Kreiranje skupova povezanih podataka

Šoštarić, Marko

Master's thesis / Diplomski rad

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Josip Juraj Strossmayer University of Osijek, Faculty of Humanities and Social Sciences / Sveučilište Josipa Jurja Strossmayera u Osijeku, Filozofski fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:142:743247

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-04-16**

FILOZOFSKI FAKULTET
SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU

Repository / Repozitorij:

FFOS-repository - Repository of the Faculty of Humanities and Social Sciences Osijek

zir.nsk.hr

SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU

dabar
DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

Sveučilište J.J. Strossmayera u Osijeku

Filozofski fakultet

Dvopredmetni diplomski studij informacijske tehnologije i nakladništva

Marko Šoštarić

Kreiranje skupova povezanih podataka

Diplomski rad

Mentor: izv. prof. dr. sc. Boris Bosančić

Osijek, 2022.

Sveučilište J.J. Strossmayera u Osijeku

Filozofski fakultet Osijek

Odsjek za informacijske znanosti

Dvopredmetni diplomski studij informacijske tehnologije i nakladništva

---

Marko Šoštarić

Kreiranje skupova povezanih podataka

Diplomski rad

Društvene znanosti, Informacijske i komunikacijske znanosti, Informacijski sustavi i informatologija

Mentor: izv. prof. dr. sc. Boris Bosančić

Osijek, 2022.

Prilog: Izjava o akademskoj čestitosti i o suglasnosti za javno objavljivanje

Obveza je studenta da donju Izjavu vlastoručno potpiše i umetne kao treću stranicu završnoga, odnosno diplomskog rada.

IZJAVA

Izjavljujem s punom materijalnom i moralnom odgovornošću da sam ovaj rad samostalno napisao/napisala te da u njemu nema kopiranih ili prepisanih dijelova teksta tuđih radova, a da nisu označeni kao citati s navođenjem izvora odakle su preneseni.

Svojim vlastoručnim potpisom potvrđujem da sam suglasan/suglasna da Filozofski fakultet u Osijeku trajno pohrani i javno objavi ovaj moj rad u internetskoj bazi završnih i diplomskih radova knjižnice Filozofskog fakulteta u Osijeku, knjižnice Sveučilišta Josipa Jurja Strossmayera u Osijeku i Nacionalne i sveučilišne knjižnice u Zagrebu.

U Osijeku, 27. 09. 2022

Marko Šoštarić, 0122223744

Ime i prezime studenta, JMBAG

Sažetak

Svrha ovog rada ogleda se u prikazu načina korištenja tehnologije povezanih podataka u domeni poslovanja knjižnica i srodnih ustanova. Rad započinje upoznavanjem s početcima weba, semantičkog weba te povezanih podataka. Kroz upoznavanje sa semantičkim webom prikazane su tehnologije koje će se koristiti u praktičnom dijelu rada: RDF, URI/IRI, SPARQL, JSON. Poseban naglasak stavljen je na prikaz koncepta povezanih podataka koji predstavlja ključan pojam u ovom radu. Opisane su i tehnologije koje su obuhvaćene konceptom povezanih podataka te se objašnjava kako svaka pojedina tehnologija funkcionira. U praktičnom dijelu rada pokazano je kako se prikupljeni podatci u tabličnom obliku mogu konvertirati u skup povezanih podataka korištenjem programa *OpenRefine.* U radu su prikazani koraci postupka kreiranja povezanog skupa podataka čiji je završni produkt RDF datoteka. Dobivena RDF datoteka strojno je čitljiva čime je psotignut jedan od ciljeva ovoga rada. Drugi cilj je ostvaren prikazom korištenja dobivene datoteke u programu *GraphDB* za potrebe naprednije analize generiranih podataka u domeni knjižničnog poslovanja. Koristeći se ovdje prezentiranom metodologijom temeljnom na tehnologiji povezanih podataka, knjižnice i srodne ustanove mogu je primijeniti za analizu bilo kojeg drugog aspekta svog poslovanja te na taj način, i u skladu s pravilnom interpretacijom dobivenih podataka, značajno podići razinu učinkovitosti vlastitih usluga.

Ključne riječi: semantički web, povezani podaci, RDF, SPARQL, *OpenRefine*, *GraphDB*

S A D R Ž A J

Sažetak 2

1. Uvod 4

2. Semantički web i povezani podaci 5

2.1 Web 5

2.1. Uvod u semantički web 6

2.2 Tehnologije semantičkog weba 9

2.3. Slojevi semantičkog weba 10

3. Povezani podaci 12

3.1. Načela povezanih podataka 12

3.2. Tehnologije povezanih podataka 14

3.2.1. RDF 14

3.2.2. URI/IRI 17

3.2.3. SPARQL 19

3.2.4. SPARQL sintaksa 20

3.2.5. JSON 22

4. Kreiranje skupa povezanih podataka na primjeru skupa podataka iz domene knjižničnog poslovanja 24

4.1. Općenito o metodologiji 24

4.2. Opis načina primjene metode modeliranja i metode prototipa u kreiranju skupa povezanih podataka 24

4.3. Kreiranje skupa povezanih podataka u programu *OpenRefine* 27

4.4. Primjena kreiranog skupa povezanih podataka 35

5. Zaključak 37

6. Literatura 39

Prilog 41

1. Uvod

Od svog nastanka ranih 2000-ih pa sve do danas, semantički web (engl. *semantic web*) pruža mogućnost razvoja weba podataka (engl. *web of data*). Tim Berners-Lee čovjek je koji nije samo izumio web, nego je i stvorio viziju weba u kojem su podaci strojno čitljivi. Njegova zamisao podrazumijevala je ostvarivanje suradnje između ljudi i računala u svijetu podataka. Jedna od najvažnijih primjena semantičkog weba su povezani podaci (engl. *linked data*). Povezani podaci temeljni su blokovi semantičkog weba koji, uz kombinaciju drugih tehnologija, u okviru uobičajenog weba omogućuju da se podaci učinkovitije procesuiraju i lakše pretražuju.

Diplomski rad daje uvid u web, semantički web te povezane podatke, opisuje povijest razvitka navedenih koncepata, kao i tehnologije koje omogućuju da se ostvari njihova funkcionalnost. U teorijskim poglavljima ključni se pojmovi objašnjavaju na primjerima kako bi se lakše razumjelo na koji način navedene tehnologije funkcioniraju.

Svrha ovoga rada ogleda se u prikazu načina korištenja tehnologije povezanih podataka u domeni poslovanja knjižnica i srodnih ustanova. Rabeći metode prototipa i modeliranja u praktičnom dijelu rada, prikazan je postupak kreiranja oglednog primjera (prototipa) skupa povezanih podataka u okviru domene knjižničnog poslovanja. Kroz detaljno opisane korake stječe se uvid u procese prikupljanja, formatiranja te konverzije podataka u strojnočitljiv skup povezanih podataka. Na kraju rada, demonstrira se i primjena tako stvorenog skupa povezanih podataka u daljnjoj analizi podataka u domeni knjižničnog poslovanja sa svrhom podizanja njegove učinkovitosti.

Postupkom stvaranja ovakvih skupova povezanih podataka pridonosi se ostvarenju ideje semantičkog weba u domeni knjižničnog poslovanja, i koji se zatim može primijeniti i u poslovanju drugih baštinskih ustanova.

2. Semantički web i povezani podaci

Prije razmatranja koncepta povezanih podataka, potrebno je dati osvrt na osnove i početke semantičkog weba kako bi se bolje razumjelo o čemu se radi kada se govori o skupovima povezanih podataka, u koju svrhu i kako se koriste; osim toga, i zašto se u današnje vrijeme spominju u raspravama o podacima i njihovu pretraživanju, povezivanju i razmjeni. Korijeni tehnologija povezanih podataka, dakako, potječu od tehnologija samoga weba.

2.1 Web

World Wide Web (*WWW*) je od trenutka svoga nastanka pa sve do danas imao strmu krivulju napretka; svake godine pojavljuju se nove inovacije i nove svrhe za koje se web može koristiti. U mnogim područjima olakšao je poslovanje i pridonio njihovu unaprjeđivanju kroz primjenu novih ideja.

World Wide Web, ili kraće web, zbirka je dokumenata, odnosno mrežnih stranica, koje su dostupne na internetu. Web se često pogrešno poistovjećuje s internetom, zbog toga je bitno uočiti razliku između tih dvaju pojmova. Dok se web odnosi na zbirku dokumenata koji su međusobno povezani putem hiperveza (engl. *hyperlinks*), internet je velika globalna mreža koja povezuje milijarde računala te ostale elektroničke uređaje omogućujući im korištenje spomenute zbirke dokumenata weba.[1]

Još davne 1989. godine Tim Berners-Lee je dobio ideju kako zasnovati tehnologiju kojom bi se znanstvenicima olakšalo dijeljenje rezultata eksperimenata i time omogućila lakša komunikacija, no nije bio ni svjestan u što će se njegova zamisao pretvoriti i koju će ulogu poprimiti u današnjem globaliziranom svijetu. T. Berners-Lee predložio je da se web temelji na tri ključne tehnologije: HTML-u (*HyperText Markup Language*), URI/IRI-ju (*Uniform Resource Identifier* / *International Resource Identifier*) te hipervezama odnosno poveznicama implementiranim u okviru HTTP-a (*HypeTtext Transfer Protocol*). Ove tehnologije čine i arhitekturu weba.

HTML standardni je označiteljski jezik za kreiranje mrežnih stranica, a sastoji se od određenog broja elemenata koji daju upute pregledniku na koji način prikazati željeni sadržaj

[1]Britannica. „World Wide Web (WWW): History, Definition & Facts“ . URL: https://www.britannica.com/topic/World-Wide-Web (2022-9-13)

nekog mrežnog mjesta.[2] HTML jezikom oblikuje se sadržaj te se stvaraju hiperveze *hipertext* dokumenata. Jezik se lako uči i jednostavan je za korištenje, pa je zbog toga postao i popularan i opće prihvaćen.

Kreiranjem sadržaja putem HTML-a i koristeći internetski rastući sustav jedinstvenih imena za svako umreženo računalo, svaki mrežni resurs dobiva jedinstvenu „lokaciju“, čijom interpretacijom upravlja jednostavan protokol. Spomenute jedinstvene „lokacije“ učinile su mrežne resurse ili izvore široko dostupnima što je omogućilo svakom osobnom računalu priključenom na internet mogućnost pregleda i preuzimanja bilo kojeg sadržaja sa weba. S druge strane, svatko tko bi zakupio prostor na serveru mogao bi na njega postaviti svoje mrežne stranice odnosno resurse. URI/IRI ne obuhvaća samo Jedinstveni lokator resursa (*Unique Resource Locator - URL*) nego i Jedinstveno ime resursa (*Unique Resource Names - URN*).[3] Kombinacijom navedenih tehnologija mrežni resursi postali su globalno dostupni.

Hiperveze su jedna od značajki weba koje omogućuju povezivanje bilo kojeg mrežnog resursa s bilo kojim drugim resursom bez obzira na njihovu lokaciju na webu.[4]

Organizacija koja se brine za razvoj weba je *World Wide Web Consortium* (*W3C*), a osnovao ju je 1994. godine Tim Berners-Lee uz podršku DARPA-e (*Defense Advanced Research Projects Agency*) i CERN-a (*Conseil Européen pour la Recherche Nucléaire*). W3C je odgovoran kako za kreiranje i objavljivanje web standarda, tako i za rast i razvoja weba u cjelini.

2.1. Uvod u semantički web

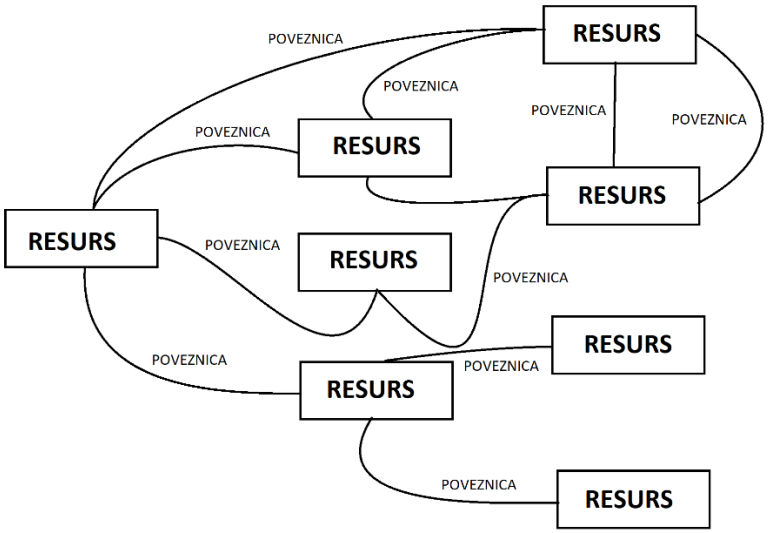
Kako su se na webu informacije pohranjivale godinama, zbog njihove ogromne količine, postajalo je sve teže pristupati im. Drugim riječima, postalo je zahtjevno pretraživati web s ogromnom količinom informacija pohranjenih u bazama podataka, bez sredstva za njihovo učinkovito pretraživanje. Stoga se počelo razmišljati o tome kako ostvariti što brži i što kraći proces pronalaska željenih informacija, i jedna od ideja je bila da se informacije učine strojnočitljivima. Tako se pojavila ideja semantičkog weba, koja je 2001. prvi puta objavljena u časopisu *Scientific American*, koja je trebala znatno unaprijediti postojeći web i učiniti

[2]W3Schools.com. „HTML Introduction“. URL: https://www.w3schools.com/html/html_intro.asp (2022-9-13)
[3]Legg, Catherine. Ontologije na semantičkom webu. // Vjesnik bibliotekara Hrvatske 53, 1(2010), str. 156-157. URL: http://www.hkdrustvo.hr/datoteke/838/vbh/God.53(2022-9-5)
[4] Isto str. 157. (2022-9-5)

informacije na webu strojnočitljivima.[5] Ujedno, semantički web trebao je omogućiti suradnju između ljudi i računala te tako otvoriti svakojake mogućnosti i pogodnosti u obradi, pohrani i pretraživanju informacija. Semantički web je naziv za novu inačicu weba koja predstavlja ekstenziju postojećem webu. T. Berners – Lee definirao je semantički web sljedećim riječima: "Ideja koja stoji iza semantičkog weba je kreirati web koji ne samo da povezuje dokumente jedne s drugima, već također prepoznaje značenje informacija u tim dokumentima".[6] Drugim riječima, cilj semantičkog weba bio je pretvoriti postojeći web koji se zasnivao na nizu međusobno povezanih, ali semantički izoliranih podataka u jednu ogromnu bazu podataka koja će se moći učinkovitije pretraživati i koju će moći koristiti i strojevi. T. Berners – Lee nije smatrao da bi svaki korisnik trebao naučiti osnove semantičkog weba i kako on funkcionira, nego je htio pomoću dostupnog i konstantno razvijajućeg softvera automatizirati proces pretraživanja podataka i informacija, dodavanjem semantičkog značenja mrežnim stranicama kako bi ih učinio semantički povezanima i strojnočitljivima. Postoji grana lingvistike pod imenom „semantika" koja se bavi proučavanjem značenja jezičnih znakova a po kojoj je, na koncu, semantički web i dobio ime. Semantički web pridodaje značenje svakoj informaciji u dokumentu čineći ga jednom smislenom cjelinom u kojoj je pristup željenim i traženim podacima brz i učinkovit.

