

# Oponentský posudek dizertační práce

## Výzkum metod optimalizace Sémantického webu s využitím ontologií

Autor disertační práce: Ing. Michal Šmiraus  
Vedoucí práce: doc. Ing. Zdenka Prokopová, CSc.  
Oponent práce: doc. Mgr. Roman Jašek, Ph.D.

Posuzované téma práce odráží velmi významný a strategický směr vyhledávání a analýzu dat, zasahující přímo podstatu globální internetové sítě a to World Wide Web. Autor práce se zabývá problematikou strojového vyhledávání smyslu obsahu různě strukturovaných dat a jejich souvislostí.

**Význam a aktuálnost práce spočívá na využití znalostních bází, spojených s vývojem obecných jazyků za využití různých postupů a metod na straně jedné a na druhé straně s konstrukcí ontologií popisujících věcné oblasti.**

Práce je sestavena z 6 kapitol v rozsahu 108 číslovaných stran.

První kapitola hodnotí současný stav a popisuje stávající a sémantický web. V oblasti stávajícího webu zdůrazňuje schéma specifikace URI identifikátorů (Uniform Resource Identifier) a sémantického webu jeho nahrazení modelem RDF (Resource Description Framework)

Druhá kapitola rozebírá cíle dizertační práce, které jsou v případě takto pojatého výzkumu vhodně formulované. Obzvláště oceňuji cíl označený jako návrh nástroje pro podporu vkládání sémantiky do obsahu webových dokumentů spojených s možnostmi zpracování dotazů v přirozeném jazyce. Diskutabilní je cílvytvoření komunity na základě dotazníkového šetření.

Třetí kapitola se zabývá teoretickými předpoklady práce, je dobře a přehledně zpracovaná, informace v ní předložené jsou relevantní k cílům práce. Kapitola je také dostatečně rozsáhlá pro poskytnutí teoretických znalostí následovníkům v uvedené oblasti aplikovaného výzkumu.

Čtvrtá kapitola je zaměřena do praktické části výzkumu a zaměřuje se na dosažení stanovených cílů práce. Vytvoření informačního portálu SemanticWeb je důležitým počinem, ale jeho výkonová odezva je bohužel velmi slabá (pokud má práce být živá a dále se rozvíjet, je třeba ji přesunout do výkonnějšího prostředí). Klíčovou část práce vnímám v kapitolách 4.3 - 4.5., kde autor vymezuje, popisuje a navrhuje konkrétní řešení. Grafická znázornění jsou uvedena v ucelených schématech 4.3 - 4.8 a grafech 4.7 - 4.8.

Pátá kapitola popisuje způsob řešení práce a zdůvodňuje které metody a proč byly dle autora práce použity.

Šestá kapitola popisuje přínosy pro vědu a praxi. Vědecký přínos autor staví na dotazníkovém šetření (což není příliš u této práce vhodné), daleko hodnotnější jsou statistická porovnání sémantických frameworků a reasonerů. Praktický přínos vývojem komponenty pro WYSIWYG editor autor zaplnil chybějící místo na "trhu" a prokázal tím

význam sémantického webu pro tvorbu a dolování dat ze znalostních bází jak intranetu, tak internetu.

Samotný závěr dizertační práce je zpracován srozumitelně a přehledně. **Jako významný výstup práce vnímám také jasný trend budování globálního webu strojově zpracovatelných dat s jejich srozumitelnou “již reálně vyhodnotitelnou” strojovou interpretací** (naštěstí bez sebeuvědomění sebe sama, což by zřejmě bylo jednou z vyhodnotitelných reálných civilizačních hrozeb).

## Připomínky

V práci se objevuje několik nepřesností a překlepů (např. titul školitelky Csc., editor WYSIWIG na str. 94 atd.)

U obrázků a schémat nebyl nikde uveden zdroj (vlastní / cizí?).

Schémata nemají jednotnou grafickou úpravu.

