



## POSUDEK OPONENTA BAKALÁŘSKÉ PRÁCE

**Jméno studenta:** Matěj Fries

**Název práce:** Sémantická analýza textů

**Autor posudku:** Ing. Karel Mls, Ph.D.

**Cíl práce:** Popsat základní způsoby sémantické analýzy textů. Vytvoření aplikace na zpracování lematizovaných textů založené na popsanych algoritmech a datových reprezentacích.

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

### Dílicí připomínky a náměty:

Konkrétní drobné připomínky k formální stránce textu: pravopisné chyby, vyskytující se zejména v závěrečné části práce (algoritmus latentní Dirichletovi alokace, ...obsahují společné významová slova, ...používá náhodných dat, a ty mohou ovlivnit). V práci zcela chybí popis vytvořené aplikace, což byl jeden z hlavních praktických cílů; popis zdrojových kódů na jedné straně a Návod k programu na dvou stranách příloh nepovažuji za dostatečný. Rovněž v Závěru autor víceméně rekapituluje představené metody a modely reprezentace dokumentů a jejich výhody, resp. nevýhody, ale prakticky nezmiňuje vlastní (programátorský) přínos.

### **Celkové posouzení práce a zdůvodnění výsledné známky:**

Autor představuje řadu existujících matematických metod používaných při analýzách přirozených textů, specificky v českém jazyce; nezabývá se však současným stavem software pro danou oblast (text-tools, Gensim od Radima Řehůřka,...). V praktické části nazvané Ukázka na datech jsou prezentovány číselné výsledky analýz – míry podobnosti vybraných dvojic dokumentů pro různé metody a jejich ad hoc nastavení. Autor bohužel nepostupuje systematicky a jednotlivé výsledky hodnotí subjektivně, není jasné vzhledem k čemu ani podle jaké metriky. Závěrečná práce až na drobné pravopisné nedostatky splňuje požadavky Metodických pokynů... Stanovený cíl práce byl splněn, avšak praktická část, týkající se vytvoření aplikace na zpracování textů, není v práci zdokumentována. Téma považuji za zajímavé a výsledky použitelné v praxi, vzhledem k celkovému posouzení práce navrhuji hodnocení stupněm D.

### **Otázky k obhajobě:**

1. Proč jste pro realizaci aplikace zvolil právě jazyk Java?
2. Podle jakého klíče jste volil konkrétní porovnávané dvojice textů a nastavení parametrů (dimenzí, témat) jednotlivých metod v praktické části?
3. Jak byste lépe a systematictěji prezentoval „čarování“ s hodnotami při hodnocení kvality výsledků v části 5? Podle jakých kritérií usuzujete, který výsledek je lepší?

**Práci doporučuji k obhajobě.**

**Navržená výsledná známka: D - velmi dobře-dobře**

**V Hradci Králové, dne 26. května 2017**

---

**podpis**