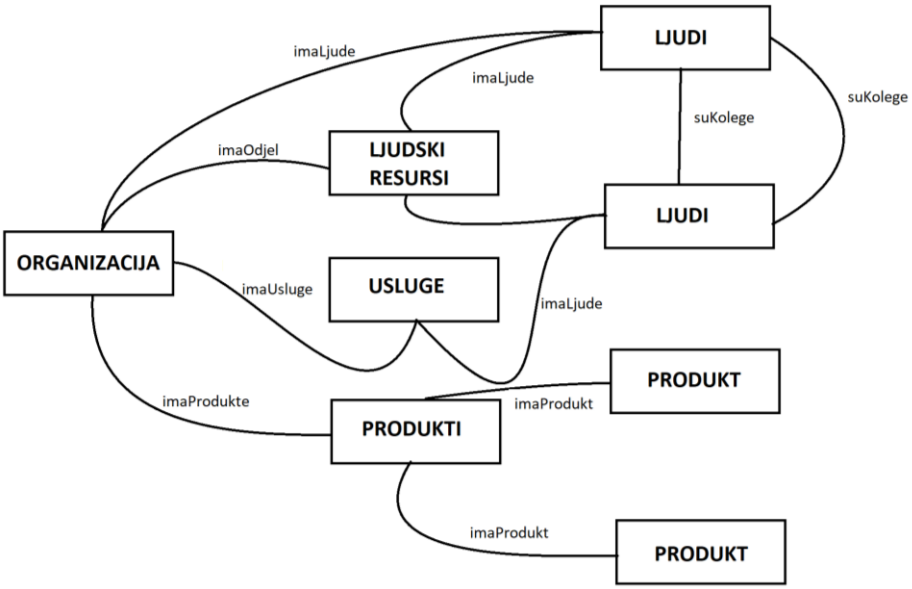
U arhitekturi „starog weba" svaki resurs imao je svoju poveznicu (engl. *link*) s jednim ili više drugih resursa, no ti resursi i poveznice nisu imale ugrađeno semantičko značenje (Slika 1).

[5] Berners-Lee, Tim; Hendler, James; Lissila, Ora. 2001. The Semantic Web. URL: http://sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21. (2022-9-5)
[6]Berners-Lee, Tim; Hendler, James; Lissila, Ora. Nav. dj. (2022-9-5)



Slika 1 - Ilustracija arhitekture weba.[7]

No, kao što je vidljivo na Slici 2, u okviru semantičkog weba, svaki resurs i svaka poveznica imaju i naziv i značenje.



Slika 2 - Ilustracija arhitekture semantičkog weba.[8]

[7] Cardoso Jose, Sheth Amit. The Semantic Web and its Applications, 2006. URL: https://link.springer.com/chapter/10.1007/978-0-387-34685-4_1#citeas (2022-9-6)
[8] Cardoso Jose, Sheth Amit. Nav. dj.

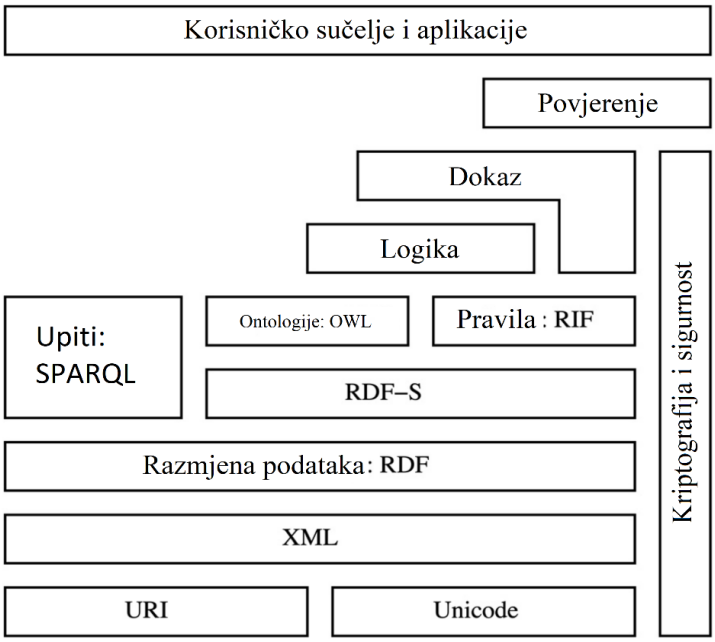
2.2 Tehnologije semantičkog weba

Kako bi se ostvarila ideja strojnočitljivog weba, bilo je potrebno pohranjivati podatke u sintaksi koja bi to omogućila. Semantički web nastoji ogroman broj informacija na webu učiniti lako pristupačnima, ljudski i strojnočitljivima te smanjiti vrijeme koje je potrebno korisnicima, odnosno ljudima da pronađu željene informacije i podatke. Tehnologije semantičkog weba nastoje dodati semantički sloj sadržajima weba. Za tu svrhu upotrjebljavaju se označiteljski jezici (engl. *markup languages*), u najvećem broju slučajeva HTML i XML. HTML je već predstavljen kao označiteljski jezik koji krajnjem korisniku prikazuje mrežni sadržaj, odnosno određuje na koji način se neki sadržaj na webu treba prikazati. HTML omogućuje razne načine definiranja semantike mrežnog sadržaja, a jedan od načina je korištenjem META oznaka ili elemenata. META element u HTML-u, zapravo, predstavlja metapodatke (engl. *metadata*) o resursu ili mrežnoj stranici, a metapodaci su podaci (informacije) o podacima. XML je također označiteljski jezik koji je razvijen kao jezik za strukturiranje dokumenata na webu. XML dokument sastoji se od ugniježđenog skupa otvorenih i zatvorenih elemenata koji se javljaju u obliku parova atribut-vrijednost.[9] Osim navedenih tehnologija, W3 konzorcij objavio je standarde semantičkog weba koji pri zajedničkom korištenju omogućuju i olakšavaju realizaciju ideje pohranjivanja i razmjene semantički obogaćenih informacija. Ti standardi su RDF (*Resource Description Framework*), OWL (*Web Ontology Language*), SPARQL (*Protocol and RDF Query Language*) te SKOS (*Simple Knowledge ORganization System*).[10]

[9] Oriental Journal of Computer Science and Technology. Semantic Web: XML and RDF Roles, 2016. URL: https://www.computerscijournal.org/vol8no3/semantic-web-xml-and-rdf-roles/ (2022-9-6)
[10] Farago, F., Bosančić, B., i Badurina. 'Povezani podaci i knjižnice', Vjesnik bibliotekara Hrvatske, 56(4), B. (2013), str. 25-52. URL: https://hrcak.srce.hr/142376 (2022-9-8)

2.3. Slojevi semantičkog weba

OSI model (engl. *Open Systems Interconnection Model*) predstavlja jedan od načina podjele komunikacijskih sustava u manje dijelove, odnosno slojeve (Slika 3). Svaki je sloj skup konceptualno sličnih funkcija koje djeluju na način da slojevi iznad rabe funkcionalnosti slojeva ispod njih.[11] Ovim se modelom koji je dobio naziv *The Semantic Web Layer Cake* prikazuju slojevi semantičkog weba, a sâm koncept ovakvog prikaza semantičkog weba osmislio je Tim Berners-Lee.



Slika 3 - Prikaz slojeva semantičkog weba.[12]

Najniži sloj u prikazu u prvoj verziji OSI modela semantičkog weba koji je objavljen 2001. sačinjavaju URI/IRI i Unicode.[13] Ovaj sloj odgovoran je za šifriranje ili kodiranje bilo kojeg znaka ili simbola napisanog na bilo kojem svjetskom jeziku. Ujedno, ovaj sloj reprezentira i jedinstveno identificiranje resursa.

[11] J. Domingue, D. Fensel, J. A. Hendler, Handbook of Semantic Web Technologies. Berlin, Germany: Springer, 2011. str. 18 URL: https://pdfdrive.to/pdfs/handbook-of-semantic-web-technologies-pdf (2022-9-8)
[12] Gopal, Pandey. „The Semantic Web: An Introduction and Issues“ // International Journal of Engineering Research and Applications, 1(2012), str. 780 URL: https://www.academia.edu/1959581/IJERA_www_ijera_com_ (2022-9-8)
[13] Unicode. URL: https://home.unicode.org/ (2022-9-23)

Drugi sloj, sloj XML-a, omogućuje ostvarivanje interoperabilnosti u komunikaciji između slojeva. Ovaj sloj pruža temelje za kombiniranje različitih elemenata iz različitih vokabulara formirajući na taj način specifične funkcije.

Prva dva sloja sastoje se od jednostavnih tehnologija weba, a od trećeg sloja počinju se rabiti tehnologije semantičkog weba. U trećem sloju nalazi se RDF, standard za opisivanje web resursa. RDF je model podataka koji se rabi za opisivanje resursa i njihove povezanosti.

Nadalje, četvrti sloj je ontološki sloj. Ontologija se može shvatiti kao instrument kojim se još preciznije opisuju resursi i njihova povezanost. Ontološki sloj podržava razvoj vokabulara za definiranje veza između raznih koncepata. Ontološkom sloju pripadaju OWL, SPARQL, RIF (engl. *Rule Interchange Format*) te tehnologije vezane uz kriptografiju koje osiguravaju da se informacije primaju iz povjerljivih resursa.

Peti sloj naziva se logički sloj (engl. *Logic*) i on je potreban za definiranje pravila prema kojima se provodi zaključivanje u ontologijama.[14]

Dokazni sloj (engl. *Proof*), kao šesti sloj, uključuje deduktivne procese kojima se provjeravaju zaključci prethodnog sloja. Glavna je uloga ovoga sloja prepoznati korisnike i njihove uloge vezane uz semantički sadržaj na webu, npr. je li autoriziran za korištenje X/Y stranice i sl.

Ako neka tvrtka X pošalje podatke tvrtki Y, a tvrtka Y ne može biti sigurna da je proslijeđene informacije poslala tvrtka X, u tom slučaju može se zaključiti da tvrtka X odnosno navedeni izvor podataka nije povjerljiv. Povjerenje se osigurava u sedmom sloju, sloju povjerenja (engl. *Trust*). Zahtijevano povjerenje može se ostvariti Digitalnim potpisom (engl. *Digital Signature*), koji predstavlja matematičku shemu za dokazivanje autentičnosti digitalnih informacija.

Posljednji sloj semantičkog weba čine aplikacije i korisnička sučelja (engl. *UI – User Interface*) koji omogućuje pristup i ljudima i strojevima kreiranoj aplikaciji semantičkog weba.[15]

---

[14] Gopal, Pandey. Nav. dj. str. 781 – 782. URL: https://www.academia.edu/1959581/IJERA_www_ijera_com_ (2022-9-8)
[15] Gopal, Pandey. „The Semantic Web: An Introduction and Issues“ // International Journal of Engineering Research and Applications, 1(2012), str. 780 – 786. URL: https://www.academia.edu/1959581/IJERA_www_ijera_com_ (2022-9-8)

3. Povezani podaci

Semantički web temelji se na poveznicama među podacima tako da ih i ljudi i strojevi mogu pretraživati, formirajući jednu veliku mrežu povezanih podataka koju nazivamo *web of data*. Jedan od temeljnih stupova semantičkog weba koji je omogućio ostvarenje ove vizije stoji iza imena povezanih podataka. Kao što je već navedeno, glavna zamisao semantičkog weba je stvoriti poveznice između skupova podataka koji su razumljivi i čitljivi ne samo ljudima već i strojevima, dok povezani podaci predstavljaju najbolju praksu omogućavanja tih veza i stvaranje smislenih skupova povezanih podataka. Tim Berners-Lee je 2006. godine povezane podatke opisao kao pojam koji se odnosi na skup najboljih praksi za objavljivanje i povezivanje strukturiranih podataka na webu koristeći se ključnim tehnologijama: URI/IRI-jem, HTTP-om i RDF-om.[16] Usvajanje najboljih praksi povezanih podataka imalo je za posljedicu proširenje globalnog podatkovnog prostora u kojem su se povezivali podaci iz različitih domena kao što su: ljudi, knjige, filmovi, glazba, znanstvene publikacije i istraživanja, online zajednica, ustanova, itd. Ova globalna mreža povezanih podataka omogućila je stvaranje novih vrsta aplikacija. Na internetu se mogu pronaći generički pretraživači povezanih podataka koji korisnicima nude mogućnost pretraživanja podataka iz različitih resursa. Korisnik, primjerice, može započeti pretraživati informacije u jednom resursu te navigirati do drugih relevantnih resursa pomoću poveznica unutar početnog resursa. Također, postoje i tražilice povezanih podataka koje pretražuju web podatke prateći poveznice koje su uspostavljene među resursima podataka pružajući mogućnost postavljanja upita, slično načinu na koji se upiti postavljaju u lokalnim bazama podataka. Načelo rada povezanih podataka temelji se na podacima koji su pohranjeni u RDF formatu.[17]

