

## Oponentský posudok habilitačnej práce

**Ing. Marián Šimko, PhD.**

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

Študijný odbor: **9.2.5 Informačné systémy**

Pracovisko: **Fakulta informatiky a informačných technológií, Slovenská technická univerzita v Bratislave**

Posudzovaná habilitačná práca prezentuje pomerne ucelený pohľad na problematiku modelovania a extrakcie informácií v kontexte inteligentných informačných systémov, s prihliadnutím na slovenčinu, čo je obzvlášť cenné v našom výskumnom priestore. Súčasťou práce sú aj dve prílohy. Príloha A stručne popisuje náplň portálu *text.fiit.stuba.sk* prezentujúceho praktické výsledky prevažne záverečných prác vedených habilitantom - Ing. Mariánom Šimkom, PhD. Príloha B uvádza základné údaje o pedagogickej a výskumnej činnosti habilitanta.

Téma habilitačnej práce je aktuálna najmä zásluhou širšieho kontextu, v ktorom sú jednotlivé čiastkové úlohy prezentované. Viaceré prezentované úlohy sú totiž študované už dlhšiu dobu (pedspracovanie textu, extrakcia informácií, semi-automatizovaná tvorba ontológií), ale autor ich približuje z pohľadu aktuálneho stavu poznania vo svete a v kontexte významných fenoménov posledných rokov (napr. používatel'mi generovaný obsah v sociálnych médiách), a najmä vlastných prínosov k riešeniu jednotlivých čiastkových vedeckých problémov.

Použitý spôsob prezentácie výsledkov svojej doterajšej vedecko-výskumnej práce považujem za veľmi vhodný. Autorovi sa v jeho habilitačnej práci podarilo podať stručný, ale výborný prehľad širokej skupiny úloh riešených v súvislosti so snahou čo najviac automatizovať extrakciu informácií z rôznych typov textových (najmä webových) zdrojov. Páči sa mi šírka záberu habilitanta aj jeho snaha prispieť vlastnými výsledkami výskumu do skoro všetkých prezentovaných typov úloh. Svoje vlastné výsledky zakaždým zhrnul na záver jednotlivých kapitol spolu s odkazmi na publikované práce, v drvivej väčšine medzinárodne uznávanej kvality (B) a v štyroch prípadoch dokonca špičkovej medzinárodnej kvality (A).

Úroveň textu habilitačnej práce napovedá, že autor má výborné predpoklady pre pedagogickú prácu. Oceňujem najmä kultivovaný jazykový prejav a dôsledné používanie slovenskej terminológie. Aj keď si viem predstaviť v niektorých prípadoch iný slovenský ekvivalent prekladaného pojmu (napr. "pomenované entity" namiesto "vlastných pomenovaní" alebo "návratnosť" namiesto "pokrytie", nakoľko existuje aj metrika *coverage*, ktorá ale znamená niečo iné ako *recall*). V niektorých momentoch uvádza viaceré vo svete používané prístupy bez priamej referencie na zdroj, v ktorom by sa čitateľ mohol dočítať podrobnosti, vrátane úspešnosti spomínaných prístupov (obzvlášť výrazné je to v časti identifikácia viac-slovných výrazov, ale občas sa to vyskytuje aj v iných častiach práce). Uvedené nedostatky sú však drobnosti, ktoré v žiadnom prípade neznižujú celkovo výborný dojem z tejto habilitačnej práce.

Veľmi oceňujem aj sprístupnenie praktických výsledkov výskumu a pedagogickej činnosti vo forme vedenia záverečných prác Ing. Mariána Šimka, PhD. na portáli <http://text.fiit.stuba.sk/>.

K práci mám niekoľko otázok:

Autor terminologicky oddeľuje pojem ontológia a znalostná báza, pričom využíva príklad znázornený na obrázku 5. Je to naozaj tak, že v konkrétnej aplikácii máte v znalostnej báze iba inštancie konceptov a relácií určitej ontológie, bez špecifikácie nej samotnej?

Na s. 16 autor uvádza, že slovníky môžu byť použité napr. na triviálny výpočet podobnosti medzi jednotlivými termami. Mohli by ste popísať, aký spôsob výpočtu podobnosti ste tu mali na mysli?

V časti perspektívy extrakcie informácií autor spomína ako prvú oblasť extrakciu vyššej sémantiky. Trocha mi v práci absentuje zmienka o početných aktivitách v oblasti vývoja "ťažkých" ontológií z prelomu milénia a prvej dekády tohto storočia, kedy bežalo aj viacero zaujímavých európskych projektov. Za všetky spomeniem NeOn. Zaujímal by ma názor habilitanta na výsledky tohto a prípadne iných podobných projektov.

Celkovo po oboznámení sa s habilitačnou prácou Ing. Mariána Šimka, PhD. a jej prílohami, ktoré uvádzajú prehľad jeho pedagogickej a výskumnej činnosti, ako aj zoznam publikácií a ich citačný ohlas môžem konštatovať, že Ing. Marián Šimko, PhD. je odborne fundovanou osobnosťou v odbore Informačné systémy, s významnými vedeckými výsledkami a pedagogickými schopnosťami. V zmysle vyššie uvedeného preto

**odporúčam,**

aby Ing. Marián Šimko, PhD. bol vymenovaný za docenta v študijnom odbore Informačné systémy.

V Košiciach, 20. januára 2017

prof. Ing. Ján Paralič, PhD.  
Katedra kybernetiky a umelej inteligencie  
Technická univerzita v Košiciach  
Letná 9, 042 00 Košice