

KONFERENCA SWIB14

SEMANTIČNI SPLET V KNJIŽNICAH

Med 1. in 3. decembrom 2014 se je v Bonnu odvijala šesta konferenca SWIB (Semantic Web in Libraries), ki se je udeležilo 165 udeležencev iz 24 držav. Konferenca je bila razdeljena v dva dela: prvi dan je bil namenjen delavnicam, drugi in tretji dan pa se je zvrstilo 17 predavanj. Namen konference je bil predstaviti novosti ter spodbuditi izmenjavo idej in izkušenj med udeleženci v zvezi z objavo in povezovanjem knjižničnih metapodatkov na spletu ter razvojem storitev in orodij, ki jih knjižnice in sorodne institucije že razvijajo na osnovi paradigme LOD (Linked Open Data). Povzetki, prosojnice in posnetki delavnic ter predavanj so na voljo na spletni strani <http://swib.org/swib14/>. Naslednja konferenca SWIB15 bo v Hamburgu med 23. in 25. novembrom 2015.

DELAVNICE

Introduction to Linked Open Data

Delavnico sta vodila Felix Ostrowski in Adrian Pohl iz nemškega podjetja *graphthinking*. Namen te uvodne delavnice je bil predstaviti bistvo pristopa povezanih podatkov (angl. *linked data*) in osnovna zakonita določila odprtih podatkov (angl. *open data*). Udeleženci smo z ustvarjanjem in izpopolnjevanjem dokumentov RDF poglobili svoje znanje o tej temi. Predstavljeni in hkrati preizkušeni so bili principi povezanih podatkov, kot je npr. objava opisov RDF v spletnem okolju. Pokazalo se je, da sta prednosti objave povezanih podatkov zbiranje podatkov iz različnih virov in ustvarjanje poizvedb, shema RDF pa je bila predstavljena kot učinkovit način podatkovne integracije.

Introducing RDFa with schema.org for web applications

Zelo praktično naravnano delavnico je vodil Dan Scott (Laurentian University, Kanada). Po kratki uvodni predstavitvi smo samostojno reševali naloge, razdeljene v štiri sklope:

1. RDFa

Leta 2008 je W3C (World Wide Web Consortium) objavil prvo specifikacijo RDFa (Resource Descrip-

tion Framework in Attributes) za razširitev sintakse XHTML s podporo specifikaciji RDF (Resource Description Framework). RDFa je takoj postal ključna metoda za povezovanje podatkov na spletu; leta 2013 je bil razširjen s specifikacijo RDFa 1.1 (www.w3.org/TR/rdfa-lite/) in novo specifikacijo RDFa Lite (www.w3.org/TR/rdfa-lite/), ki zajema minimalno podmnožico atributov za predstavitev podatkov RDF v spletnih dokumentih HTML, SVG in XML. To je oživilo zanimanje za združevanje semantičnega in dokumentnega spleta. Na delavnici smo na vzorčnem dokumentu HTML z dodajanjem strukturiranih podatkov na praktičnih primerih preizkusili bistvene koncepte specifikacije RDFa Lite.

2. shema.org (<https://schema.org/>)

V drugem delu delavnice smo v naš vzorčni dokument dodajali še elemente iz strukture shema.org za podatke o knjižnicah, kot so npr. naslov, lokacija, delovni čas. Schema.org je besednjak, ki so ga leta 2011 pripravili Google, Yahoo in Yandex, da bi s tem omogočili preprost način objave metapodatkov o priljubljenih iskalnih temah, kot so dogodki, ljudje, proizvodi itd. Razvoj besednjaka je leta 2013 prevzel W3C. Prednosti uporabe besednjaka shema.org za knjižnične aplikacije je več: knjižnica lahko ponudi natančnejše in bogatejše rezultate iskanja po svojih metapodatkih, izboljša lahko algoritme razvrstitve rezultatov iskanja po pomembnosti (angl. *search ranking*), omogoči pa lahko tudi podrobnejše iskanje in predstavitev svojih organizacijskih informacij.

