



## POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

**Jméno studenta:** Bc. Radek Salay

**Název práce:** Zpracování nestrukturovaných dat pomocí IBM Watson.

**Autor posudku:** Ing. Tomáš Nacházel, Ph.D.

**Cíl práce:** Diplomová práce seznámí čtenáře s problematikou zpracování nestrukturovaného textu. Dále práce představí produkt Watson a jeho podprodukty pro zpracování anglického jazyka. V závěru práce bude ukázané praktické využití těchto poznatků v knihovní aplikaci.

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 checked="" type="checkbox"/>	<input checked="" 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 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 type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" 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 type="checkbox"/>	<input checked="" 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 type="checkbox"/>	<input type="checkbox"/>	<input checked="" 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"/>

### Dílicí připomínky a náměty:

Anglická verze anotace má velmi špatnou jazykovou úroveň.

Ilustrační věty a příklady měly být spíše v uvozovkách než pouze kurzívou, aby se lépe oddělily od ostatního textu.

Celá kapitola 6 cituje téměř pouze z jednoho zdroje a za úplně každým odstavcem. Bylo by vhodnější citace lépe zakomponovat do textu. To samé se týká i kapitoly 7.

Překlepy a chyby - např. strana 15 („modlu“), 22 (chybí čárka), 23 („pod typ“), 32 („IBM Cloude“), 50 („Blížší seznáme se třídou je níže v textu.“), 69 (malé písmeno u příjmení autorky [4]) a tak podobně.

Nadužívání poznámek pod čarou – např. zkratku GPE by bylo vhodnější vysvětlit přímo v textu než poznámkou pod čarou; když už jsou ve zdrojích stránky webu cloud.ibm.com (18, 20), mohly tam být i odkazy z poznámek pod čarou 10 a 11.

Bylo by vhodné zkratku jednou uvést (v textu, ne poznámkou pod čarou) a pak ji používat, v kapitole 6.1 byla představena zkratka NLP, na začátku kapitoly 7 začala být používána nezkrácená verze a od 7.6 je znovu využívána zkratka.

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

Autor se zabýval zpracováním textu pomocí platformy IBM Watson. Práce je strukturovaná přehledně a má odpovídající jazykovou úroveň.

V teoretické části (3. až 7. kapitola) autor seznamuje čtenáře s historií a obecnými základy zpracování textu a představuje platformu IBM Knowledge Studio se službou Natural Language Understanding. V této části mi chybí alespoň stručná kapitola o ontologiích a expertních systémech, jelikož výsledné modely odpovídají strukturou právě jim. Pojmy jako entita a vazba jsou představeny teprve až u konkrétní platformy. Také zde chybí alespoň stručná charakteristika algoritmu pro vytváření modelu, například zda využívá klasifikátor, neuronovou síť nebo jiné metody.

Hlavním problémem teoretické části je ovšem citování převážně webových stránek, diplomová práce by měla čerpat z odbornějších zdrojů. U kapitoly popisující software a odkazující se na dokumentaci je to samozřejmě pochopitelné, ale články na webu by neměly být jedním z hlavních zdrojů pro kapitoly teoretické části práce.

Původním záměrem praktické části práce bylo vytvoření vlastního modelu, který by sloužil k analýze knih. Tento záměr se podle autora později ukázal jako příliš náročný a neúměrně pracný pro potřeby diplomové práce. S tímto závěrem souhlasím – původní cíl byl příliš ambiciózní. Místo vlastního modelu byl tedy využit existující obecný model od IBM. Vlastní přínos práce je tak omezený na aplikaci využívající služby a model od IBM.

V rámci zdůvodnění, proč nebyl vytvořen vlastní model, se píše „Výsledný model byl tedy velice obecný, a přesto nedosahoval kvality obecného modelu IBM.“ Znamená to tedy, že nějaký model přeci jen vytvořený byl, pouze nedosahoval dostatečné kvality? Myslím, že alespoň menší ukázka a zhodnocení vytvořeného modelu by bylo zajímavé i přesto, že se model nakonec ukázal jako nedostatečný.

Téma a cíl práce odpovídá oboru aplikovaná informatika. Práci doporučuji k obhajobě.

#### **Otázky k obhajobě:**

- 1) Jak jsou naučené modely v IBM Watson závislé na jazyce textu? Předpokládám, že by vytvořená aplikace nedokázala zpracovat i text v češtině, je to tak? Nebo je to jen otázka použití jiného slovníku?
- 2) Myslíte, že kdyby bylo pro vlastní model vybráno nějaké užší zaměření textů (místo obecné knihovny např. jen články o zpracování textu), bylo by jeho vytvoření reálné, anebo by to i tak překračovalo rámec diplomové práce kvůli pracnosti anotování textů?

**Práci doporučuji k obhajobě.**

**Navržená výsledná známka: C**

**V Hradci Králové, dne 29. prosince 2020**



---

**podpis**