3.1. Načela povezanih podataka

Tim Berners-Lee je 2006. godine osmislio načela povezanih podatka. Postoje četiri načela na kojim se zasniva objavljivanje i povezivanje podataka koja se temelje na postojećoj infrastrukturi, pravilima i standardima weba. Načela povezanih podataka su:

[16] Berners-Lee, Tim; Bizer, Christian; Heath, Tom. Linked Data: The Story so Far. // International journal on Semantic Web and information systems, 5(2009), 1-22. URL: https://www.researchgate.net/publication/225070216_Linked_Data_The_Story_so_Far (2022-9-8)
[17] Berners-Lee, Tim; Bizer, Christian; Heath, Tom. Nav. dj. (2022-9-8)

- Koristiti URI/IRI-je za imenovanje stvari i resursa.
- Koristiti HTTP URI/IRI-je kako bi korisnici mogli pretraživati mrežno dostupne resursee povezanih podataka.
- Osigurati korisne informacije u resursima povezanih podataka na koje vode poveznice, rabeći RDF i SPARQL; drugim riječima, sve „stvari" odnosno resursi moraju biti opisani putem RDF modela podataka u jednom od RDF formata pohrane (Turtle, RDF/XML, itd.).[18]
- U opisima resursa i „stvari" uključiti i poveznice do drugih URI/IRI-ja kako bi korisnik mogao istraživati više resursa povezanih podataka.[19]

Po pitanju prvog načela, postoje dodatne smjernice za oblikovanje imenovanja resursa HTTP URI/IRI-jima. Te smjernice postavio je, također, Tim Berners-Lee, a URI/IRI-je oblikovane po njima nazvao je „*Cool URIs"*. Načela za formiranje takvih URI/IRI-ja mogu se svrstati u tri kategorije:

- **Jednostavnost** (engl. *simplicity*) – kratki, takozvani „mnemonički" URI/IRI-ji imaju manju vjerojatnost neispravnosti u otvaranju kada se šalju elektroničkom poštom te su općenito lakše pamtljivi.
- **Stabilnost** (engl. *stability*) – kada se formira URI/IRI za određeni resurs, bitno je promisliti kako ga učiniti što trajnijim. Razmišljajući unaprijed, mogu se spriječiti nadolazeće prepreke dolaskom novih tehnologija, pa je preporuka da se implementacijsko-specifični dijelovi URI/IRI-ja izostave (npr. .php, .asp), jer će se u budućnosti možda promijeniti tehnologije za osiguravanje njihove pristupačnosti.
- **Upravljivost** (engl. *manageability*) – bitno je učiniti URI/IRI-je lakima za modeliranje, odnosno da se lako mogu vršiti lokalne izmjene, bez utjecaja na ostatak URI/IRI-ja. Jedna dobra praksa za tu svrhu je uvrstiti trenutnu godinu u putanju URI/IRI-ja tako da moguća promjena URI/IRI-ja iz godine u godinu ne utječe na ostale.[20]

Razmatrajući povezane podatke samo iz perspektive navedenih načela može se potvrditi da su URI/IRI i RDF ključne tehnologije na kojima se temelje povezani podaci. Koncept povezanih

---

[18] Bosančić, Boris. Otvoreni povezani podatci i metapodatci. // Otvorenost u znanosti i visokom obrazovanju, 2018, 187-191. URL: https://repozitorij.ffzg.unizg.hr/en/islandora/object/ffzg%3A1487/datastream/FILE0/view#page=182 (2022-9-8)
[19] Berners-Lee, Tim; Bizer, Christian; Heath, Tom. Nav. dj. (2022-9-8)
[20] Cool URIs for the Semantic Web. URL: https://www.w3.org/TR/cooluris/ (2022-9-9)

podataka temelji se tako na metodi stvaranja poveznica između pojedinih „dijelova“ RDF dokumenta kao i drugih RDF dokumenata koji opisuju resurse. Kako bi to bilo ostvarivo, svaki dio RDF dokumenta koji bi htjeli povezati s nekim drugim dijelom postojećeg ili nekog drugog RDF dokumenta mora biti identificiran HTTP URI/IRI-jem. Tako se potvrđuje da su povezani podaci zaista skup dobrih praksi za povezivanje strukturiranih podataka te objavljivanje istih koristeći se ključnim tehnologijama: URI/IRI-jem, HTTP-om i RDF-om.

3.2. Tehnologije povezanih podataka

3.2.1. RDF

RDF predstavlja model podataka za prikaz informacija (podataka) na webu. Svatko tko se koristi RDF-om može opisivati web resurse, dodjeljivati im dodatna značenja i atribute, kao što su autor, datum kreiranja resursa, subjekt i autorsko pravo. Identificiranje web resursa provodi se korištenjem URI/IRI-ja, dok RDF predstavlja model podataka koji se može i grafički prikazati. RDF je prvi puta objavljen u sklopu W3 konzorcija 1999. godine, ali se tijekom prvih godina postojanja neprestano razvijao te je 2004. godine „evoluirao“ u verziju koju danas smatramo RDF standardom. Unatoč tomu što je ova verzija smatrana stabilnom, RDF se nastavlja razvijati i nadograđivati te je 2010. godine W3 konzorcij sazvao sjednicu kako bi se raspravile buduće smjernice vezane uz daljnji razvoj RDF-a.[21] RDF model podataka predstavlja rješenje za svaki informacijski portal ili mrežnu stranicu čija je svrha predstavljanje neke vrste podataka ili baze podataka na semantičkoj razini. Korištenjem URI/IRI-ja te referenciranjem na već postojeće resurse podataka, svatko može dodati neku informaciju i značenje o bilo čemu na webu. Uz to, koristeći se RDFS-om (RDF Schemom) osobe mogu definirati klase i svojstva specifična za određenu domenu kako bi opisali web resurse te ih hijerarhijski organizirali. RDFS predstavlja rječnik za modeliranje RDF podataka.[22] Osim što pruža način za modeliranje podataka kroz grafički prikaz, RDF također predstavlja model za dijeljenje podataka. Pored toga, RDF omogućuje i druge funkcionalnosti: postavljanje upita (engl. *querying*), ugrađivanje (engl. *embedding*) i rasuđivanje (engl. *reasoning*).[23] Postavljanje upita nad RDF dokumentima provodi se već spomenutom tehnologijom SPARQL, koja će biti detaljnije objašnjena u

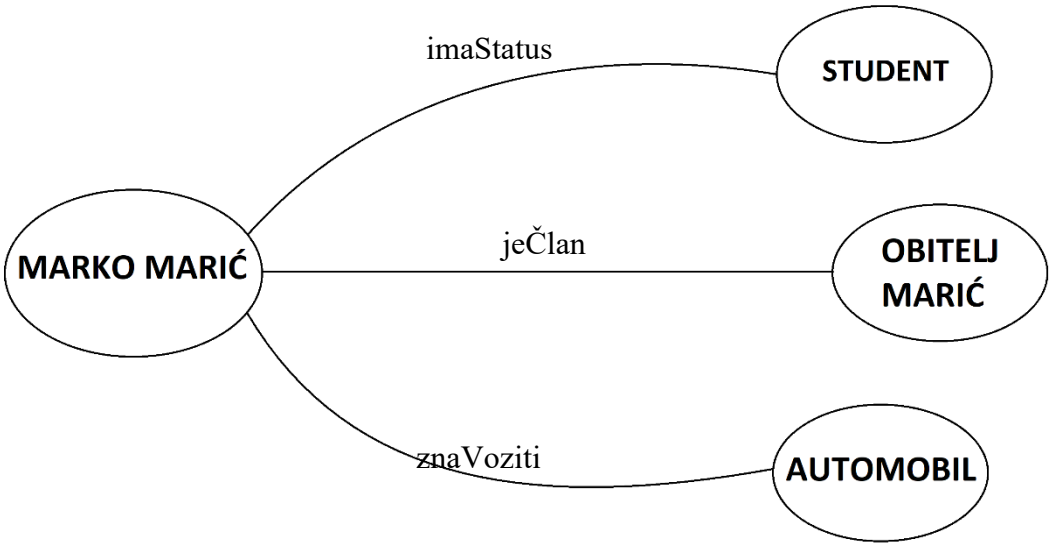
[21] J. Domingue, D. Fensel, J. A. Hendler, Handbook of Semantic Web Technologies. Berlin, Germany: Springer, 2011. str. str 119. URL: https://pdfdrive.to/pdfs/handbook-of-semantic-web-technologies-pdf (2022-9-8).
[22] RDF Schema 1.1, 2014-2-25. URL: https://www.w3.org/TR/rdf-schema/
[23] J. Domingue, D. Fensel, J. A. Hendler, Handbook of Semantic Web Technologies. Berlin, Germany: Springer, 2011. str. str 120. URL: https://pdfdrive.to/pdfs/handbook-of-semantic-web-technologies-pdf (2022-9-8)

sljedećem poglavlju. Ugrađivanje se provodi korištenjem RDFa (ekstenzije HTML5 koja služi za pomoć pri označivanju dijelova podataka kao što su ljudi, događaji, mjesta u okviru HTML dokumenata)[24] i GRDDL (*Gleaning Resource Descriptions from Dialects of Languages*), koji predstavlja alat za dohvaćanje RDF podataka iz XML dokumenata, preciznije, XHTML stranica.[25] RDF je svojim mogućnostima otvorio mnoga vrata za upravljanje podacima te tako postao široko prihvaćen alat za istu svrhu od strane raznih industrijskih i upravnih tijela. Organizacije koje se bave golemim skupinama podataka trebale su interoperabilni, fleksibilni i proširivi model koji sadrži funkcionalnosti za sve potrebne radnje nad podacima, a sve to posjeduje RDF. U načelu, RDF model podataka se sastoji od:

- Resursa (engl. *resource*) - bilo što na što se netko može pozvati smatra se resursom; to može biti mrežna stranica, slika, ali isto tako i osoba, mjesto, događaj, organizacija, itd. Tehnički gledano, sve što se može identificirati URI/IRI-jem smatra se resursom (npr. https://www.w3.org/) .
- Svojstva (engl. *property*) resursa - svojstva resursa predstavljaju skupovi atributa, značajke koje se odnose na resurs (npr. „preuzeto od“, „spremljeno u“ i slično). Svako pojedino svojstvo također mora imati URI/IRI.
- Vrijednosti svojstva (engl. *property value*) mogu biti slovna vrijednost (engl. *literal*), tip podataka (engl. *datatype*) pa čak i drugi resurs, odnosno izvor.[26]

Temeljna struktura sintakse RDF-a počiva na RDF tripletima (engl. *triples*) ili trojkama. Svaki RDF triplet sastoji se od subjekta, predikata i objekta. Subjekt se odnosi na resurs, predikat na svojstvo, a objekt RDF tripleta na vrijednost svojstva.

[24] RDFa. Linked Data in HTML. URL: https://rdfa.info/ (2022-9-8)
[25] W3C. Gleaning Resource Descriptions from Dialects of Languages (GRDDL). URL: https://www.w3.org/2001/sw/wiki/GRDDL (2022-9-8)
[26] J. Domingue, D. Fensel, J. A. Hendler, Handbook of Semantic Web Technologies. Berlin, Germany: Springer, 2011. str. str 121. URL: https://pdfdrive.to/pdfs/handbook-of-semantic-web-technologies-pdf (2022-9-10).



Slika 4 - RDF graf.

Na Slici 4 prikazan je jednostavan primjer RDF grafa koji se sastoji od tri RDF tripleta. U primjeru naveden je jedan subjekt (Marko Marić) na koji se vežu tri predikata (imaStatus, jeČlan i znaVoziti) i tri objekta (STUDENT, OBITELJ MARIĆ, AUTOMOBIL). Gledajući sliku na kojoj je prikazan graf, može se reći da Marko Marić ima status studenta, član je obitelji Marić te zna voziti automobil. Vidljivo je da su na ovaj način putem RDF-a podaci povezani i da je ta povezanost i strojno razumljiva. Nadalje, RDF graf može se nastaviti granati i tako povezivati sve općenitije resurse i pojmove stvarajući globalnu i povezanu mrežu RDF podataka.