3. Spletni iskalnik

Google nam, poleg indeksiranja vsebine, ki je dostopna preko njegovega iskalnika google.com, omogoča kreiranje svojega lastnega, osebno prilagojenega spletnega iskalnika CSE (Custom Search Engine, <https://cse.google.com/cse/>), ki ponuja hitro nameščeno in hkrati zelo prilagodljivo iskalno storitev. V tem delu delavnice smo ustvarili enostaven spletni iskalnik za izbrane knjižnične kataloge s strukturiranimi podatki po besednjaku shema.org, na osnovi katerih smo, z zajemanjem in filtriranjem bibliografskih podatkov, zgradili svoj vzajemni katalog.

4. Zajemanje podatkov

V tem delu delavnice je bil predstavljen enostaven

odprtokodni spletni pajek RDFLib, napisan v pythonu, za zajemanje strukturiranih podatkov (na primeru podatkov RDFa po besednjaku shema.org) z uporabo sitemaps.org (<http://www.sitemaps.org/>) za specifikacijo seznama naslovov URL-strani knjižničnih katalogov, od koder smo črpali podatke (VuFind, Koha, Evergreen, Blacklight, Islandora). Na ta način smo indeksirali zelene bibliografske podatke in omogočili lažje iskanje in prikazovanje podatkov iz shranjenih indeksnih datotek.

PREDAVANJA

Using linked data to annotate semantically the BBC's content

Tom Grahame (BBC, Združeno kraljestvo) je predstavil uporabo povezanih podatkov za semantično označevanje BBC-jevih vsebin. Povezane podatke so vpeljali kot skupke idej, metod in tehnologij pri gradnji njihovih spletnih strani, kar je poenostavilo potek dela ter ponudilo nove zanimive avtomatizirane agregacije podatkov za njihove uporabnike. Uspeh uporabe te tehnologije pri pokrivanju online vsebin pomembnih športnih dogodkov je pokazal, kakšne so možnosti za ponovno uporabo semantične infrastrukture kot osrednjega dela BBC-jevega produkcijskega procesa. Vizijo semantičnega objavljanja so razvili s povezovanjem uporabnikom zanimivih vsebin z različnih področij, npr. politika, šport, glasba, ter vsebin v zvezi z različnimi dogodki, mesti, temami itd. BBC vsak dan ustvari obilo vsebin (od člankov, programov, izobraževalnih priročnikov do odlomkov in receptov), ki se med seboj povezujejo in na različne načine prikazujejo v ločenih proizvodih, namenjenih različnim tipom uporabnikov – vsebina pa je vedno enaka. Tako je novinarjem in vsebinskim urednikom omogočen enoten dostop do heterogenih in doslej izoliranih kreativnih del. Predavatelj je opisal, kako je BBC Sport razširil in razvil uporabo povezanih podatkov od izdelave spletne strani, ki je pokrivala svetovno prvenstvo v nogometu leta 2010, do dinamičnega zbiranja dnevnih športnih prispevkov in pokrivanja velikih dogodkov, kot so bili London 2012, Soči 2014 in Svetovno prvenstvo v nogometu leta 2014. Predstavil je še, kako isti tehnološki pristop uvajajo znotraj BBC-ja na različnih področjih, kot so izobraževanje, novice, radio in glasba, in kako se lahko pristop povezanih podatkov uporabi pri podobnih izzivih v bibliotekarskem okolju.

schema.org: machine-readable cataloguing for the open web

Dan Scott (Laurentian University, Kanada) je predstavil prizadevanja za uporabo besednjaka schema.org v bibliografskem okolju. Medtem ko so ga spletne strani navdušeno prevzele v upanju po boljših iskalnih rezultatih in