## Otázky

- Vysvětlete způsoby získávání dat ze sémantického webu v porovnání se současnými způsoby.
- Jakým způsobem bude vyhodnocována kvalita dat a získávaných informací ze sémantického webu?
- Kde je dle Vás hranice mezi inženýrským a vědeckovýznamným přístupem ve vaší práci?
- Vysvětlete možné způsoby testování a vyhodnocování výkonnosti výpočetního zdroje pro analýzu dotazu, která bude pravděpodobně velmi náročná. Nebude toto klíčovou překážkou pro budoucí nasazení vámi řešené technologie?

## Závěr

Předložená práce je zpracována na dostatečné úrovni, splňuje požadavky kladené na dizertační práci. Práce také nabízí možnosti návazných témat, která by mohla být naplní dalšího aplikovaného výzkumu v dané oblasti.

Práci **doporučuji** k obhajobě před příslušnou komisí a v případě úspěšného obhájení navrhuji Ing. Michalu Šmirausovi udělení vědecké hodnosti Ph.D..

Ve Zlíně dne 11.5. 2015

doc. Mgr. Roman Jašek, Ph.D.  
Ústav informatiky a umělé inteligence  
Fakulta aplikované informatiky  
Univerzita Tomáše Bati ve Zlíně

**Oponentní posudek závěrečné disertační práce**  
**Univerzita Tomáše Bati ve Zlíně,**  
**Fakulta aplikované informatiky**  
**Obor: Inženýrská informatika**  
**Školitelka: doc. Ing. Zdenka Prokopová, CSc.**

**Autor doktorské disertační práce: Ing. Michal Šmiraus**  
**Téma práce: „Výzkum metod optimalizace Sémantického webu s využitím ontologií“**

**Aktuálnost tématu disertační práce**

Předkládaná disertační práce, z pohledu vysoké dynamiky rozvoje vědy, techniky a aplikací prostředků informačních a komunikačních technologií (ICT), je vysoce aktuální a potřebná pro nové prostředí civilizovaného světa a nových technických, právních a dalších kyberprostorů používaných znalostních bází s precizním využitím moderních technologií Sémantického webu a ontologií v informační a znalostní společnosti.

Dále je aktuálnost řešených úkolů, vyplývajících z tématu disertační práce, dána systémově vymezeným pohledem na integrační procesy v oblasti vývoje obecných jazyků, metodik a nástrojů programového zabezpečení moderních ICT a konstrukcí samotných ontologií.

Proto uvedená řešení, této zajímavé oblasti výzkumu metod optimalizace, mají svá opodstatnění. Tuto práci chápu také jako podnět k řešení nových zajímavých a potřebných úkolů v uvedené oblasti pro inženýrskou informatiku.

Zpracované téma, této velmi aktuální a potřebné oblasti, je již vyjádřeno správným vymezením a hierarchickým členěním kapitol předkládané práce.

**Splnění stanovených cílů v disertační práci**

Cílem předkládané závěrečné práce jsou, vedle hlavního velmi náročného cíle, dále uvedeny dílčí cíle a postupy řešení vázané zejména na výzkumnou oblast a to správně vymezenými profily teoretické části, výzkumného profilu a praktické oblasti v implementační části softwarového nástroje pro sémantickou anotaci.

Komplexní pohled autora je vyjádřen v celé práci a svědčí o velmi kvalitním vedení disertanta zkušenou školitelkou a také o kvalitě a zaměření vědecké práce na tomto pracovišti.

Splnění stanovených cílů v disertační práci je vyjádřeno jednak stručně závěry provedenými z výzkumu a stručně interpretované v této práci.

Stanovené cíle byly splněny a hodnotím jejich průběžné a celkové splnění za pozitivní a přispívající k úspěšnému plnění dalších kritérií při hodnocení celé disertační práce.