Za potrebe prikaza i objašnjenja RDF sintakse na osnovi primjera RDF grafa (Slika 4), izjava „Marko Marić je član obitelji Marić“ prikazat će se u nastavku u RDF/XML notaciji kôda koji se pohranjuje u RDF dokumentu.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns="https://example.org#marko-maric#"
xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:foaf="http://xmlns.com/foaf/0.1/"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

<rdf:Description rdf:about="http://example.org#marko-maric">
```

```
<dcterms:isPartOf>
        <foaf:family_name
        rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
                Obitelj Marić
        </foaf:family_name>
</dcterms:isPartOf>
</rdf:Description>
</rdf:RDF>
```

Na početku dokumenta mogu se uočiti deklaracije rječnika koji se koriste u dokumentu, a to su RDF, DC (*Dublin Core*) i FOAF (*Friend of a Friend*). Subjekt je definiran pomoću atributa `rdf:about` koji se nalazi unutar elementa `rdf:Description` koji obuhvaća sve predikate i objekte, odnosno svojstva i vrijednosti svojstva RDF dokumenta. U atribut `rdf:about` unosi se URI/IRI identifikatora resursa (`http://example.org#marko-maric`). Iz DC vokabulara korišten je element `dcterms:isPartOf` koji naznačuje kako je nešto dio nečega, u ovom slučaju da je osoba dio, odnosno član obitelji, a obitelj je izražena pomoći elementa iz FOAF rječnika `foaf:family_name`. Unutar FOAF elementa nalazi se još jedan RDF atribut `rdf:datatype` koji definira tip podatka koji će se prikazati unutar elementa.

3.2.2. URI/IRI

Već spominjana tehnologija, URI/IRI, najvažnija je vrsta identifikatora za podatke odnosno resurse podataka. RDF koristi URI/IRI-je za jasnu identifikaciju različitih resursa. URI/IRI-ji su nastali generalizacijom URL-a. Svaki URL je ujedno i validan HTTP URI te se URL-ovi mogu koristiti i kao identifikatori u RDF dokumentima. Iako se URI/IRI-ji i URL-ovi u nekim slučajevima mogu naizmjence koristiti, oni nisu identični. URI/IRI identificira resurse te ih razlikuje od drugih resursa pomoću naziva ili lokacije dok URL identificira web adresu, odnosno lokaciju jedinstvenog izvora resursa.[27] URI/IRI-ji se mogu karakterizirati prema sljedećim svojstvima i konceptima:

[27] URI vs URL: Differences and When to Use Them, 2022-8-15. URL: https://www.hostinger.com/tutorials/uri-vs-url (2022-9-12)

- Ujednačenost (engl. *uniform*) – omogućuje različitim tipovima identifikatora resursa da se koriste u istom kontekstu, iako se mehanizmi i proces pristupanja tim resursima razlikuju, omogućujući tako novim aplikacijama i protokolima da se koriste već postojećim mnoštvom skupova identifikatora resursa.
- Resurs (engl. *resource*) – pojam resurs, u općenitom smislu, odnosi se na bilo što, što se može identificirati URI/IRI-jem. Resurs nije nužno samo dostupan putem interneta, resurs se može odnositi na ljudsko biće, korporacije te knjige koje se pohranjuju u knjižnicama. Također, resurs može biti i apstraktan, npr., matematički operator, tipovi ljudskih odnosa ili neke unaprijed zadane vrijednosti (kao u kontroliranim rječnicima termina).
- Identifikator (engl. *identifier*) – identifikator čini specifične informacije unikatnima, on utjelovljuje informacije koje je potrebno razlikovati i odvaja ih od svih ostalih pojmova i stvari u nekoj određenoj domeni ljudskog znanja. Cilj URI/IRI-ja je identificirati, te u isto vrijeme razlikovati resurs od drugih resursa.[28]

URI/IRI je identifikator koji se sastoji od sekvence znakova koji odgovaraju istoimenoj generičkoj sintaksi URI/IRI-ja. Generička sintaksa URI/IRI-ja sastoji se od hijerarhijski uređenog niza komponenti koje se nazivaju shema, autoritet, put, upit i fragment. Generalna konstrukcija sheme URI/IRI-ja može se sažeti u kombinaciju navedenih elemenata, a izgled te konstrukcije prikazan je na Slici 5.

shema: [//autoritet] putanja [?upit] [#fragment]

Slika 5 - Prikaz sheme URI-ja.[29]

Shema URI-ja sastoji se od sljedećih elemenata:

- Shema (engl. *scheme*) – u ovom elementu definira se tip URI/IRI-ja, odnosno naziv sheme ili protokola. Element sheme može pružiti dodatne informacije o tome kako rukovati URI/IRI-jima u aplikaciji. Neki od naziva shema ili protokola su http, mailto,

[28] Berners-Lee, Tim; R. Fielding; L. Masinter. Uniform Resource Identifier (URI) : generic syntax. 2005, str. 4-6. URL: https://greenbytes.de/tech/webdav/rfc3986.pdf (2022-9-15)
[29] Hitzler, Pascal; Krotzsch, Markus; Rudolph, Sebastian. Foundations of Semantic Web Technologies. Philadelphia, Chapman & Hall / CRC, 2011. URL: https://freepdf-books.com/foundations-of-semantic-web-technologies/ str. 23 (2022-9-15)

gopher, itd. No, za potrebe povezanih podataka bitna je samo http shema ili protokol.

- Autoritet/vlasništvo (engl. *authority*) – odnosi se na ime domene s mogućim dodatnim podacima o korisniku i mrežnom priključku (engl. *port*) (primjer: korisnik.org:8080). Ovaj element nije obavezan te se može prepoznati po prefiksu „//“ koji stoji ispred elementa.
- Putanja (engl. *path*) – element putanje je glavni dio mnogih URI/IRI-ja, iako je moguće ovaj element ostaviti prazan, kao na primjer u email adresama. Putanje se mogu hijerarhijski organizirati koristeći se znakom „/“ kojim se razdvajaju dijelovi putanje. Primjer: /ffos/knjiznica/upiti, putanja/bez/autoriteta (putanje koje ne počinju znakom „/“ su pristupačne samo ako nema zadanog autoriteta/vlasništva).
- Upit (engl. *query*) – upiti su neobavezni dio URI/IRI-ja, pružaju dodatne informacije o resursima, a prepoznaju se po znaku „?“ koji stoji ispred elementa.
- Fragment (engl. *fragment*) – opcionalni element koji pruža novu razinu identifikacije resursa te se prepoznaje po znaku „#“ koji se nalazi ispred elementa. URI/IRI-ji s različitim fragmentima se razlikuju iako vode do istog dokumenta.[30]

Za formiranje URI/IRI-ja koriste se znakovi latiničnog pisma uključujući i brojeve. Korištenje ne-latiničnih znakova koji su uobičajeni u drugim jezicima te se često pojavljuju također su validni znakovi za formiranje URI/IRI-ja u svim trenutnim formatima semantičkog weba. URI/IRI-ji koji se proširuju navedenim znakovima nazivaju se IRI-jima (engl. *International Resource Identifiers*) te se mogu koristiti na svim područjima identifikacije resursa kao i URI/IRI.[31]

3.2.3. SPARQL

Kreiranjem što većeg broja povezanih, strojnočitljivih skupova povezanih podataka ostvaruje se prvi korak do ostvarenja semantičkog weba. No, potrebno je i pokazati i korist tog postupka. Kako bi se demonstrirale mogućnosti tehnologije povezanih podataka u pretraživanju i analizi istih, potrebno je poznavati korištenje RDF-ovog jezika za upite (engl. *query langauge*) koji se zove SPARQL. SPARQL nije akronim, puno ime ovog jezika je *SPARQL Protocol and RDF Query Language*. Prema tome, SPARQL predstavlja upitni jezik za pretraživanje RDF dokumenata, kao što je SQL upitni jezik koji se rabi za pretraživanje relacijskih baza podataka.

[30] Hitzler, Pascal; Krotzsch, Markus; Rudolph, Sebastian. Nav. dj. str. 23 (2022-9-15)
[31] Isto. str. 23 (2022-9-15)

Sintakse SPARQL-a i SQL-a djelomično se podudaraju, ali to je jedina sličnost koju dijele.[32] Glavne razlike SPARQL-a i SQL-a su:

- SPARQL ne sadrži specifične odredbe jezika za definiranje podataka, odnosno DDL (engl. *Data Definition Language*), jer su sheme predstavljene i u RDFS-u i OWL-u (*Web Ontology Language*) kao standardni RDF grafovi, te stoga ne zahtijevaju specifični jezik.[33]
- U kontrastu sa SQL-om, SPARQL upiti nisu ograničeni na rad unutar jedne baze podataka. Udruženi upiti mogu pristupati višestrukim spremištima podataka.[34]

SPARQL koristimo kako bi provodili upite bilo nad lokalnom RDF datotekom s podacima u obliku RDF trojki, ili na dokumentima bilo gdje na webu. Moguće je, dakle, kombinirati bezbroj upita, postavljenih lokalno ili „udaljeno“, kako bi se dobili traženi podaci u aplikaciji što predstavlja istaknuto svojstvo SPARQL-a i povezanih podataka.[35]

Zbog korištenja SPARQL-a u drugom dijelu rada, u sljedećem će se poglavlju dati i osvrt na njegovu sintaksu kroz konkretan primjer.

3.2.4. SPARQL sintaksa

U procesu formiranja upita u SPARQL-u koristimo se varijablama. Varijable služe kao spremnici rezultata pretrage nad RDF dokumentom te se mogu privremeno koristiti u svrhu filtriranja rezultata. Svaka varijabla mora biti označena znakom upitnika ispred imena te ih nije potrebno unaprijed deklarirati. SPARQL ima definirane standardne naredbe za dobivanje rezultata i postavljanje upita:

- **SELECT** – ovom naredbom se prikazuju svi podaci koji odgovaraju kriterijima koji su zadani.
- **CONSTRUCT** – naredba koja omogućuje modificiranje podataka na temelju već

[32] Wang ,Yongming; Q. Yang, Sharon. Linked Data Technologies and What Libraries Have Accomplished So Far. // International Journal of Librarianship, Vol. 3, 1(2018), str. 10- 12 URL: https://journal.calaijol.org/index.php/ijol/article/view/62 (2022-9-18)
[33] Domingue, D. Fensel, J. A. Hendler, Handbook of Semantic Web Technologies. Berlin, Germany: Springer, 2011. str. 241 URL: https://pdfdrive.to/pdfs/handbook-of-semantic-web-technologies-pdf (2022-9-18)
[34] Ontotext. What is SPARQL. URL: https://www.ontotext.com/knowledgehub/fundamentals/what-is-sparql/ (2022-9-23)
[35] Wang ,Yongming; Q. Yang, Sharon. Linked Data Technologies and What Libraries Have Accomplished So Far. // International Journal of Librarianship, Vol. 3, 1(2018), str. 10- 12 URL: https://journal.calaijol.org/index.php/ijol/article/view/62 (2022-9-18)

postojećih podataka u skupu podataka.

- **ASK** – ovom naredbom provjerava se postoji li određeni uzorak, odnosno primjerak u skupu podataka; naredba ASK kao rezultat upita vraća samo dvije vrijednosti: „true“ ili „false“.
- **DESCRIBE** – naredba koja daje uvid u RDF graf koji opisuje specifični resurs.[36]

Osim glavnih naredbi SPARQL također nudi i neke dodatne opcije:

- **PREFIX** – definira imenske prostore rječnika koji se rabe u SPARQL upitima.
- **FROM** – definira skupove podataka ili grafova nad kojima se vrše upiti.
- **WHERE** – specificira uzorak upita za skup podataka tako da prikazuje podatke koji se podudaraju na temelju oblika upita.[37]

Kombinacijom navedenih naredbi i dodatnih opcija formiraju se upiti koji odgovaraju određenim kriterijima i željama. U svrhu dodatnog objašnjenja sintakse SPARQL upitnog jezika, prikazat će se jednostavan primjer u kojem se provode jednostavni upiti nad lokalnim RDF dokumentom. U primjeru se koristi datoteka pod imenom *pero.rdf* sa sljedećim sadržajem:

```
prefix foaf: http://xmlns.com/foaf/0.1/ .
prefix rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns .
prefix schema: http://schema.org/ .
prefix xsd: http://www.w3.org/2001/XMLSchema# .

http://example.org#pero rdf:type foaf:person .
http://example.org#pero schema:birthdate "1996-04-01"^^xsd:date .
http://example.org#pero foaf:knows http://example.org#marica.
http://example.org#pero foaf:knows http://example.org#leon.
```

RDF dokument sadrži podatke o Peri, njegovom rođendanu i ljudima koje poznaje. Pomoću SPARQL-a, može se formirati upit kojim se pronalaze i prikazuju svi Perini prijatelji u datoteci *pero.rdf* i to se postiže sljedećim upitom:

---

[36] Cambridge Semantics. SPARQL Query and Command Clauses. URL: https://docs.cambridgesemantics.com/anzograph/v2.4/userdoc/sparql-queries.htm (2022-9-18)
[37] Isto. (2022-9-18)

```
PREFIX foaf: http://xmlns.com/foaf/0.1/ .

SELECT ?x

FROM pero.rdf

WHERE {

http://example.org#pero foaf:knows ?x ;

