



## POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

**Jméno studenta:** Bc. David Illner

**Název práce:** Extrakce témat z nestrukturovaného textu

**Autor posudku:** Mgr. Jiří Haviger, Ph.D.

**Cíl práce:** Cílem práce je vytvořit model pro extrakci témat z nestrukturovaného textu v analytické platformě KNIME a zhodnotit výsledky tohoto modelu.

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 checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zpracování teoretických aspektů tématu	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zpracování praktických aspektů tématu	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Adekvátnost použitých metod, způsob jejich použití	<input checked="" type="checkbox"/>	<input 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 type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Práce s literaturou	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Logická stavba a členění práce	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" 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 checked="" type="checkbox"/>	<input type="checkbox"/>	<input 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 checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

### Vyjádření k výsledku anti-plagiátorské kontroly

Celková podobnost 0%.

### Díličí připomínky a náměty:

Kapitoly po úvodu

2. Cíl práce

3. Základy textové analýzy

4. Základy dolování z textů

5. Zpracování přirozeného jazyka

6. Detekce témat

7. Analytická platforma KNIME

8. Modelování extrakce témat

9. Závěr

Strukturu nepovažuji za ideální – není jasné, kde končí rešeršní část pro zjištění současného stavu poznání a kde začíná popis vlastního výzkumu. Není jasné, ve které kapitole jsou popsány datové sady na kterých bude extrakce probíhat, není jasné, kde je popsáno jak budou výsledky

evaluovány. Doporučil bych kapitoly 3-6 sloučit do jedné (current state of knowldge), naopak kapitolu 8 rozdělit na dvě části:

1/ část obsahující jaké texty se zpracovaly, jakými metodami se témata extrahovaly a jak se výsledky evaluovaly - metriky a kritéria kvality (methods) a

2/ část popisující co vyšlo, grafy, tabulky, wolrdcloudy, výsledky evaluace (results).

Obsahově pak není zřejmé, jak se některé kapitoly (3.1 analýza konverzace, 5.3 NER, ...) promítají do vlastního výzkumu extrakce témat.

Kladně hodnotím popis KNIME (str 35-39) a to jak obsah, tak rozsah.

Dále kladně hodnotím i samotný model a popis jeho vytvoření v prostředí KNIME.

Jinak celkový model dosahuje úspěšnosti 67% což jak sám autor přiznává není ideální, ale srovnává jej s citačně podpořeným standardem 70%. Pro DP bych však doporučil porovnání dvou různých modelů, nebo tohoto modelu s různým nastavením parametrů  $\alpha, \beta$

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

Práce splňuje požadavky kladené na diplomovou práci.

### **Otázky k obhajobě:**

Jaké jsou alternativní moderní přístupy k extrakci témat?

**Práci doporučuji k obhajobě.**

**Navržená výsledná známka: B**

**V Hradci Králové, dne 14. května 2024**

---

**podpis**