**Metodologie a metody použité při vypracování disertační práce**

Celá metodologie výzkumu a vyjádření kvalitativních a kvantitativních údajů byla správně vedena především klasickými metodami vědecké práce a to na požadované úrovni.

Uvedené metody v práci byly postupně využívány, jak je velmi dobře vyjádřeno již v obsahu této práce. Pozornost autora byla především správně směřována k vyjádření technologií Sémantického webu, ontologii a znalostnímu inženýrství a uvedeným programovým nástrojům a sémantickým vyhledávačům.

Autor velmi stručně a přehledně vyjádřil, také v kapitole čtvrté, výsledky výzkumu a velmi správně také vymezil použitou metodiku výzkumu a dobře vyjádřil komparace sémantických inferenčních mechanismů.

Zvolené metody zpracování zadaných úkolů autor správně popsal na str.93.

Vysoce kladně hodnotím autorovu snahu o identifikaci podstatných profilů uvedených informačních zdrojů citovaných v práci.

### **Přínos a originalita práce**

Vědecký přínos autora, v souladu s uvedeným popisem v kapitole šesté a při řešení zadaných úkolů považuji za odpovídající, zejména z pohledu uvedených a osvědčených přístupů k řešení tohoto aktuálního problému.

Vymezení vývoje v této oblasti vidím jako velmi dobrý výsledek této práce a systémově integrované popisy o technologiích a standardech Sémantického webu vidím jako další vhodný příspěvek k mimořádně důležitému kroku v informovanosti odborníků.

Autor práce velmi dobře využil stávajících teoretických znalostí a praktických zkušeností především z citovaných informačních zdrojů i chápání takto pojatého tématu zadané práce.

Dále přínos předkládané práce vidím pro rozvoj vědního oboru a z toho i možných aplikačních přínosů, které spatřuji především v rámci pojatém a systémově vyjádřeném modelu.

Přínos práce je nutné posuzovat také z pohledu celé systémové koncepce vysokoškolského vzdělávání a to z pohledu uvedeného pojetí Sémantického webu a jeho technologie.

Význam pro praxi spatřuji ve zpřístupnění techniky sémantické anotace webových dokumentů. Pozitivně hodnotím využití této práce také v komerčním použití ať již ve firemních portálech nebo v publikačních aktivitách řešitelů úkolů vědy, výzkumu a vzdělávání.

### **Formální úprava disertační práce**

Formálně je disertační práce rozdělena do odpovídajících oblastí.

V práci se vyskytují některé nepřesnosti jako např. méně pečlivé vyjadřování popisu os grafů, citacích obrázků apod.

Práce je jinak napsána přehledně, systémově dobře orientována a splňuje obsahově nároky na současné doktorské disertační práce v uvedeném vědním oboru.

### **Publikační činnost**

Uvedená publikační činnost autora je zatím skromná.

### **Otázka do rozpravy:**

Jaká řešení předpokládáte v systémové integraci prostředků ICT v aplikaci Sémantického webu (při současném velkém množství samostatných doménových ontologií?)

### **Závěr**

**Předkládanou práci doporučuji k obhajobě před příslušnou komisí a po úspěšném jejím obhájení udělit jmenovanému titul Ph.D. v uvedeném oboru.**

V Brně 15. května 2015



prof. Ing. Jiří Dvořák, DrSc.  
profesor technické kybernetiky,  
vědeckovýzkumný pracovník  
Vysoké učení technické v Brně  
Fakulta podnikatelská

## **Oponentní posudek doktorské disertační práce**

**Název pojednání:** Výzkum metod optimalizace Sémantického webu s využitím ontologií

**Doktorand:** Ing. Michal Šmiraus

**Školitel:** doc. Ing. Zdenka Prokopová, Csc.

Předložená disertační práce se zabývá výzkumem v oblasti zvolna se rozšiřujícího, avšak dosud plně nspecifikovaného mezioborového odvětví Sémantického webu, přičemž v uvedené práci je zřetelný převažující detailnější záběr zejména na technologickou oblast související s využitím ontologií a RDF schémat při vytěžování dat z rozsáhlých znalostních databází. Vlastní posudek se nejprve zabývá obsahem a technickou stránkou práce, s následným komentářem k definovaným cílům práce, postupu jejich řešení a praktickým výsledkům se závěrečným celkovým zhodnocením celé práce a posouzením její disertability.