}
```

PREFIX definira rječnik koji se rabi, a u ovom slučaju to je FOAF, strojnočitljiva ontologija čija je svrha opisivanje ljudi i njihovih odnosa. Nadalje, definirana je varijabla koja će predstavljati rezultate, a to je varijabla „x“ ispred koje stoji znak upitnik. Konačno, u naredbi „WHERE“ unosi se kriterij po kojem se rezultati trebaju prikazati, odnosno specificira se predikat (*knows*) koji povezuje subjekt (Peru) s osobama koje on poznaje. Pritom je Peri dodijeljen HTTP URI/IRI „http://example.org#pero“. Unutar naredbe, „?x“ predstavlja varijablu u koju se spremaju rezultati upita. U rečeničnom obliku, upit bi glasio ovako: „Prikaži sve osobe (objekte) koje Pero (subjekt) poznaje (predikat). Provedbom upita dobiva se sljedeći rezultat:

x
http://example.org#marica
http://example.org#leon

Ovo je jednostavan primjer SPARQL upita nad jednostavnom RDF datotekom, ali u radu s većim skupinama podataka, ovi upiti mogu se formirati na puno kompleksniji način uključujući puno veći broj uvjeta za filtriranje rezultata.

3.2.5. JSON

JSON (*JavaScript Object Notation*) predstavlja format za serijalizaciju podataka i „slanje poruka“ među dokumentima.[38] U postupku kreiranja skupa povezanih podataka u drugom dijelu rada JSON će se koristiti za konverziju tabličnih u povezane podatke. JSON koristi konvencije koje se koriste u jezicima kao što su C, C++, Java i sl., što ga čini idealnim jezikom za zadovoljavanje potreba za razmjenom podataka. JSON objekti sastoje se od dvije vrste

[38] W3C. JSON-LD 1.1. URL: https://www.w3.org/TR/json-ld11/#the-context (2022-9-26)

strukture podataka, a to su:

- Zbirka parova u obliku svojstvo-vrijednost.
- Uređen popis vrijednosti.[39]

JSON svojom jednostavnošću pruža prikladan format podataka pomoću kojeg je moguće prikazivati podatke u obliku RDF grafova. Novije izdanje JSON-a, JSON-LD, pruža mogućnost dodavanja konteksta podacima i time formiranje povezanih podataka koji se vrlo lako procesuiraju.[40]

[39] Pandey, Mrinal; Pandey, Rajiv. JSON and its use in Semantic Web. // International Journal of Computer Applications (0975-8887), 11(2017). Str. 10-11. URL: https://www.researchgate.net/profile/Mrinal-Pandey-3/publication/316250952_JSON_and_its_use_in_Semantic_Web/links/5925af4baca27295a8ebb055/JSON-and-its-use-in-Semantic-Web.pdf (2022-9-28)
[40] Isto. Str. 11 (2022-9-28)

4. Kreiranje skupa povezanih podataka na primjeru skupa podataka iz domene knjižničnog poslovanja

4.1. Općenito o metodologiji

Nakon upoznavanja s tehnologijama semantičkog weba i povezanih podataka, u ovom poglavlju putem odgovarajuće metodologije pokazat će se način kreiranja skupa povezanih podataka (u RDF formatu) od nekoliko tabličnih skupova podataka (objedinjenih u jednom MS Excel dokumentu), a zatim pokazati kako se kreirani skup povezanih podataka može iskoristiti u daljnjoj analizi za potrebe ostvarivanja učinkovitijeg poslovanja knjižnica i srodnih ustanova. Ova dva postupka u skladu sa svrhom rada predstavljaju i njegove ciljeve. U radu se kombinirano rabe metoda modeliranja i metoda prototipa. Metoda modeliranja rabi se u postupku odabira vrste podataka koja će se prikupljati i biti zastupljena u skupu povezanih podataka, dok se primjena metode prototipa ogleda u primjenjivosti kreiranog skupa povezanih podataka i na druge aspekte knjižničnog poslovanja odnosno na mogućnost kreiranja drugih skupova povezanih podataka s različitim vrstama drugih podataka, ali na osnovi istog, u radu opisanog postupka. U postupku kreiranja skupa povezanih podataka korišteni su programi MS Excel, *OpenRefine* i *GraphDB*.

4.2. Opis načina primjene metode modeliranja i metode prototipa u kreiranju skupa povezanih podataka

Odabrani primjer vezan je uz prikupljanje podataka o posjećenosti mrežnih stranica deset hrvatskih knjižnica tijekom mjeseca lipnja 2022. Zbog jednostavnosti, podaci o posjećenosti mrežnim stranicama knjižnica preuzeti su sa *Semrush*[41] stranice, ali važno je napomenuti da se ovim postupkom mogla obuhvatiti i fizička posjećenost istim knjižnicama, kao i bilo koji drugi aspekt poslovanja koji se želi analizirati s ciljem podizanja njegove učinkovitosti.

Prvi korak odnosio se na definiranje podataka koje se željelo prikupiti u skupu podataka. Drugim riječima, uz pretpostavku da će ovakvu vrstu podataka prikupljati osoba bez naprednijeg tehničkog znanja (npr. knjižničar sa srednjom stručnom spremom), kreirana je tablica u programu MS Excel, sa zaglavljem u kojem su navedene vrste podataka, i u koju će se prikupljati podaci. Za praćenje posjećenosti mrežnih stranica knjižnica, odabrana su sljedeće vrste podataka: naziv knjižnice, akronim knjižnice, URI/IRI knjižnice, posjećenost knjižnice po

[41] Semrush. URL: https://www.semrush.com/?l=en&1664208785(2022-9-18)

danu te konačno datum mjerenja posjećenosti. Za potrebe prilagodbe konverzije tabličnih u RDF podatke, u prvom polju tablice nužno je postavljen identifikator svakog zapisa. Identifikator zapisa mora biti jedinstven i mora imati oblik HTTP URI/IRI-ja. Poželjno je i da leksički naznači svrhu podatkovnog zapisa (npr. knjiznice-posjecenost) (Slika 6).

	A	B	C	D	E	F
1	Posjecenost	Naziv_knjiznice	Akronim_knjiznice	URI_knjiznice	Posjecenost_knjiznice_po_danu	Datum_posjecenosti
2	https://oziz.ffos.hr/knjiznice-posjecenost#1	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	225	01/06/2022
3	https://oziz.ffos.hr/knjiznice-posjecenost#2	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	225	02/06/2022
4	https://oziz.ffos.hr/knjiznice-posjecenost#3	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	230	03/06/2022
5	https://oziz.ffos.hr/knjiznice-posjecenost#4	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	231	04/06/2022
6	https://oziz.ffos.hr/knjiznice-posjecenost#5	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	231	05/06/2022
7	https://oziz.ffos.hr/knjiznice-posjecenost#6	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	231	06/06/2022
8	https://oziz.ffos.hr/knjiznice-posjecenost#7	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	231	07/06/2022
9	https://oziz.ffos.hr/knjiznice-posjecenost#8	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	230	08/06/2022
10	https://oziz.ffos.hr/knjiznice-posjecenost#9	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	230	09/06/2022
11	https://oziz.ffos.hr/knjiznice-posjecenost#10	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	230	10/06/2022
12	https://oziz.ffos.hr/knjiznice-posjecenost#11	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	230	11/06/2022
13	https://oziz.ffos.hr/knjiznice-posjecenost#12	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	230	12/06/2022
14	https://oziz.ffos.hr/knjiznice-posjecenost#13	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	230	13/06/2022
15	https://oziz.ffos.hr/knjiznice-posjecenost#14	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	230	14/06/2022
16	https://oziz.ffos.hr/knjiznice-posjecenost#15	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	216	15/06/2022
17	https://oziz.ffos.hr/knjiznice-posjecenost#16	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	216	16/06/2022
18	https://oziz.ffos.hr/knjiznice-posjecenost#17	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	216	17/06/2022
19	https://oziz.ffos.hr/knjiznice-posjecenost#18	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	236	18/06/2022
20	https://oziz.ffos.hr/knjiznice-posjecenost#19	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	245	19/06/2022
21	https://oziz.ffos.hr/knjiznice-posjecenost#20	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	245	20/06/2022
22	https://oziz.ffos.hr/knjiznice-posjecenost#21	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	246	21/06/2022
23	https://oziz.ffos.hr/knjiznice-posjecenost#22	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	244	22/06/2022
24	https://oziz.ffos.hr/knjiznice-posjecenost#23	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	244	23/06/2022
25	https://oziz.ffos.hr/knjiznice-posjecenost#24	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	238	24/06/2022
26	https://oziz.ffos.hr/knjiznice-posjecenost#25	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	239	25/06/2022
27	https://oziz.ffos.hr/knjiznice-posjecenost#26	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	241	26/06/2022
28	https://oziz.ffos.hr/knjiznice-posjecenost#27	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	232	27/06/2022
29	https://oziz.ffos.hr/knjiznice-posjecenost#28	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	233	28/06/2022
30	https://oziz.ffos.hr/knjiznice-posjecenost#29	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	231	29/06/2022
31	https://oziz.ffos.hr/knjiznice-posjecenost#30	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	228	30/06/2022

Slika 6 - Prikaz tablice s podacima o posjećenosti mrežnih stranica hrvatskih knjižnica.[42]

Na Slici 6 prikazana je tablica u programu MS Excel s prvih 30 zapisa o posjećenosti mrežne stranice Gradske i sveučilišne knjižnice Osijek (GISKO). Za 10 knjižnica i 30 dana mjeseca lipnja prikupljeno je sveukupno 300 zapisa, odnosno 30 zapisa po knjižnici. Kao što je vidljivo iz tablice, vrlo je bitno da se na kraju svakog identifikatora zapisa nalazi redni broj zapisa.

Za potrebe kreiranja skupa povezanih podataka na osnovi podataka iz navedene tablice s objedinjenim manjim skupovima podataka koje je mogla generirati i svaka knjižnica zasebno, najprije je potrebno kreirati predložak (engl. *template*) za pretvaranje podataka iz tabličnog zapisa u RDF zapis. Predloškom se ujedno definiraju i rječnici koji će se koristiti za opisivanje podataka koji će se prikazati u skupu povezanih podataka. Isto tako, u predlošku se definira i kako će se opisani podaci prikazati nakon kreiranja skupa povezanih podataka. Ovaj oblik

[42] Tablica s podacima o posjećenosti mrežnih stranica knjižnica može se pronaći na github repozitoriju: https://github.com/msostaric-hub/Posjecenost-knjiznica-RDF

predloška specifičan je po tome što ga zahtjeva program *OpenRefine* kako bi tablični zapis pretvorio u RDF zapis povezanih podataka. Predložak koji se rabi u ovom primjeru prikazan je u nastavku.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns="https://oziz.ffos.hr/knjiznice-posjecenost#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:skos="http://www.w3.org/2004/02/skos/core#"
xmlns:foaf="http://xmlns.com/foaf/0.1/"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
```

U prvom dijelu predloška definiraju se rječnici koji će se koristiti. Navedeni su imenski prostori (engl. *namespace*) nekoliko rječnika, uključivo i FOAF i SKOS (*Simple Knowledge Organization System)* koji se nisu koristili. Nadalje, u sljedećem dijelu predloška formira se XML kôd za predstavljanje podataka u redovima (engl. *rows*). U ovom dijelu važno je pravilno konstruirati sintaksu kojom će se u konačnici program služiti kako bi konstruirao redove podataka u RDF datoteci.

```
<posjecenost_knjiznice
rdf:about="{{jsonize(cells["Posjecenost"].value).replace('"', '')}}">
      <naziv>{{jsonize(cells["Naziv_knjiznice"].value).replace('"',
'')}}</naziv>
      <akronim>{{jsonize(cells["Akronim_knjiznice"].value).replace('"',
'')}}</akronim>
      <uri
rdf:resource="{{jsonize(cells["URI_knjiznice"].value).replace('"', '')}}" />
      <posjecenost rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">
      {{jsonize(cells["Posjecenost_knjiznice_po_danu"].value).replace('"',
'')}}</posjecenost>
      <datum
rdf:datatype="http://www.w3.org/2001/XMLSchema#date">{{jsonize(cells["Datum
_posjecenosti"].value).replace('"', '')}}</datum>
</posjecenost_knjiznice>
```

U predlošku se kao nositelj svakog zapisa rabi element <posjecenost_knjiznice> u kojoj se navodi URI/IRI subjekta zapisa, preuzet iz prvog stupca tablice. Radi se o URI/IRI-ju koji je osmišljen specifično za ovaj primjer: https://oziz.ffos.hr/knjiznice-posjecenost# i kojemu se pridodaju redni brojevi zapisa. Nadalje se definiraju podaci koji će se prikazivati nakon ekstrahiranja iz tablice. Element <naziv> prikuplja podatke o nazivima knjižnica (polje Naziv_knjiznice u tablici), element <akronim> podatke o akronimu knjižnice (polje Akronim_knjiznice u tablici), element <uri> URI/IRI-je knjižnice (polje URI_knjiznice u tablici), element <posjecenost> posjećenost svake knjižnice u jednom danu (polje Posjecenost_knjiznice_po_danu) te element <datum> koji ekstrahira podatke iz tablice koji se odnose na datum mjerenja posjećenosti (polje Datum_posjecenosti u tablici). Broj posjetitelja mrežne stranice knjižnice izražava se u obliku broja (engl. *integer*) pa se za element <posjecenost> može definirati tip podatka koji će biti unesen (rdf:datatype="http://www.w3.org/2001/XMLSchema#integer"). Isto tako, i za element <datum> definiran je tip podatka koji će biti unesen (rdf:datatype=http://www.w3.org/2001/XMLSchema#date). Nakon što je stvoren predložak, isti je zajedno s tablicom u MS Excelu potrebno učitati u program *OpenRefine* kako bi se kreirao skup povezanih podataka.

4.3. Kreiranje skupa povezanih podataka u programu *OpenRefine*

OpenRefine je besplatna aplikacija otvorenoga kôda koja služi za upravljanje svim tipovima podatkovnih dokumenata. *OpenRefine* je započeo kao projekt pod imenom *Freebase Gridworks* koji je otkupio Google i preimenovao ga u *GoogleRefine* 2010. godine. Službena potpora Googlea projektu prestala je 2012. godine te je *OpenRefine* prešao u aplikaciju otvorenog kôda pod istim imenom. *GoogleRefine* i *OpenRefine* identične su aplikacije pa se većina praktičnih vodiča (engl. *tutorial*) za *GoogleRefine* može koristiti i za *OpenRefine*. *OpenRefine* je moćan alat koji znatno olakšava rad s tzv. neurednim podacima (end. *messy data*), on ih pročišćava, pretvara iz jednog formata u drugi i istovremeno proširuje i povezuje s web uslugama i vanjskim podacima. Program štiti podatke korisnika, čineći ih privatnima na korisnikovim osobnim računalima sve dok se korisnik ne odluči podijeliti ih ili surađivati s ostalim korisnicima ili ustanovama. Naime, *OpenRefine* za svaki projekt stvara privatni server na računalu na kojem je projekt kreiran te se može koristiti na webu, ali ga samo vlasnik projekta može vidjeti i manipulirati njime dok ne odluči drugačije. *OpenRefine* baziran je na Javi te radi na bilo kojem operativnom sustavu u mrežnom pregledniku. Također, cijeli program je dostupan na više od

