



národní
úložiště
šedé
literatury

Portál poznámek aneb polidštění semantického webu

Řimnáč, Martin
2010

Dostupný z <http://www.nusl.cz/ntk/nusl-41754>

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 19.04.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní [nusl.cz](http://www.nusl.cz) .

Portál poznámek aneb polidštění sémantického webu

doktorand:

ING. MARTIN ŘIMNÁČ

Ústav informatiky AV ČR, v. v. i.
Pod Vodárenskou věží 2

182 07 Praha 8

rimnacm@cs.cas.cz

školitel:

ING. JÚLIUS ŠTULLER, CSC.

Ústav informatiky AV ČR, v. v. i.
Pod Vodárenskou věží 2

182 07 Praha 8

stuller@cs.cas.cz

obor studia:
Databázové systémy

Práce byla částečně podpořena projektem 1M0554 Ministerstva školství, mládeže a tělovýchovy ČR "Pokročilé sanační technologie a procesy" a záměrem AV0Z10300504 "Computer Science for the Information Society: Models, Algorithms, Applications".

Abstrakt

Příspěvek¹ se s ohledem na aktuální dění v oblasti webových technologií věnuje možnosti implementace portálu, kde by popis objektů byl realizován formou krátkých, sémanticky ohraničených poznámek. Taková forma zajišťuje čitelnost jak pro koncového uživatele (člověka), tak pro softwarové agenty pracující s dokumenty sémantického webu, neboť tyto poznámky mohou být převedeny na RDF data. Další zásadní výhodou použití sémanticky ohraničených poznámek je jejich přímočarý překlad do jiných jazyků, kterého lze využít při tvorbě multijazyčných webových prezentací.

Jedním z aktuálně řešených problémů v oblasti sémantického webu je hledání způsobu, jak v této oblasti zajistit dostatečný objem reálných dat, po kterých by se poptávali také běžní uživatelé webových technologií. Taková data je možné získat buď semiautomatickou extrakcí z klasických webových stránek do formátu sémantického webu anebo manuálně tvorbou ontologií a následnou anotací existujících dat. Zatímco tvorba ontologií je doménou ontologických inženýrů (čemuž odpovídají i podpůrné aplikace), u mnohých jiných formátů (blogy, sociální sítě, sdílení dat) se o tvorbu starají přímo uživatelé bez podrobnějšího obeznámení s použitými technologiemi. Příspěvek se proto snaží poukázat na možnost zadávání RDF dat ve formě sémanticky ohraničených poznámek, kdy daný subjekt (resource) je popsán pomocí (potenciálně anotovaných) hodnot, respektive klíčových slov. K podobné formě dnes přistupují i mnohé sociální sítě; jejich omezení

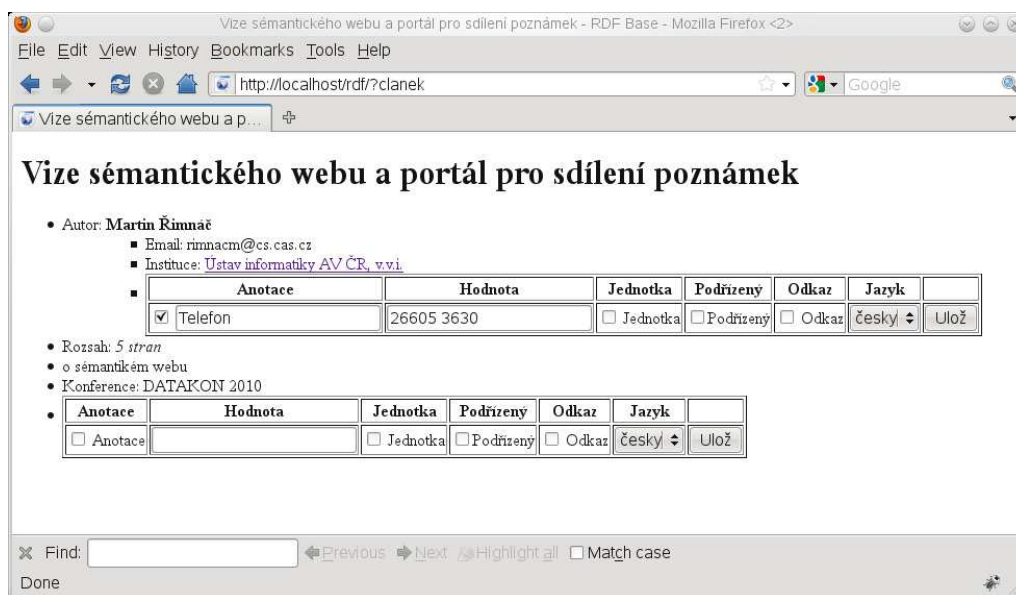
délky sdělení vede uživatele k návyku, aby namísto strojově složitě zpracovatelných vět v přirozeném jazyce použili pouze několik klíčových slov ve stejném významu.