### **Obsah a struktura práce**

V celkovém rozsahu necelých 110 stran je pojednáváno o problematice užití technik znalostního inženýrství při rozvoji a budování konceptuálních datových schémat s využitím ontologického přístupu, jehož hlavní a fundamentální výhodou oproti klasickým znalostním bázím je vysoká nezávislost na dané aplikaci. Zvolené téma disertační práce je aktuální, stejně tak výzkumná a vývojová činnost v dané oblasti, která i po letech od prvotního představení celé koncepce Sémantického webu neustále nabírá na intenzitě, a to celosvětově. Vlastní strukturu disertační práce tvoří 6 sekcí přehledně členěných od obecných úvodních informačních částí až po obsahově detailněji zpracované úseky, doplněné o názorná schématická vyobrazení popisovaných procesů a přehledové grafy. V průřezu celé práce je obsažen jak stručný teoretický úvod do problematiky, tak i praktické možnosti návrhu procesu pro zpracování dotazu v přirozeném jazyce na konkrétní aplikaci, dále experimentální měření na doménové ontologii a univerzálně použitelný zásuvný modul pro CMS systém Wordpress.

### **Splnění cílů disertační práce**

Samotná práce řeší několik dílčích cílů, které jsou nejprve obecně stanoveny v sekci 2 a později konkrétně, přesto stále stručně a věcně, popisovány v sekci 4 již jako dosažené výsledky, které jsou ve většině případů také podloženy odpovídajícími praktickými výstupy, tudíž lze tyto body považovat za splněné. Ze stanovených stěžejních cílů práce vyplývá také její hlavní motivace a tou je snaha o vytvoření v praxi snadno použitelného programového prostředku pro zajištění sémantické anotace webových dokumentů, která je nezbytným předpokladem pro následné zpracování dat a jejich vzájemnou provázanost v prostředí Sémantického webu.

Návrhu a implementaci komponenty pro všeobecně rozšířený WYSIWYG editor předcházeli teoretický průzkum dotazníkovým šetřením na vzorku necelé tisícovky respondentů, převážně z oblasti IT, který poukázal na skutečnost, že myšlenka Sémantického webu se sice od svého vzniku stala již dostatečně známou, avšak jejímu dalšímu využití brání především absence snadno použitelného a uživatelsky přívětivého nástroje pro jeho tvorbu, ideálně v online prostředí a bez předchozí vývojářské znalosti zdrojového kódu (nikoli však jako desktopová aplikace). Zajímavým cílem je bezesporu i snaha o komplexní zpracování informací dané problematiky v rámci vlastního encyklopedicky koncipovaného portálu v českém jazyce, neboť kvalitně sepsaných a do češtiny přeložených přehledových materiálů je zatím stále nemnoho.

Návrh procesu pro zpracování dotazů v přirozeném jazyce má za cíl užití pouze v rámci konkrétní doménové ontologie. Stálo by však za zvážení, zda by daný koncept byl uplatnitelný i v jiných oblastech a co by takové univerzální řešení vyžadovalo, případně jaké potenciální problémy by v takovém případě mohly nastat. S tím souvisí i závěrečné zkoumání a porovnávání sémantických inferenčních mechanismů. Naměřené výsledky sice mohou být pomocným ukazatelem pro výzkumníky v dané oblasti, kteří se rozhodují o (ne)vhodnosti použití toho kterého odvozovacího modulu, ale je téměř jisté, že vždy bude záviset na daném způsobu použití a konkrétní ontologii, nad kterou se tyto moduly budou uplatňovat, tudíž je nelze brát jako univerzální komparační výstup. Nicméně chápu to tak, že tato závěrečná oblast experimentálního výzkumu je v rámci disertace a jejím celkovém konceptu myšlena spíše jako podružný (nikoli hlavní) cíl a jako takový je možno jej taktéž považovat za splněný.

