

# Oponentský posudek habilitační práce

Autor práce: Ing. Marián Šimko, Ph.D.

Název práce: **Modelovanie a extrakcia informácií pre inteligentný (slovenský) web**

## Aktuálnosť tématu habilitační práce

Předložená habilitační práce se zabývá extrakcí informace z webu se zaměřením na informace sloužící k vytváření modelů/reprezentací sloužících k porozumění obsahu webu. Mezi takové modely patří doménové ontologie, které mohou být využívány inteligentními informačními systémy. Práce poskytuje přehled přístupů k řešení ve všech základních etapách extrakce informací pro využití inteligentními informačními systémy, včetně způsobů reprezentace extrahovaných informací. V návaznosti na soudobé prostředky vytváření obsahu webu jde nesporně o problematiku aktuální. Navíc zaměření se na jazykové prostředí slovenštiny tuto aktuálnost ještě zvyšuje.

## Dosažené výsledky a přínos práce

V práci autor prokazuje velmi dobrou orientaci ve všech klíčových oblastech souvisejících s extrakcí informací pro využití inteligentními informačními systémy. Jde zejména o oblast konceptualizace a přístupů k reprezentaci sémantiky, zpracování přirozeného jazyka pro extrakci informací a samotnou extrakci a budování ontologických struktur. Výsledky své výzkumné činnosti publikoval jak ve kvalitních časopisech, tak na mezinárodních konferencích. O přínosech těchto výsledků svědčí i řada citací evidovaných ve WoS a Scopus. Ve srovnání s tím považuji přínos samotné habilitační práce za nižší. Autor podle mne mohl do práce v podstatně větší míře promítnout vlastní výsledky. Nepříliš velký rozsah 57 stran bez referencí k tomu dával dostatečný prostor. Případně mohl alespoň klíčové práce dát do přílohy.

## Formální stránka a struktura práce

Habilitační práce je členěna do šesti kapitol. V úvodu autor neformálně vysvětluje motivaci výzkumu v oblasti extrakce informací pro využití inteligentními informačními systémy. Cíl práce explicitně stanovený není, ale z uvedeného jej lze chápat jako „představení automatizované extrakce v širším kontextu inteligentních informačních systémů“. Tomu odpovídají i následující kapitoly pokrývající klíčové kroky procesu extrakce.

Druhá kapitola blíže zavádí některé základní pojmy související se symbolickou reprezentací informací a konceptualizací, podrobněji vysvětluje pojem inteligentní webový informační systém a uvádí příklady takových systémů a jejich architektur. V závěru kapitoly vysvětluje význam extrakce informací pro tvorbu doménového modelu využívaného inteligentními webovými informačními systémy.

Třetí kapitola je věnována ontologiím jako prostředku konceptualizace. Provádí jejich kategorizaci a uvádí formální definice ontologie při přístupu založeném na těžké a lehké sémantice. Ve druhém případě prezentuje vlastní přístup použitý u adaptivního webového vzdělávacího systému ALEF. Je to jedno z mála míst, kde je prezentován vlastní přístup ve větším rozsahu. V závěru kapitoly je uveden souhrn příspěvku autora v této oblasti zahrnující několik publikací. Jedno- až dvouřádková charakteristika příspěvku ale mnoho nevyovídá a posouzení by vyžadovalo si publikace vyhledat.

Čtvrtá kapitola se zabývá zpracováním přirozeného jazyka pro potřeby extrakce informací. Autor zde stručně prezentuje základní kroky předzpracování textu a přístupy k jazykové analýze. Odkazuje se zde na výsledky některých závěrečných prací z oblasti zpracování přirozeného jazyka, které vedl. Hlavní oblasti jsou opět sumarizovány v závěru kapitoly. Vzhledem k určitému deficitu prostředků pro

zpracování textů ve slovenštině je třeba hodnotit aktivity Ing. Šimka, Ph.D. a jím vedených studentů, prezentované na portálu *text.fiit.stuba.sk*, velmi pozitivně.

Pátá kapitola je věnována samotné extrakci informací z textu pro účely budování ontologií. Autor zde poskytuje přehled dílčích úloh a přístupů k jejich řešení. Zaměřuje se zejména na automatický přístup k tvorbě ontologií, možnost využití různých forem obsahu webu vytvářeného uživateli a na podporu ruční tvorby ontologií. Jedna podkapitola je zde věnována přístupům a mírám používaným k hodnocení kvality vytvořených ontologií a metod jejich vytváření. I v této páté kapitole autor na několika místech zmiňuje relevantní publikace, kde je autorem nebo spoluautorem, a na závěr kapitoly svůj příspěvek sumarizuje výčtem čtyř oblastí a odpovídajících publikací.

Šestá kapitola je ve skutečnosti závěrem práce, kde autor vymezuje několik perspektivních oblastí výzkumu souvisejících s tématem habilitační práce a hodnotí stručně aktuální stav v nich.

Habilitační práce má dvě přílohy. První představuje úvod k portálu zaměřenému na zpracování textu *text.fiit.stuba.sk*. Ve druhé autor shrnuje podstatné informace o své osobě vztahující se k profesní kariéře, pedagogické a výzkumné činnosti.

Celkově mohu konstatovat, že kapitoly jsou uspořádány logicky, dobře na sebe navazují a jejich obsah je až na některé případy zmíněné níže dobře srozumitelný. To svědčí o dobrých pedagogických schopnostech autora.

Počet překlepů a nepřesností, které jsem v práci zjistil, je nízký. Jako příklad lze uvést Definici 2 na str. 18, kde je uvedena množina  $T_C$ , naopak chybí  $T_S$  z Definice 2.2. Pravděpodobně jde o překlep. V některých případech převzatých obrázků, tabulek či algoritmu bez komentáře je porozumění obtížné. Příkladem může být Tabulka 2 na str. 40. Podobně u Algoritmu 1 na str. 43 by upozornění na použití značení z Definice 2.2 na str. 17 určitě usnadnilo pochopení. Dalším příkladem je Algoritmus 2 na str. 45 a obr. 14.

### **Publikace**

Jak již jsem uvedl v hodnocení přínosu, výsledky své výzkumné činnosti publikoval autor jak ve kvalitních časopisech, tak na mezinárodních konferencích. Na ně je evidována řada ohlasů v podobě citací.

### **Otázky do diskuse**

1. Jakým způsobem byl vytvářen doménový model systému ALEF a jak v něm byl využíván?
2. Jednou z výzev v oblasti habilitační práce, kterou uvádíte na str. 57, je zkoumání tvůrců struktur informačních struktur. Jakým způsobem může přispět sledování pohledu, kterým jste se zabývali v publikaci (Šimko, Bieliková, 2015)?

### **Závěr a doporučení**

Závěrem konstatuji, že předložená habilitační práce Ing. Mariána Šimka, Ph.D. v kontextu jeho dalších publikací tvoří ucelené dílo spadající do oboru informačních systémů a že považuji autora za uznávanou vědeckou osobnost v tomto oboru. Splňuje tak podle mého názoru podmínky vedoucí k udělení vědecko-pedagogického titulu docent. Proto práci k obhajobě doporučuji a po úspěšné obhajobě navrhuji tento titul udělit.

V Brně dne 13. 12. 2016

Doc. Ing. Jaroslav Zendulka, CSc.  
Fakulta informačních technologií VUT v Brně