Abstrakt článku z Semantic Web Journal

Map-On: A web-based editor for visual ontology mapping (1)

S rozvojem principů sémantického webu a propojených dat roste poptávka po nástrojích, které by v kterékoliv etapě tvorby vazeb postup zjednodušily, či rovnou zautomatizovaly. Autoři v rámci představení svého nástroje zdůrazňují, že je v současné době stále běžné většinu ontologických vazeb tvořit ručně. Snahou tak je, aby nebylo nutné být odborníkem jak v práci s databází, tak i v tvorbě ontologií.

Nynější nástroje většinou používají znázornění ve stromu pro znalosti a tabulku pro DB záznamy. Limit tabulky tkví v nemožnosti znázornit komplikovanou strukturu dat či vazeb. Zároveň má omezený prostor a bojuje tak s nepřehledností. Zobrazení v grafu je naproti tomu nejvíce přirozená a používaná vizualizace, avšak je náročné jej udržovat přehledný, pokud přesáhne kritický počet vzájemně propojených bodů.

Stále běžná ruční tvorba R2RML (RDB do RDF) mapování vyžaduje od člověka perfektní znalost struktury databáze a zároveň cílové ontologie pro nalezení vazeb. Uživatel musí tvořit SQL dotazy a zároveň IRI vzory pro subjekty a objekty. Právě zde přichází na řadu nástroj Map-On nabízející vizuální reprezentaci mapování a klikatelný interface pro spojení sloupců s koncepty, rolemi a atributy. IRI vzory a SQL dotazy jsou automaticky generovány, což znamená výraznou úsporu času a nároků na uživatele.

Prostředí Map-On se vyznačuje 4 vlastnostmi – interface pro ovládání počítačovou myší, mapování na základě ontologií, tzv. top-down vizualizace a mapovací mezery. Editor je ovládán hlavně kurzorem myši, což je hlavním benefitem pro snadné ponoření se do problematiky bez znalosti mapovacích jazyků jako R2RML. Mapování na základě ontologií je jeden z možných přístupů (druhý je řízený databází), který je uživatelům zběhlým v práci s databází bližší. Top-down vizualizace třída – data – tabulka/sloupec působí velmi přehledně a pomáhá k snadné kontrole databázových i ontologických konstrukcí. Navíc je v levém sloupci aplikace box pro každou třídu jasně ukazující, s čím je spojena. Přirozeně se pracuje i s hover efektem pro zdůraznění vazeb. Logická tabulka mapy trojic je generována automaticky ve formě SQL dotazů, které vycházejí z vazeb a jejich názvů a jsou tak velmi srozumitelné.

Architektura systému se skládá z několika částí - rozhraní pro koncového uživatele, správa zdrojů, správa ontologií, mapovací modul, R2RML procesor, logování a konečně databáze. O rozhraní byla již zmínka výše, jednoduchost ovládání pomáhá uživateli tvořit přehlednější mapy pouhým přetažením prvku či prvků na jiné místo. Samozřejmostí je našeptávač. Správa zdrojů a ontologií slouží k načtení databázových schémat či OWL ontologií v RDF/XML do systému. R2RML procesor pak na základě vstupních dat a mapy uživatele tvoří R2RML dokument.

Postup práce v programu je pak vcelku přirozená. Na začátku je nutné nahrát databázi a ontologie, pak definovat mapovací mezery. Po vytvoření jednotlivých mapování a propojení tříd je možné vygenerovat R2RML dokument a ten si pak posléze nejen otevřít, ale i upravit, k čemuž pomáhá zvýrazněná turtle syntaxe.

Během vývoje programu proběhlo i uživatelské testování na 5 lidech pro demonstrování uživatelské přívětivosti i pro amatéry v práci s ontologiemi, kteří však měli znalost databází a SQL. Uživatelé neměli k dispozici žádné materiály o konceptu sémantického webu a nebyli seznámeni se speciálními jazyky jako OWL, RDF, SPARQL či R2RML. 2 z nich byli profesoři z technických škol, 2 odborníci z průmyslu a jeden z vědecké komunity.

Testování probíhalo jednotlivě průměrně 50 minut na běžném notebooku s připojenou myší. Během toho byla obrazovka a pohyby myší zaznamenávány. Na začátku proběhlo stručné seznámení s principy Semantic Web a výhodami RDF. Uživatelé se seznámili s ovládáním Map-On editoru. Testeři vyplnili jeden dotazník na začátku ověřující úroveň odborných znalostí a druhý na konci pro zpětnou vazbu z používání. Téma, na kterém v programu pracovali, byly výzkumné konference, což má být snadno představitelná a pochopitelná realita.

Cílem testu bylo i na malém počtu účastníků vyzkoušet co nejlépe reálné fungování a očekávání, které program nabízí. 3 úkoly, se kterými se museli testeři vypořádat a které měly vzestupnou náročnost, se dařilo plnit se sestupnou přesností. Již druhá úroveň obtížnosti dávala tvůrcům potřebnou zpětnou vazbu v nedostatečných chybových hláškách, která se uživatelům zobrazovala. U 3. úkolu pak chyběla znalost práce s ontologiemi, na druhou stranu již bylo znát, že se testeři na druhém úkolu naučili, jak na věc.

Z dotazníků na konci testu vyplynulo, že je Map-On vnímán jako rychle pochopitelný nástroj, který nevyžaduje velké znalosti na začátku. Pocity z používání však byly pozitivní, i negativní. Uživatelům chyběl průvodce, který by je na začátku navedl na to, jak tvořit vlastní mapování. Z analýzy chování pak vyplynuly další body, co by bylo dobré zlepšit, ačkoliv s nimi nebyl problém u všech.

Na závěr pak autoři zkusili ukázat reálné uplatnění editoru použitím energetických dat třech velkých měst v Evropě. Map-On byl použit pro tvorbu globální ontologie společné pro všechny tři databáze. Dále by měl být využíván k aktualizování mapování při změnách zdrojů. Pokus byl úspěšný včetně toho hlavního – žádný SQL dotaz nemusel být psán ručně, vše bylo generováno automaticky.

1. **Sicilia, Álvaro, Nemirovski, German a Nolle, Andreas.** Map-On: A web-based editor for visual. *Semantic Web Journal.*