### **Stanovisko k postupu řešení, k výsledkům a přínosům disertační práce**

Přiložený seznam literatury obsahuje značné množství odborných článků. Je zřejmé, že je autor četl, kriticky zhodnotil a že je schopen se jimi inspirovat ve vlastní práci, což hodnotím velmi kladně. Jádro praktické části, obsahující výsledky této práce se vymezuje na pokrytí oblasti CMS systémů, které v souvislosti s dalším využitím Sémantického webu nebyly dosud tolik hojně diskutovány, nicméně je zde patrný jasný cíl a záměr o přiblížení technologií automatické či manuální sémantické anotace běžnému uživateli, což v konečném důsledku může výrazně napomoci k dalšímu širšímu rozšíření Sémantického webu jako takového.

Celkový přínos disertace vidím jak ve vědecké, popř. akademické sféře, tak i v oblasti praktické, zejména týká-li se možného dalšího rozvoje a užití komponenty zásuvného modulu uživatelského editoru v rámci komunity celosvětově nejužívanějších CMS systémů, což by mohlo i běžné uživatele bez znalosti kódu motivovat k tvorbě sémanticky anotovaných webových dokumentů a časem (podobně jako po stále ještě neutichající "SEO horečce") bychom se mohli dočkat rostoucího počtu takto optimalizovaných stránek, které následně dopomohou k větší provázanosti a relevantnosti dat na objemově neustále se zvětšujícím Webu.

### **Formální a jazyková stránka práce**

Vlastní disertace je svou strukturou a obsahem po formální stránce zpracována na dobré úrovni. V textu se nevyskytují žádné zásadní věcné nesrovnalosti či nejasně vyjádřená tvrzení. Co se týče jazykové stránky práce, vyskytují se zde občas drobné gramatické nesrovnalosti v interpunkci, stejně tak i několik v závěru nevysvětlených zkratk a místy též přehlédnuté či dosud neopravené překlepy.

### **Otázky k obhajobě**

1. Zdůvodněte použití standardu serializace RDFa pro sémantickou anotaci v praktické části práce. V čem vidíte výhody a naopak nevýhody takového řešení?
2. Jaký je Váš názor na řešení otázky integrity a důvěryhodnosti dat v rámci struktury technologických vrstev Sémantického webu?
3. Lze nějakým způsobem výhledově řešit „mnohojazyčnost“ některých webových aplikací a ontologií při zachování jejich vzájemné interoperability?

## **Celkové zhodnocení a disertabilita**

Předkládaná disertační práce je výsledkem systematického experimentálního studia doktoranda v aktuální a neustále se rozvíjející oblasti s prokazatelným akademickým i praktickým přínosem. Ing. Michal Šmiraus prokázal schopnost samostatné vědecké práce, jejíž výsledky průběžně publikoval na české i mezinárodní úrovni, na konferencích i v časopisech. Práce jako celek může být obstojně a čtivě zpracovaným komplexním přehledem v oblasti užitých technologií Sémantického webu.

Závěrem a na základě výše uvedeného hodnocení mohu konstatovat, že předložená disertace věcně i formálně plní zákonná ustanovení z hlediska obecně uznávaných požadavků na disertační práce, a proto ji

**doporučuji**

**přijmout k obhajobě a po úspěšném obhájení udělit Ing. Michalu Šmirausovi titul Ph.D.**

V Žeranovicích 18.5. 2015

Ing. Jaroslav Klimek, Ph.D.



---