Abstrakt článku

“Seed, an End-user Text Composition Tool for the Semantic Web”

Martin Svatoš

Práce “Seed, an End-user Text Composition Tool for the Semantic Web” je rozdělena na pět částí: úvod, architerturu a implementaci programu Seed, zkušební experiment, vyhodnocení experimentu a závěr. Článek se zabývá aplikací Seed, což je zkratka ze spojení semantic editor. Jedná se o aplikaci pro koncové uživatele, která má zprostředkovat možnost přístupu k sémantickému webu i těm, kteří s ním jinak neumí pracovat. Program umožňuje jednoduché procházení a zároveň vytváření nového obsahu sémantického webu.

V rámci úvodu se článek nejprve zabývá vývojem sémantického webu z dokumentů s propojenými vazbami až na propojená data jako taková, s obrovskými úložišti jako DB-Pedia a Freebase. Zmiňuje se také o menších soukromých úložištích, například pro firemní účely. Poukazuje však na to, že přes veškerý pokrok a rozvoj v této oblasti řídí tvorbu těchto dat skoro zcela pouze vysoce trénovaní jednotlivci. Tuto skutečnost popisují jako sémantickou propast.

V druhé části úvodu je představeno několik prací, které se zabývají podobnou problematikou, jmenovitě OntoWiki, SAHA, One Click Annotator a RDFaCE. Poukazuje se na to, že tyto nástroje se snaží vytvořit dokumenty dvojího typu, a to: a.) plně sémantické b.) založené na nesémantické reprezentaci dat obohacené sémantickou reprezentací v průběhu psaní. Z těchto SCA (Semantic Content Authoring) nástrojů se poté zdůrazňují nástroje One Click Annotator a RDFaCE, které mají podobný cíl jako Seed. Dále se článek zaměřuje na přínos této aplikace a její popis. Jedná se o rozsáhlý, znalostmi podpořený, webově založený nástroj pro skladbu textu v přirozeném jazyce, jehož cílem je především zaplnit propast mezi nezkušenými uživateli a sémanticky anotovaným textovým obsahem na webu. Aplikace umožňuje automatickou či polo-automatickou tvorbu mikrodaty anotovaného HTML textového obsahu bez jakýchkoliv požadavků na vědomosti o technologii propojených dat, či jejich anotačním formátu. Toho je docíleno tím, že se aplikace zaměřuje především na skladbu textu, než na analýzu anotací. V rámci aplikace jsou uživateli v textu samy nabízeny možné anotace na základě toho, co píše, než aby je musel sám hledat. Díky tomu je umožněna interakce s daty a možnost jejich objevování a procházení (pokud například uživatel do textového editoru napíše slovo “Prague”, jsou mu nabídnuty různé varianty a možnosti, co si pod tímto slovem v textu představit. Tyto možnosti může postupně proházet, včetně všech o nich dostupných informací, a nakonec k textu některou z nich přiřadit).

Druhá část článku se zaměřuje na architekturu a implementaci aplikace. Hlavními třemi komponentami aplikace jsou: a.) znalostní vrstva b.) back-end c.) web front-end. Jejich komunikaci zajišťuje standardní webové API. Znalostní vrstva vychází především z informací dostupných na DBPedii a Freebase, ale také z různých ontologických slovníků. Back-end se skládá ze dvou podčástí - NLP komponenta, která se stará o rozpoznávání různých entit zastoupena na serverové straně Stanford Core NLP a Apache OpenNLP a LOD komponenty, která vyhledává informace o rozpoznaných entitách. Web frond-end je poté vytvořen kombinací JavaScriptu a html.

Třetí část porovnává Seed s již zmíněnými programy One Click Annotator a RDFaCE. Tyto programy jsou vybrány především proto, že se stejně jako Seed alespoň

částečně zaměřují na koncové uživatele. Porovnány jsou nejprve tabulkou a poté i volným textem, který je jakýmsi shrnutím tabulky. Jako první je rozebráno hledisko potřebných znalostí. Oba porovnávané programy na rozdíl od Seedu požadují od autora alespoň základní znalosti trojic a znalosti alespoň některých specifických slovníků. u Seedu se o tyto aspekty stará aplikace, nikoli uživatel. Druhé hledisko se týká anotací v reálném čase, RDFaCE požaduje od autora, aby anotace dělal ručně, Seed na druhou stranu sám hledá možná klíčová slova, která by mohla být anotacemi a uživateli je zvýrazňuje. To má za cíl povzbudit autora k tomu, aby anotace sám více specifikoval a dále vytvářel. Z pohledu nativních anotací je poté u RDFaCE velice obtížné anotaci změnit, jakmile je jednou v textu vytvořena. Seed tyto změny podporuje libovolně. Zaměření na znalosti - Seed umožňuje uživateli různé náhledy na entity označené v textu, jejich volné procházení a studování. Z pohledu posudku aplikace poté upozorňují na to, že Seed byl v rámci vývoje vyzkoušen 120 nezávislými účastníky z různých vrstev společnosti, na rozdíl například od RDFaCE, který zkoušelo pouze 16 účastníků se zkušenostmi se sémantickým webem.

Kapitola zkušebního experimentu se zaměřuje právě na představení nástroje 120 účastníkům, kteří mají různé úrovně znalostí práce s propojenými daty. Cílem kapitoly je a.) ukázat, že Seed je snadno použitelný nástroj pro práci se sémantickým obsahem b.) i uživatel bez znalostí sémantického webu je díky němu schopen vytvořit jednoduchý anotovaný text c.) pomáhá uživateli ovlivňovat a prozkoumávat sémantický web během psaní textu. Demo použité účastníky je dostupné na stránce http://tiny.cc/seed-demo.

Z experimentu vyplynulo, že SUS (System Usability Score) Seedu se pohybuje okolo 73,5. SUS bylo zjištěno pomocí dotazníku, který účastníci vyplnili po dokončení dema. Dále z dotazníku vyplývá, že 55% účastníků aktivně anotovali během psaní textu, 39% anotovalo po dokončení textu a zbytek neměl vůči anotování textu preferenci. V rámci objevování nových znalostí 95% účastníků uvedlo, že ve vedlejším panelu Seedu našli informace, které hledali. V závěru je poté shrnut obsah článku a autoři se zaobírají dalšími možnostmi aplikace Seed, jako je například její přeložení do dalších jazyků mimo němčiny a angličtiny. Také se zabývají otázkou budoucího vývoje této aplikace a potenciálu, který může navždy zaplnit propast mezi neosvíceným koncovým uživatelem a sémantickým webem.

ELDESOUKY, Bahhaa, BARKY, Menna, MAUS, Heiko a DENGEL, Andreas, *Seed, an End-User Text Composition Tool for the Semantic Web,* V: *15th International Semantic Web Conference, Kobe, Japan, October 17–21, 2016, Proceedings, Part I,* Springer International Publishing, ISBN: 978-3-319-46523-4, str.: 218-233, 2016