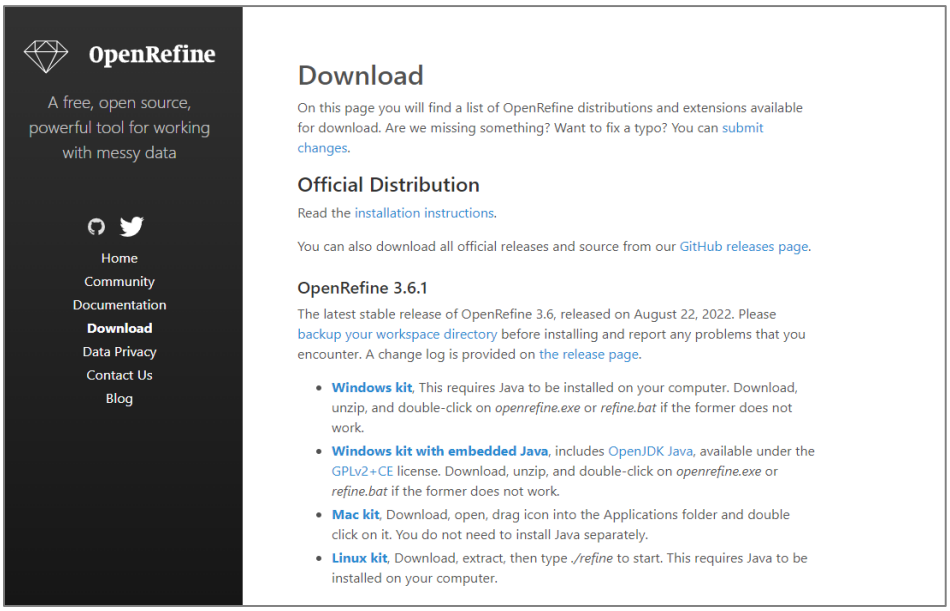
15 jezika.[43] *OpenRefine* dio je neprofitne organizacije CS&C (engl. *Code for Science and Society*) čija je misija unaprjeđivati kvalitetu i vrijednost podataka s ciljem poboljšanja socioekonomskih uvjeta života ljudi kroz javno obrazovanje, znanstvena istraživanja, razvoj tehnologija i njihovu implementaciju[44]. Kao što je već spomenuto, korištenjem *OpenRefinea* može se istraživati, analizirati i povezivati podatke u velikom razmjeru, a navedene funkcionalnosti uključuju:

- Normalizaciju podataka
- Reorganizaciju tabličnih stupaca
- Fasetiranje i grupiranje podataka
- Operacije praćenja promjena
- Izvoz podataka.[45]

U program se mogu uvesti različite vrste datoteka: datoteke s podacima odvojenim tabulatorima (.tsv), datoteke s podacima odvojenim zarezima (.csv), Excel (.xls), JSON, XML, RDF/XML te Google-ove proračunske tablice. Ostale vrste datoteka mogu se dodati pomoću *OpenRefine* ekstenzija.

Za početak, potrebno je preuzeti i instalirati program *OpenRefine*, i to se može obaviti preko mrežne stranice istoimenog programa; preuzimanje i instalacija su besplatni.

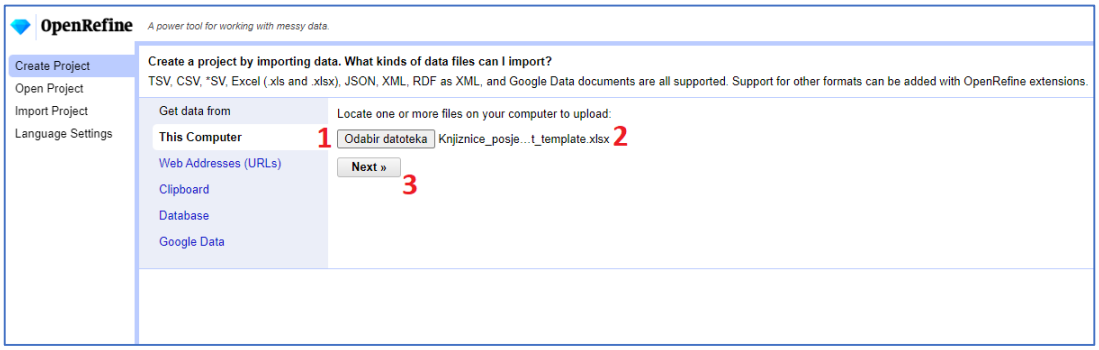
[43] OpenRefine. Documentation. URL: https://openrefine.org/documentation.html (2022-9-18)
[44] Code for Science & Society. URL: https://codeforscience.org/about/ (2022-9-18)
[45] Illinois Library. OpenRefine: About OpenRefine. URL: https://guides.library.illinois.edu/openrefine/about (2022-9-18)



Slika 7 - Isječak mrežne stranice za preuzimanje programa *OpenRefine*.[46]

Kao što se može vidjeti na Slici 7, na mrežnoj stranici programa postoje instrukcije za instalaciju kao i dokumentacija za korištenje, te se svaka informacija vezana uz program može lako pronaći.

Nakon uspješne instalacije program se može pokrenuti. Program se otvara u mrežni pregledniku i kompatibilan je sa sljedećim mrežnim preglednicima: Google Chrome, Chromium, Opera, Microsoft Edge, i Safari. U ovom primjeru, koristit će se mrežni preglednik Google Chrome. Nakon pokretanja, u mrežnom pregledniku se prikazuje početna stranica programa kao što je prikazano na slici 8.

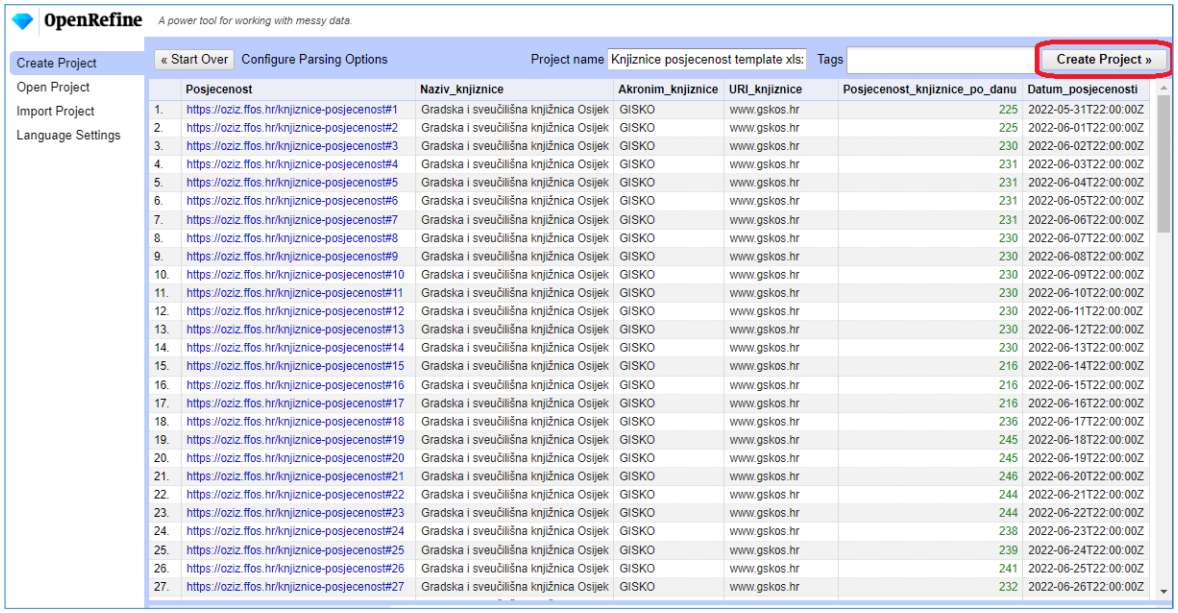


Slika 8 - Početni zaslon *OpenRefine* programa.

[46] OpenRefine. Download. URL: https://openrefine.org/download.html (2022-9-28)

Na početnom zaslonu *OpenRefine* programa podaci se unose u program klikom na gumb Odabir datoteka (1). Podaci se mogu unijeti preko datoteke s našeg računala, web adrese (URL-a), direktno iz međuspremnika (engl. *clipboard*), iz baze podataka ili iz Googleove proračunske tablice (engl. *Google Spreadsheet*). U primjeru dodana je datoteka *Knjiznice_posjecenost_template.xlsx* koja sadrži tablični prikaz prikupljenih podataka (Slika 6). Nakon uvoza datoteka je prikazana pored gumba Odabir datoteka (2) na Slici 8. Nakon provjere ispravnosti datoteke potrebno je kliknuti na gumb *Next* (3) na Slici 8.

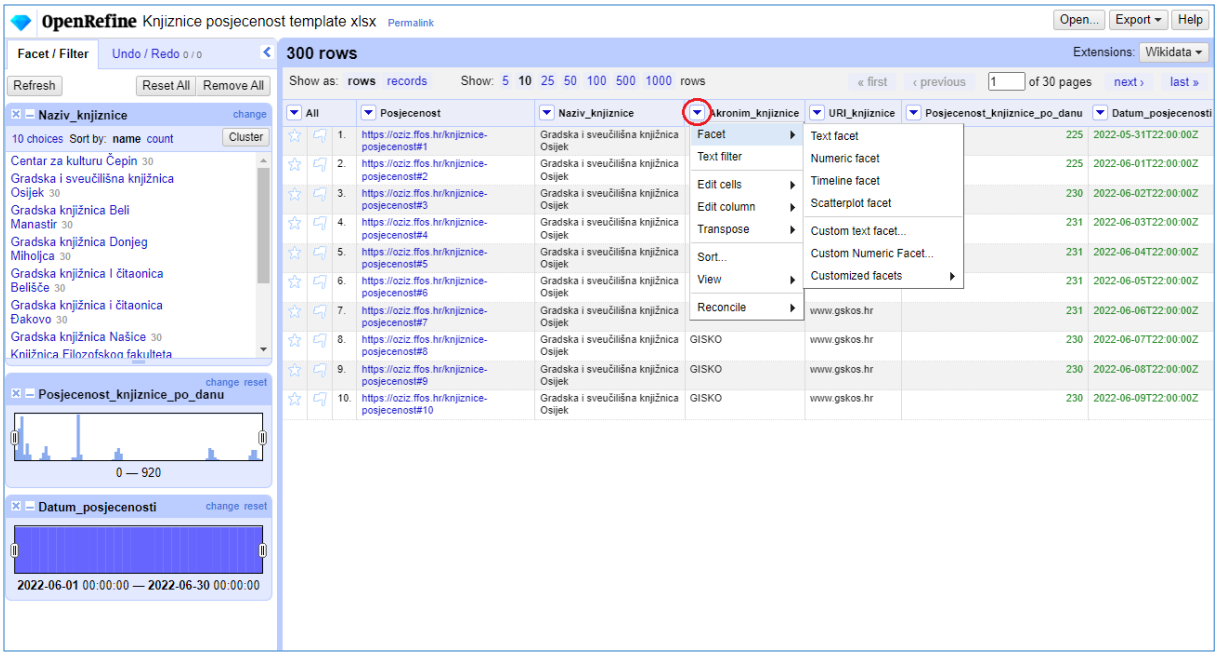
Nakon uspješnog učitavanje datoteke, slijedi zaslon na kojem se mogu vidjeti učitani podaci iz tablice. Nakon provjere odgovaraju li stupci u programu stupcima podataka u tablici, potrebno je kliknuti na crvenom bojom zaokružen gumb *Create Project*. (Slika 9).



	Posjecenost	Naziv_knjiznice	Akronim_knjiznice	URI_knjiznice	Posjecenost_knjiznice_po_danu	Datum_posjecenosti
1.	https://oziz.ffos.hr/knjiznice-posjecenost#1	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	225	2022-05-31T22:00:00Z
2.	https://oziz.ffos.hr/knjiznice-posjecenost#2	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	225	2022-06-01T22:00:00Z
3.	https://oziz.ffos.hr/knjiznice-posjecenost#3	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	230	2022-06-02T22:00:00Z
4.	https://oziz.ffos.hr/knjiznice-posjecenost#4	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	231	2022-06-03T22:00:00Z
5.	https://oziz.ffos.hr/knjiznice-posjecenost#5	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	231	2022-06-04T22:00:00Z
6.	https://oziz.ffos.hr/knjiznice-posjecenost#6	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	231	2022-06-05T22:00:00Z
7.	https://oziz.ffos.hr/knjiznice-posjecenost#7	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	231	2022-06-06T22:00:00Z
8.	https://oziz.ffos.hr/knjiznice-posjecenost#8	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	230	2022-06-07T22:00:00Z
9.	https://oziz.ffos.hr/knjiznice-posjecenost#9	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	230	2022-06-08T22:00:00Z
10.	https://oziz.ffos.hr/knjiznice-posjecenost#10	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	230	2022-06-09T22:00:00Z
11.	https://oziz.ffos.hr/knjiznice-posjecenost#11	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	230	2022-06-10T22:00:00Z
12.	https://oziz.ffos.hr/knjiznice-posjecenost#12	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	230	2022-06-11T22:00:00Z
13.	https://oziz.ffos.hr/knjiznice-posjecenost#13	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	230	2022-06-12T22:00:00Z
14.	https://oziz.ffos.hr/knjiznice-posjecenost#14	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	230	2022-06-13T22:00:00Z
15.	https://oziz.ffos.hr/knjiznice-posjecenost#15	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	216	2022-06-14T22:00:00Z
16.	https://oziz.ffos.hr/knjiznice-posjecenost#16	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	216	2022-06-15T22:00:00Z
17.	https://oziz.ffos.hr/knjiznice-posjecenost#17	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	216	2022-06-16T22:00:00Z
18.	https://oziz.ffos.hr/knjiznice-posjecenost#18	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	236	2022-06-17T22:00:00Z
19.	https://oziz.ffos.hr/knjiznice-posjecenost#19	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	245	2022-06-18T22:00:00Z
20.	https://oziz.ffos.hr/knjiznice-posjecenost#20	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	245	2022-06-19T22:00:00Z
21.	https://oziz.ffos.hr/knjiznice-posjecenost#21	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	246	2022-06-20T22:00:00Z
22.	https://oziz.ffos.hr/knjiznice-posjecenost#22	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	244	2022-06-21T22:00:00Z
23.	https://oziz.ffos.hr/knjiznice-posjecenost#23	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	244	2022-06-22T22:00:00Z
24.	https://oziz.ffos.hr/knjiznice-posjecenost#24	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	238	2022-06-23T22:00:00Z
25.	https://oziz.ffos.hr/knjiznice-posjecenost#25	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	239	2022-06-24T22:00:00Z
26.	https://oziz.ffos.hr/knjiznice-posjecenost#26	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	241	2022-06-25T22:00:00Z
27.	https://oziz.ffos.hr/knjiznice-posjecenost#27	Gradska i sveučilišna knjižnica Osijek	GISKO	www.gskos.hr	232	2022-06-26T22:00:00Z

