména její rozsah a dále množství položených dotazů v jednotlivých sledovaných kategoriích (viz. kapitola 7.1)
- V práci vyhodnocujete přesnost (relevanci) vrácených výsledků vůči vzorovému přiřazení očekávaných výsledků k dané vyhledávané frázi. Rozved'te, jakým způsobem (popř. kým) byl rejstřík očekávaných výsledků pro jednotlivé vyhledávané fráze sestaven.

Práci doporučuji k obhajobě.

Navržená výsledná známka: C

V Hradci Králové, dne 19. 5. 2018

podpis