Navrhovaná forma může interaktivním způsobem nabádat uživatele, aby rozšířil popis daného objektu o vlastní postřehy (stejný model tvorby obsahu je znám například z wikipedie). Každý uživatel může buď označit poznámku, kterou navrhuje smazat (se kterou nesouhlasí), může poznámku potvrdit nebo opravit, případně může vložit poznámku novou. Poznámka může být anotována názvem vlastnosti, na rozdíl od klasických ontologií je však většina vlastností definována pouze lokálně. Ukázka takového zadávání RDF dat je na obrázku 1.

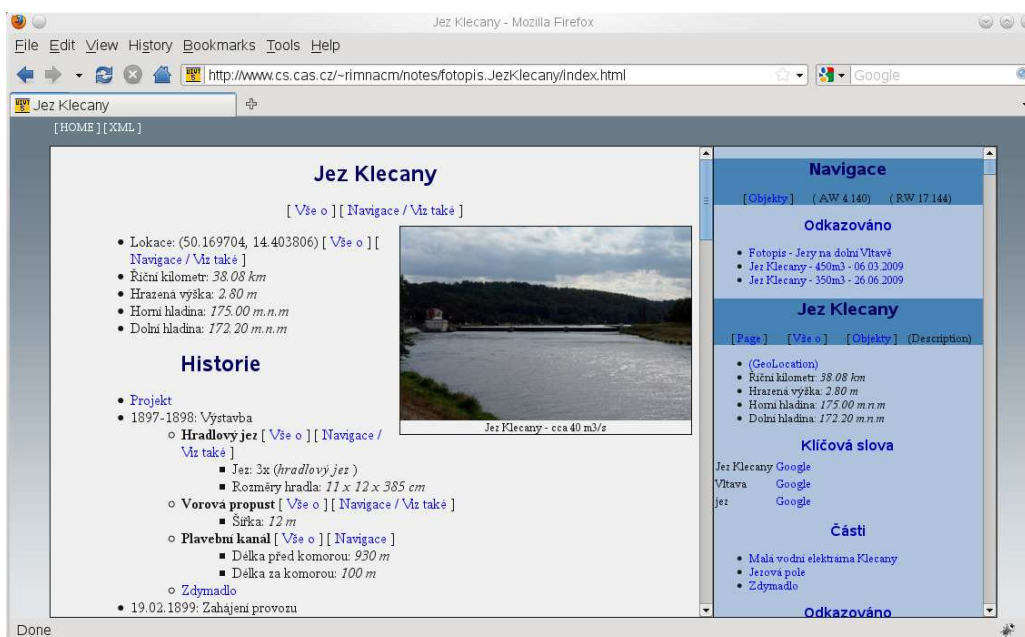
Takto koncipovaná RDF data mohou být transformována XSLT procesorem do XHTML kódu, ideálně s RDFa anotacemi. Taková transformace zajistí indexaci dokumentů publikovaných na portálu dnešními klasickými fulltextovými nástroji - tak ani člověk ani fulltextový nástroj "nepozná", že se vůbec jedná o dokument sémantického webu. Portál tímto způsobem umožňuje spolupráci mezi fulltextovými vyhledávacími a nástroji pro zpracování dokumentů sémantického webu.

S ohledem na transformaci je vhodné zavést i specializované, například multimediální prvky, které obohatí strohý seznam poznámek o grafiku. Tím se portál dále přibližuje současným uživatelským požadavkům na webovou prezentaci. Ukázka takového výstupu včetně pokročilých navigačních nástrojů je ilustrována na obrázku 2.

¹Příspěvek je rozšířeným abstraktem příspěvku přijatému na mezinárodní konferenci DATAKON 2010.



Obrázek 1: Studie interaktivního vkládání strukturovaných poznámek



Obrázek 2: Ukázka prezentující RDF data ve formě poznámek

Vzhledem k rozličné kvalitě vkládaných poznámek je nanejvýše vhodné každou poznámku ohodnotit fuzzy mírou udávající její důvěryhodnost. Ta zajistí, že při vyhledávání budou nalezeny všechny varianty k dané poznámce, avšak každá z nich bude ohodnocena pouze určitým stupněm (varianty mohou být seřazeny podle této míry). Analogicky lze postupovat i u hodnocení uživatelů a vhodně tak iniciovat váhy uživatelem nově vložených poznámek, popřípadě tyto váhy využít stejným způsobem jako v sociálních sítích.

Z pohledu ostatních zdrojů RDF dat může být portál

chápán jako jedna součást distribuovaného prostředí vzájemně spolupracujících zdrojů. Takové prostředí umožňuje značně omezit redundanci i duplicitu prezentovaných dat, kdy namísto kopie dat je použito pouze odkazu na (externí) zdroj. Data v takovém prostředí bývají označována jako tzv. linked data. Portál tuto aktuálně se rozvíjející technologii navíc rozšiřuje o aspekt důvěryhodnosti těchto dat.

Návrh a parametry portálu jsou předmětem několika posledních prací a jsou součástí závěrečné části di-sertační práce autora.