

Posudek oponenta diplomové práce

Jméno a příjmení autora posudku: Jakub Klímeck

Jméno a příjmení autora práce: Karel Klíma

Název práce: Customizable Linked Data Browser

Předkládaná práce se zabývá implementací přizpůsobitelného prohlížeče propojených dat (Customizable Linked Data Browser). Práce je dobře a srozumitelně strukturovaná a dobře se čte. Vyvinutý software lze nainstalovat a funguje dle zadání, které je tedy beze zbytku splněno. Práci ale chybí finální korektura, takže práce obsahuje překlepy, např. v sekci 3.3.7 „browsed“ místo „browser“, v sekci 4.1.1 „specify of a“ místo „specify a“ a další. V úvodu práce zabývající se úvodem do problematiky a přehledem existujících řešení se vyskytují faktografické nepřesnosti, na které je třeba upozornit zejména s ohledem na obor předkládané práce:

- 3.1 Za otce internetu je označován Vint Cerf, nikoliv Tim Berners-Lee
- 2.1.1 Tim Berners-Lee není autorem Internetu, nýbrž sítě WWW, což je podstatný rozdíl
- 1.1. Čtvrtý problém Linked Data – Autor tvrdí že entity propojených dat jsou potenciálně nekonečné, což je špatně. Sice není limit na počet relací mezi entitami, tento počet ale bude vždy konečný. I kdyby veškerá úložná kapacita na světě (konečná) byla zaplněna vztahy mezi 2 entitami, pořád to bude konečné číslo. Jedná se samozřejmě o nepřesnost, kterou by se na jiné škole nikdo nezabýval, ale zrovna na MFF je třeba ji vytknout.

Věcné připomínky

- 3.1.2 Jsou identifikovány různé způsoby publikace Linked Data, zejména SPARQL endpoint, poskytování RDF dat při dereferencování URI a poskytování RDF dat v HTML stránce při dereferencování URI. V ospravedlnění podpory pouze přístupu přes SPARQL endpoint ale autor argumentuje jen nepohodlností scrapování obsahu z HTML stránek, s čímž lze souhlasit, nicméně se už nezabývá variantou, kdy při přístupu na URI jsou přímo poskytnuty RDF data, což je jedním z hlavních principů Linked Data, nevyžaduje scrapování a jeho podpora by v Linked Data browseru byla očekávána. S podporou přístupu pouze přes SPARQL endpoint je problém ten, že u cizích URI nevím, kde/zda je SPARQL endpoint přístupný.
- Chybí podpora SPARQL endpointů jiných než Openlink Virtuoso. Argumenty pro toto omezení shledávám nedostatečné, konkrétně:
 - 5.8 – Feature pro práci s prefixy by neměla být kritériem pro volbu podporovaných endpointů. Tuto službu (navíc využitelnou při použití pouze některých RDF serializací jako Turtle nebo JSON-LD) poskytuje nezávislý server prefix.cc. Využití funkcionality v Openlink Virtuoso může být i kontraproduktivní, pokud si její správce nadefinuje jiné než obvyklé významy prefixů.
 - 5.8 – Feature pro fulltext search by také neměla být kritériem pro volbu podporovaných endpointů. Je to hezká funkcionality, která by v případě podpory endpointu (v Openlink Virtuoso) mohla být volitelně zapnutá a uživatel by měl být srozuměn s tím, že v případě chybějící podpory bude práce s Browserem pomalejší, nicméně by to nemělo znamenat, že browser s jinými SPARQL 1.1 kompatibilními endpointy nebude fungovat vůbec.
- 6. – V sekci Evaluation chybí srovnání s jinými řešeními. Jsou zde měření hodnotící vyvíjené řešení, ale chybí měření stejných úloh například v Openlink Faceted Browseru

Otázky k obhajobě

1. Existují ještě nějaké jiné důvody krom těch uvedených v sekci 5.8 pro nepodporu SPARQL endpointů jiných než Openlink Virtuoso a představují nějaký zásadní problém?
2. Jak si výkonostně stojí vaše řešení oproti Openlink Faceted Browseru?

Doporučení k obhajobě:

Z výše uvedených důvodů práci *doporučuji* k obhajobě.

Vynikající práce vhodná pro soutěž studentských prací	ANO <input type="checkbox"/>
---	------------------------------

V Praze dne: 27.8.2015

Podpis: