

POSUDEK OPONENTA BAKALÁŘSKÉ PRÁCE

Autor práce: Ján Káva
Název práce: Analýza využívanosti modelu Schema.org
Studijní program: AI - AI
Termín odevzdání: 5/2018
Vedoucí práce: Doc. Ing. Vojtěch Svátek, Dr.
Oponent práce: Mgr. Jindřich Mynarz, Ph.D.

Celkové hodnocení práce

- Obtížnost zvoleného tématu a stanoveného cíle vzhledem ke studijnímu programu je adekvátní.
- Zvolené téma a stanovený cíl je vzhledem ke studijnímu programu zcela relevantní.
- Téma je stále aktuální a řešení nebo výsledek je původní částečně.
- Doporučuji předloženou bakalářskou práci k obhajobě.
- Navrhuji klasifikační stupeň dobře.

Hodnocení práce v dílčích kritériích

Dílčí kritéria	Výborně (90 % a více)	Velmi dobře (75 % až 89 %)	Dobře (60 % až 74 %)	Nevyhovující (méně než 60 %)
Jasnost a srozumitelnost formulace tématu a cíle práce			x	
Míra naplnění cíle		x		
Logická stavba práce a vzájemná konzistence jednotlivých částí			x	
Rozsah a relevance popisu současného stavu			x	
Adekvátnost metod k řešenému problému, korektnost výběru a aplikace			x	
Rozsah, hloubka a preciznost popisu výsledku			x	
Relevance a správnost diskuze výsledku			x	
Korektnost zpracování soupisu zdrojů a metody odkazování			x	
Relevance informačních zdrojů				x
Korektnost použití informačních zdrojů (citace a parafráze)				x
Gramatika, jazykový styl, terminologie			x	
Úprava práce	x			

Doplňující komentář hodnotitele

Odevzdaná práce je třetí iterací na dané téma. Jako takové se jí dostalo důkladnější zpětné vazby než pracím, které byly obhájeny v první iteraci. Tato iterace práce některé nedostatky vytýkané předchozím iteracím opravuje, ale jiné zůstávají; např. sčítání vzájemně se překrývajících procentních podílů do 100 %.

Užitečnost práce trpí neodbornou formou, která obsahuje nepodložená tvrzení nebo chybějící či nepřesné citace.

Oceňuji, že autor našel (na s. 31) potenciální chybu ve statistikách Web Data Commons pro extrakci JSON-LD a nahlásil jí na mailing listu projektu. Bohužel je jasné, že se autorovi nestihlo dostat odpovědi, když dotaz zaslal den před termínem odevzdání bakalářské práce. Nicméně v práci je překlep: uvádí 2 689 746 domén se Schema.org JSON-LD, zatímco ve zdrojových datech Web Data Commons (a v emailu autora práce) je to 2 689 736 domén.

Dílicí připomínky

- (s. 4) Pokud sémantický web označen jako web 3.0, pak se jeho myšlenka datuje do doby vzniku WWW v raných 90. letech. Pokud jde o označení "web 3.0", pak souhlasím, že je novějšího data.
- (s. 4) A List Apart není blogem Jeffreyho Zeldmana.
- (s. 9) Citovaná definice sémantického webu na odkazované stránce (<https://www.w3.org/standards/semanticweb>) není. Odkud citace pochází?
- (s. 10) Architektura sémantického webu uvedená na obrázku 1 je zastaralá. XML syntaxe už dávno není preferovanou nebo jedinou standardní syntaxí pro RDF.
- (s. 27) N-Quads neobsahují „označenie RDF tripletu“, ale IRI pojmenovaných grafů.
- (s. 29) „Väčšina SEO best practice blogov a tutoriálov práve odporúča využívanie formátu Microdata spolu s modelom schema.org.“ Chtělo by podložit citací nebo alespoň příkladem takového blogu.
- (s. 31) „Výlučnosť modelu schema.org vyplýva z toho, že JSON-LD sa hlavne využíva pri tvorbe webových aplikácií pre prenos dát z backendu na frontend v rámci RESTového rozhrania, kde sú následne tieto dáta vykreslené do DOMu (Document Object Model).“ Toto tvrzení je třeba podložit buď citací nebo vysvětlením. Následně je podáno lepší zdůvodnění využití modelu Schema.org v JSON-LD: většina dokumentace JSON-LD uvádí příklady se Schema.org.
- (s. 31) „pozrieme na oficiálnu dokumentáciu JSON-LD napr. v Node.JS prostredí“ Chybí odkaz na dokumentaci.
- (s. 36) Práce pokračuje s chybným předpokladem vytýkaným minulým verzím, že přítomnost jednotlivých rozšíření Schema.org na stránkách se vzájemně vylučuje. Stránka může využívat více než jedno rozšíření. Potom nemá smysl procenta výskytů rozšíření na stránkách počítat do 100 %.
- (s. 33) Regulární výraz je nepřesný: lomítka pro oddělení protokolu jsou povinná, subdomény se vzájemně vylučují (tj. např. <http://health-lifesci.auto.schema.org> je dle výrazu validní).
- (s. 40) „V súčasnosti však máme javascript frameworky, ktoré sú rýchle a hlavne nevyužívajú klasickú DOM štruktúru ale virtuálny DOM“ Virtuální DOM je stále transformován na DOM, takže rozlišení je irelevantní. Jde spíše o klasický problém dat generovaných na straně klienta, přičemž crawlers typicky nespouštějí JavaScript, pomocí něhož klient data generuje. Lze to řešit např. nasměrováním crawlerů na HTML generované na straně serveru spuštěním JavaScriptu kódu pomocí Node.js.

Otázky k obhajobě

- (s. 9) Proč je sémantika značkovacích elementů a znalostí na webu odlišná?
- (s. 40) „Určite by bolo zaujímavé zistiť nakoľko sú komaptibilné“ Jak definujete kompatibilitu mezi slovníky?

Datum: 17.5. 2018

(podpis hodnotitele)