uvrstitvah, so knjižnice pri integriranju tega besednjaka nekoliko bolj zadržane. Namen uporabe besednjaka schema.org je zagotoviti točke dostopa za knjižnične vire v velikih spletnih iskalnikih preko bibliografskih metapodatkov, podatkov o zalogi in podatkov o knjižnicah. Poudaril je prednosti integriranih knjižničnih sistemov Evergreen in Koha, odprtokodnih vmesnikov Blacklight in VuFind ter repozitorijev, kot sta Islandora in ScholarSphere, pridobljenih z objavo metapodatkov schema.org (od nestrukturiranih do strukturiranih in povezanih podatkov). Omenjena je bila vloga skupine *W3C Schema Bib Extend Community Group* (<https://www.w3.org/community/schemabibex/>) pri zapolnjevanju vrzeli v besednjaku schema.org in dokumentiranju najboljših praks za uporabo tega besednjaka v knjižnicah. Nazadnje je bilo predstavljeno, kako se lahko običajni delovni procesi v knjižnicah (npr. kreiranje vzajemnega kataloga in preverjanje dostopnosti gradiva), ki so trenutno implementirani s specifičnimi protokoli za posamezne knjižnice/institucije, poenostavijo s standardnimi spletnimi orodji za zajemanje strojno berljivih podatkov, strukturiranih po besednjaku schema.org.

Moving from MARC: How BIBFRAME moves the Linked Data in Libraries conversation to large-scale action

Eric Miller (Zepheira, ZDA) je predstavil tri vidike migracije s formata MARC na model povezanih knjižničnih podatkov z uporabo modela in besednjaka BIBFRAME (<http://www.loc.gov/bibframe/>). To so kontekstni vidik, vidik s stališča orodij za transformacijo ter vidik kreiranja vsebine in prepoznavnosti na spletu. Knjižnice veže dolga zgodovina z bibliografskim formatom MARC kot sofisticiranim in zelo prilagojenim opisnim besednjakom z milijardami zapisov; le-ti so zelo razširjeni preko različnih knjižničnih sistemov in ponudnikov. Z namenom ohraniti vrednost te zapuščine in jo predstaviti na nov, sodoben način je Kongresna knjižnica na principih povezanih podatkov in standardov semantičnega spleta razvila BIBFRAME; pri tem je bil njen cilj, da omogoči razširljivost besednjakov glede na različne profile knjižnic, fleksibilna orodja za transformacijo metapodatkov, iterativni proces povezovanja podatkov iz različnih virov in izboljšavo spletne prepoznavnosti knjižnic in njihovega notranjega delovanja. Praktični nasveti, ki jih je navedel v zvezi z zagotovitvijo pomoči knjižnicam pri prehodu na BIBFRAME, so:

- vrednotenje trenutnih podatkov knjižnic,
- določanje prioritet lokalnih podatkov,
- izvedba obsežne konverzije podatkov,
- specifikacija besednjakov za izvirno vsebino,
- prepoznavanje možnosti povezovanja podatkov,

- prehod z enostavnega prikaza obstoječih podatkov na prikaz, ki izkorišča vse prednosti principa povezanih podatkov z uporabo besednjakov BIBFRAME in schema.org.

Entification: The Route to 'Useful' Library Data

Richard Wallis (OCLC, Združeno kraljestvo) je predstavil nekatere prednosti kreiranja knjižničnih podatkov kot opisov entitet resničnega sveta v modelu WorldCat Entities ter težave, ki se pri tem pojavljajo. Pri povezovanju podatkov gre za prepoznavanje "stvari", opisovanje le-teh in njihovih zvez v mreži drugih "stvari" ali entitet. Britanska knjižnica (z verzijo britanske nacionalne bibliografije, ki temelji na povezanih podatkih) je bila pionir pri oblikovanju entitet resničnega sveta kot osnove za njihov podatkovni model. Podobna raziskava znotraj OCLC-ja, ki je vodila k izdaji entitet kot povezanih odprtih podatkov WorldCat.org, je pokazala prednosti tovrstnega pristopa. Raziskava dokazuje tudi, da je za splet entitet resničnega sveta treba storiti mnogo več kot samo konvertirati zapis za zapisom. Pri tem igrajo pomembno vlogo procesi podatkovnega rudarjenja, odprte normativne podatkovne baze (kot so VIAF, FAST, LCSH) ter uporaba prilagodljivih in širše sprejetih besednjakov.

Nevenka Zupančič, Gordana Budimir