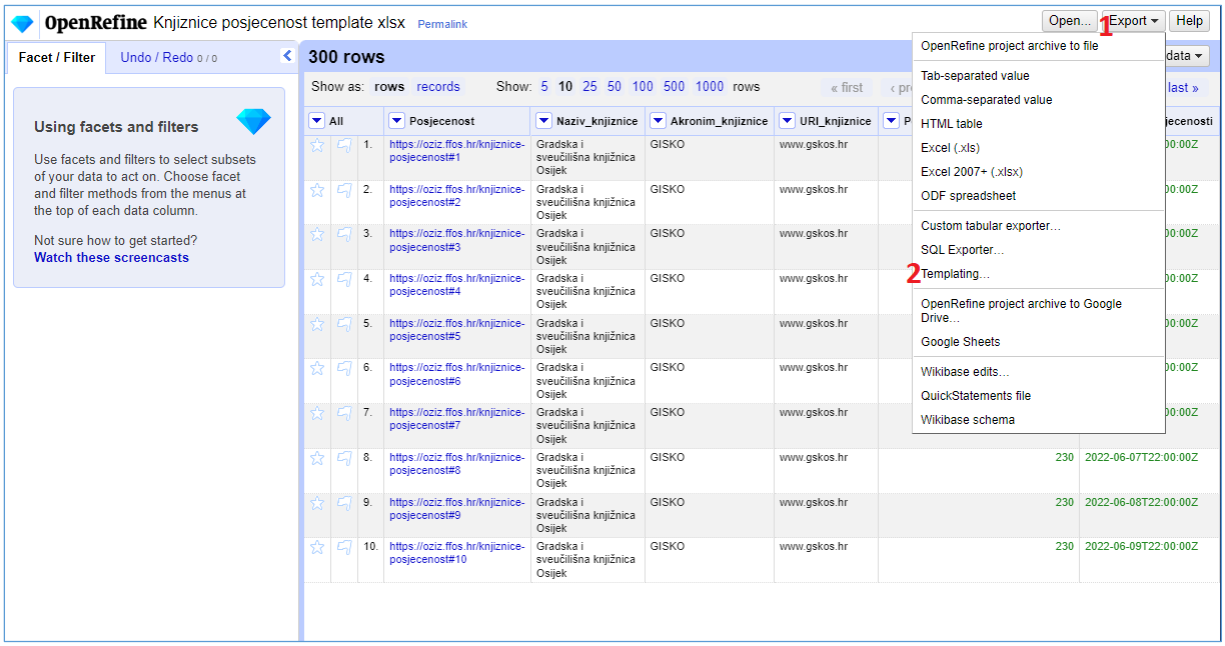
Slika 9 - Prikaz zaslona nakon postavljanja datoteke s podacima.

Na novom zaslonu prikazan je učitan projekt na kojem su podaci prikazani na sličan način kao i na prošlom zaslonu, no razlika je u tome što na ovom zaslonu postoji više mogućnosti za njihovo manipuliranje. U naslovu stupca nalazi se izbornik koji omogućuje izbor vrste upravljanja nad podacima koji su učitani. Tako je moguće filtrirati, sortirati i na druge načine upravljati učitanim podacima, a jedna od naprednijih mogućnosti je kreiranje faceta podataka (Slika 10).



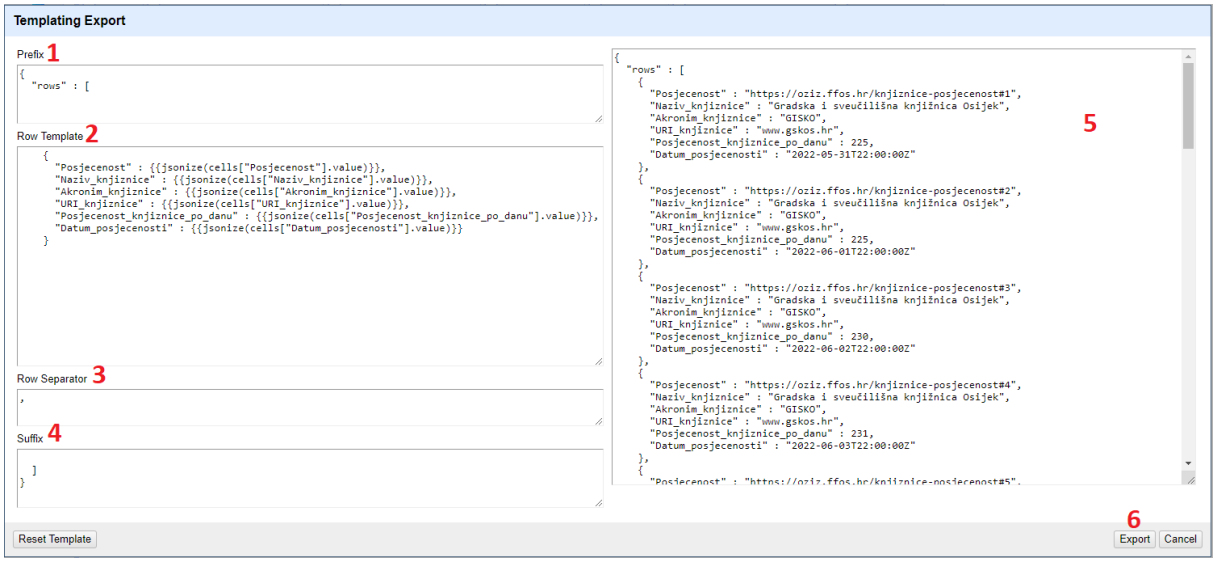
Slika 10 - Prikaz zaslona projekta, izbornik za manipulaciju podataka.

Nakon detaljne analize podataka slijedi korak primjene prethodno kreiranog predloška za konverziju tabličnih u povezane podatke. Klikom na gumb označen brojem 1 (*Export*) otvara se izbornik projekta, i vidljivom postaje opcija „*Templating*…“ koja je označena brojem 2 na Slici 11. Odabirom navedene opcije otvara se prozor u koji se u odgovarajuća polja kopiraju dijelovi kôda iz prethodno kreiranog predloška kako slijedi.



Slika 11 – Koraci unosa predloška za konverziju povezanih podataka u *OpenRefine*.

Nakon što se otvori prozor prikazan na Slici 12, dijelovi kôda predloška unose se u odgovarajuća polja. U polje prefiksa (engl. *prefix*) (1) unosi se deklaracijski kôd u kojem se, kao što je već spomenuto, definiraju imenski prostori upotrjebljenih rječnika. Znakovi „< ! - - " i "- - >" su znakovi za komentiranje dijelova kôda i u ovom slučaju u komentar su stavljeni svi rječnici koji neće biti korišteni.



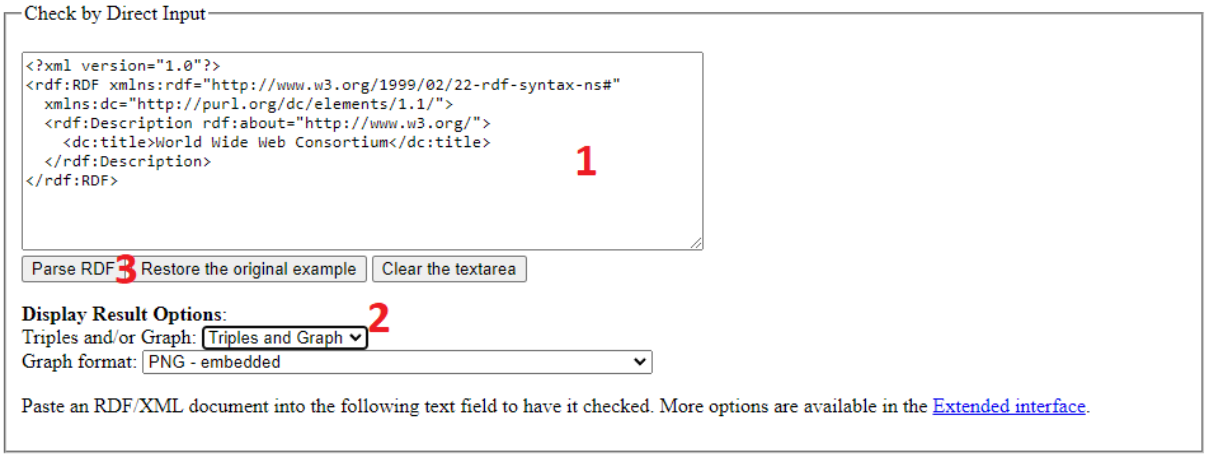
Slika 12 - Prozor za kreiranje predloška povezanih podataka.

Iz predloška za kreiranje skupa povezanih podataka se zatim u obrazac označen brojem 2 kopira kôd koji propisuje način kreiranja RDF tripleta u skupu povezanih podataka.

Nakon toga, u obrascu za razdvajanje redova (engl. *Row Seperator*), označen brojem 3, potrebno je izbrisati zarez kako bi obrazac ostao prazan. Također, bitno je u obrascu za sufiks (engl. *Suffix*) unijeti završnu oznaku kojom se zatvara RDF dokument (4). Ako su svi podaci uspješno uneseni u odgovarajuća polja, na desnoj strani, u obrascu označenim brojem 5 na Slici 13, trebao bi se prikazati rezultat.

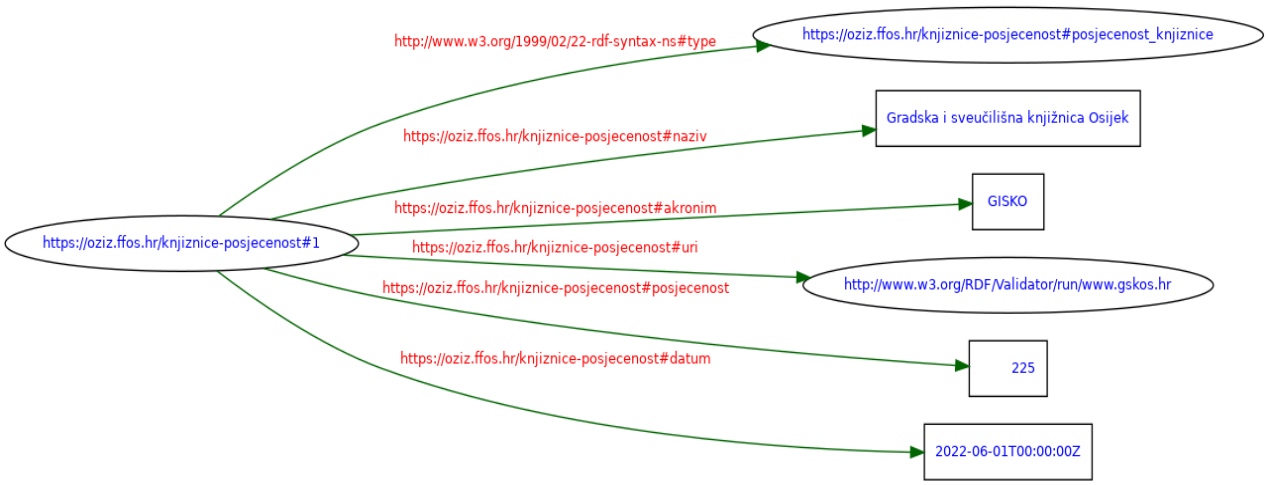
Slika 13 - Polje s rezultatom predloška u programu *OpenRefine*.

U obrascu s rezultatom prikazuju se preoblikovani tablični podaci u RDF-u. Svaki zapis iz tablice pretvoren je u semantičku izjavu koja ima svoju semantičku vrijednost. Na ovom mjestu, važno je pregledati obrazac s rezultatom, odnosno utvrditi prikazuju li se svi podaci ispravno. Do grešaka može doći ako predložak nije ispravno formiran ili u tablici nisu uneseni svi podaci (deklaracija tipa podataka, deklaracija rječnika i sl.). Nakon svih navedenih koraka, preostalo je izvesti (engl. *Export*) RDF datoteku koju smo ovim koracima formirali. Za izvoz datoteke potrebno je nakon pregleda rezultata kliknuti na gumb *Export* koji je naznačen brojem 6 na slici 12. Nakon klika na gumb, otvara se prozor u kojem se može odabrati mjesto spremanja datoteke. Pritom, važno je u polju za vrstu datoteke odabrati „sve tipove“ (engl. *All types*) i zatim nakon naziva datoteke dopisati nastavak „.rdf“ kako bi se naznačilo da završni produkt mora biti RDF datoteka. Time je postignut jedan od ciljeva ovoga rada, a to je kreiranje skupa povezanih podataka u RDF strojnočitljivom formatu. Kako bi se utvrdila ispravnost stvorene RDF datoteke, potrebno je datoteku validirati putem RDF validatora koji se nalazi na mrežnoj stranici W3 konzorcija (https://www.w3.org/RDF/Validator/).



Slika 14 - RDF validator, unos podataka.

Na Slici 14. prikazan je dio zaslona za unos podataka s početnim primjerom. U obrazac označen brojem 1 na Slici 14 unose se podaci iz RDF datoteke koji su generirani pomoću *OpenRefinea*, nakon toga u izborniku označenim brojem 2 odabire se način na koji korisnik želi da mu se rezultat validacije prikaže, u primjeru odabrana je opcija prikaza i tripleta i grafa (engl. *Triples and Graph*). I konačno, potrebno kliknuti na gumb *Parse RDF* koji je označen brojem 3. Nakon validacije, prikazuje se zaslon s rezultatom. Ako u datoteci i podacima nema pogrešaka, pojaviti će se poruka koja govori da je datoteka ispravna, a ako nije, prikazat će opis pogreške. Ako je validacija uspješna, prikazat će se RDF tripleti stvorenog skupa povezanih podataka u tabličnom obliku. Također, prikazuje se i graf koji je generiran na temelju povezanih podataka koji su uneseni. Za potrebe generiranja grafičkog prikaza RDF zapisa, samo primjera radi, u validatoru je unesen i prikazan prvi zapis.



Slika 15 - Prikaz RDF grafa s podacima unesenim u validator.

Na Slici 15 može se vidjeti na koji je način svaki podatak reprezentiran RDF tripletom. Ovom vrstom reprezentacije podataka omogućen je bolji uvid u njihovu međusobnu povezanost .

4.4. Primjena kreiranog skupa povezanih podataka

Nakon uspješno stvorenog skupa povezanih podataka, postavlja se pitanje što se s njim može učiniti. U ovom radu prikazat će se primjer grafičkog uspoređivanja podataka iz kreiranog skupa povezanih podataka u besplatnoj inačici programa *GraphDB*. *GraphDB* predstavlja grafičku bazu podataka ili RDF triplet-spremište (engl. *RDF Triplestore*) u kojem se, između ostalog, nakon uvoza datoteke mogu postavljati SPARQL upiti. Program je potrebno preuzeti s mrežne stranice programa.[47]

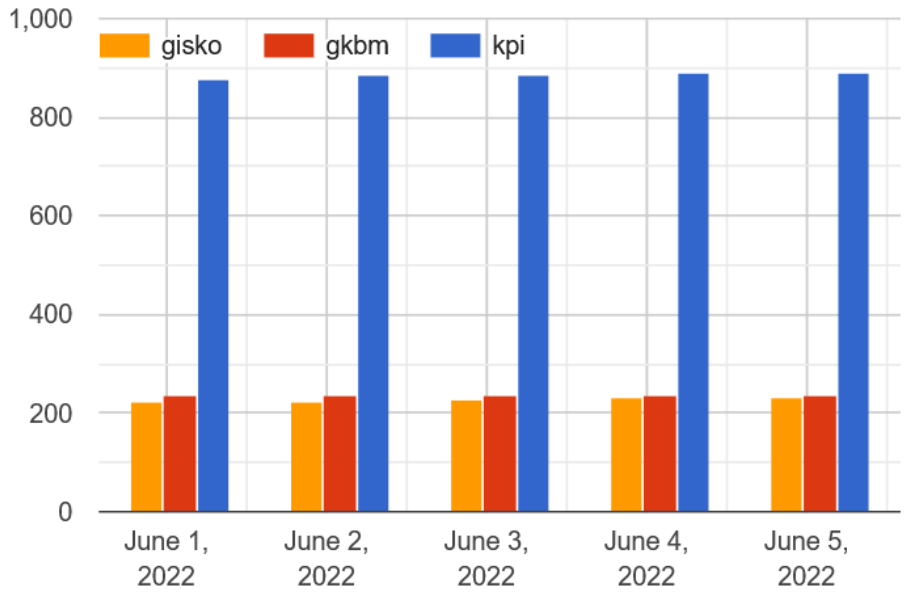
Na početnom zaslonu programa pojavit će se kratki vodič koji vodi korisnika kako da učita datoteku te na kojem mjestu je moguće postaviti SPARQL upit. Nakon što je RDF datoteka skupa povezanih podataka učitana, u programu je moguće postaviti SPARQL upit. Za potrebe demonstracije provedbe SPARQL upita nad datotekom kreiranom u ovom radu i koja predstavlja prototip skupa povezanih podataka, formiran je sljedeći upit koji na prirodnom jeziku glasi: prikaži grafički usporedbu posjećenosti mrežnih stranica Gradske i sveučilišne knjižnice u Osijeku, Gradske knjižnice Beli Manastir i Knjižnice Poljoprivrednog instituta za prvih pet dana mjeseca lipnja 2022. U SPARQL-u upit glasi:

```
PREFIX : <https://oziz.ffos.hr/knjiznice-posjecenost#>
select ?da ?gisko ?gkbm ?kpi
where
{
        ?s :naziv "Gradska i sveučilišna knjižnica Osijek" ;
           :posjecenost ?gisko ;
           :datum ?da .
        ?d :naziv "Gradska knjižnica Beli Manastir" ;
```

[47] GraphDB. URL: https://graphdb.ontotext.com/ (2022-9-28)

```
            :posjecenost ?gkbm ;
            :datum ?da .
        ?b :naziv "Knjižnica Poljoprivrednog instituta";
            :posjecenost ?kpi ;
            :datum ?da .
}
LIMIT 5
```

Navedenim upitom traži se, dakle, da se prikažu podaci o posjećenosti za tri knjižnice, a podaci koji se prikazuju su akronim knjižnice, broj posjetitelja te datum kada se mjerila posjećenost. Zadnjom linijom naredbe (LIMIT 5) ograničava se broj rezultata na 5, odnosno na prvi pet dana lipnja ili pet zapisa o posjećenosti mrežne stranice knjižnice. Provedbom navedenog upita najprije se dobivaju tablični rezultati u kojem su ispisani svi traženi podaci. U *GraphDB*-u je zatim moguće u tim tablicama filtrirati podatke. No, postoji i opcija kreiranja grafa. Graf s podacima generiranim postavljenim upitom prikazan je na Slici 16.



Slika 16 - Prikaz grafa generiranog u *GraphDB*-u.

Generirani graf vizualno prikazuje podatke koji su rezultirali provođenjem SPARQL upita te može dati drugačiji pogled na podatke i pomoći pri analizi istih.

5. Zaključak

Globalna mreža *World Wide Web* uzrok je revolucijske promjene u domeni dijeljenja podataka među ljudima. U prvom dijelu razdoblja postojanja, web je služio kao spremište podataka i informacija kojima se moglo pristupiti tako da se direktno uputi na izvor istih. Kroz godine razvoja, došlo je do revolucionarnih promjena, a za većinu je zaslužan Tim Berners-Lee koji je imao svoju viziju weba na kojem bi bila omogućena puna suradnja između ljudi i računala, a ta nova verzija weba naziva se semantički web. Ostvarivanjem navedene suradnje ljudi bi mogli puno lakše i brže doći do traženih podataka, a računala bi jasnije i kvalitetnije razumijevala podatke koje korisnici postavljaju na web. Slijedom te ideje došlo je do razvoja mnogih tehnologija semantičkog weba kao što su RDF, RDFS, ontologijski rječnici (OWL), SPARQL i mnoge druge. Kako bi ta zamisao postala ostvariva, Tim Berners-Lee je osnovao još jednu bitnu tehnologiju, a to su povezani podaci. Povezani podaci formiraju se na način da su svi podaci međusobno povezani, ljudski definiranim vezama u obliku koji je strojnočitljiv. Kreiranjem takvih skupova podataka, računalima se otvara puno mogućnosti i načina za pretraživanje, modeliranje te u konačnici prikaz rezultata podataka koji se temelje na tim povezanim skupovima.

U radu je prikazan postupak stvaranja skupa povezanih podataka, kroz korake u programu *OpenRefine*. Detaljno su opisani koraci kojima se može dobiti uvid u svaki dio postupka, od prikupljanja i modeliranja podataka u oblik koji odgovara svim standardima potrebnima za realizaciju datoteke povezanih podataka do završnog produkta u kojem se jasno vide definirane poveznice među podacima. Korišten je rječnik RDF-a, u nekim slučajevima definirali su se i tipovi podataka i veza te su se u konačnici podaci mogli prikazati u obliku RDF grafa.

Osim toga, stvorena RDF datoteka može se dalje koristiti za potrebe različitih načina prikaza odabranih podataka, njihova modeliranja i filtriranja. U radu je korišten program *GraphDB* u kojem se provode SPARQL upiti na generiranim podacima. Upitima se prikazuju željeni podaci, ovisno o zadanim kriterijima.

Svrha ovoga rada bila je prikazati cjelokupni postupak stvaranja skupa povezanih podataka putem tehnologije povezanih podataka te dokazati da i osoba koja nije u potpunosti upoznata sa svim navedenim tehnologijama može to napraviti ako se posluži ovim diplomskim radom kao vodičem. U radu se koristio primjer posjećenosti mrežnih stranica hrvatskih

knjižnica. Provedbom sličnih upita vezanih uz ispitivanje posjećenosti vlastitih mrežnih stranica, knjižnice bi mogle dobiti detaljan uvid u promet na njihovim stranicama. Djelatnici knjižnice mogli bi uočiti porast i opadanje broja korisnika koji posjećuju mrežnu stranicu te na temelju detaljne analize zaključiti koji načini poslovanja privlače korisnike, a koji ih odbijaju. Pored toga, ako knjižnice imaju pristup podacima o posjećenosti mrežnih stranica drugih knjižnica, moguće je i uspoređivati podatke. Koristeći se ovdje prezentiranom tehnologijom povezanih podataka, knjižnice i srodne ustanove mogu je primijeniti za analizu bilo kojeg drugog aspekta svog poslovanja (poput broja posuđenih a nevraćenih knjiga, dobi korisnika koji posjećuju knjižnicu i sl.) te na taj način, i u skladu s pravilnom interpretacijom dobivenih podataka, značajno podići razinu učinkovitosti vlastitih usluga.

6. Literatura

1. Berners-Lee, Tim; Bizer, Christian; Heath, Tom. Linked Data: The Story so Far. // International journal on Semantic Web and information systems, 5(2009), 1-22. URL: https://www.researchgate.net/publication/225070216_Linked_Data_The_Story_so_Far (2022-9-8)
2. Berners-Lee, Tim; R. Fielding; L. Masinter. Uniform Resource Identifier (URI) : generic syntax. 2005, str. 4-6. URL: https://greenbytes.de/tech/webdav/rfc3986.pdf (2022-9-15)
3. Bosančić, Boris. Otvoreni povezani podatci i metapodatci. // Otvorenost u znanosti i visokom obrazovanju, 2018, 187-191. URL: https://repozitorij.ffzg.unizg.hr/en/islandora/object/ffzg%3A1487/datastream/FILE0/view#page=182 (2022-9-8)
4. Britannica. „World Wide Web (WWW): History, Definition & Facts“ . URL: https://www.britannica.com/topic/World-Wide-Web (2022-9-13)
5. Cambridge Semantics. SPARQL Query and Command Clauses. URL: https://docs.cambridgesemantics.com/anzograph/v2.4/userdoc/sparql-queries.htm (2022-9-18)
6. Cardoso Jose, Sheth Amit. The Semantic Web and its Applications, 2006. URL: https://link.springer.com/chapter/10.1007/978-0-387-34685-4_1#citeas (2022-9-6)
7. Code for Science & Society. URL: https://codeforscience.org/about/ (2022-9-18)
8. Farago, F., Bosančić, B., i Badurina. 'Povezani podaci i knjižnice', Vjesnik bibliotekara Hrvatske, 56(4), B. (2013), str. 25-52. URL: https://hrcak.srce.hr/142376 (2022-9-8)
9. Gopal, Pandey. „The Semantic Web: An Introduction and Issues“ // International Journal of Engineering Research and Applications, 1(2012), str. 780 – 786. URL: https://www.academia.edu/1959581/IJERA_www_ijera_com_ (2022-9-8)
10. GraphDB. URL: https://graphdb.ontotext.com/ (2022-9-28)
11. Illinois Library. OpenRefine: About OpenRefine. URL: https://guides.library.illinois.edu/openrefine/about (2022-9-18)
12. J. Domingue, D. Fensel, J. A. Hendler, Handbook of Semantic Web Technologies. Berlin, Germany: Springer, 2011. str. 18 URL: https://pdfdrive.to/pdfs/handbook-of-semantic-web-technologies-pdf (2022-9-8)
13. Legg, Catherine. Ontologije na semantičkom webu. // Vjesnik bibliotekara Hrvatske 53,

1(2010), str. 156-157. URL: http://www.hkdrustvo.hr/datoteke/838/vbh/God.53(2022-9-5)

14. Ontotext. What is SPARQL. URL: https://www.ontotext.com/knowledgehub/fundamentals/what-is-sparql/ (2022-9-23)
15. OpenRefine. Documentation. URL: https://openrefine.org/documentation.html (2022-9-18)
16. OpenRefine. Download. URL: https://openrefine.org/download.html (2022-9-28)
17. Oriental Journal of Computer Science and Technology. Semantic Web: XML and RDF Roles, 2016. URL: https://www.computerscijournal.org/vol8no3/semantic-web-xml-and-rdf-roles/ (2022-9-6)
18. Pandey, Mrinal; Pandey, Rajiv. JSON and its use in Semantic Web. // International Journal of Computer Applications (0975-8887), 11(2017). Str. 10-16. URL: https://www.researchgate.net/profile/Mrinal-Pandey-3/publication/316250952_JSON_and_its_use_in_Semantic_Web/links/5925af4baca27295a8ebb055/JSON-and-its-use-in-Semantic-Web.pdf (2022-9-28)
19. RDF Schema 1.1, 2014-2-25. URL: https://www.w3.org/TR/rdf-schema/
20. RDFa. Linked Data in HTML. URL: https://rdfa.info/ (2022-9-8)
21. URI vs URL: Differences and When to Use Them, 2022-8-15. URL: https://www.hostinger.com/tutorials/uri-vs-url (2022-9-12)
22. Wang ,Yongming; Q. Yang, Sharon. Linked Data Technologies and What Libraries Have Accomplished So Far. // International Journal of Librarianship, Vol. 3, 1(2018), str. 10- 12 URL: https://journal.calaijol.org/index.php/ijol/article/view/62 (2022-9-18)
23. W3C. Gleaning Resource Descriptions from Dialects of Languages (GRDDL). URL: https://www.w3.org/2001/sw/wiki/GRDDL (2022-9-8)
24. W3C. JSON-LD 1.1. URL: https://www.w3.org/TR/json-ld11/#the-context (2022-9-26)
25. W3Schools.com. „HTML Introduction“. URL: https://www.w3schools.com/html/html_intro.asp (2022-9-13)

Prilog

Predložak za modeliranje povezanog skupa podataka

```
<!-- For Prefix -->

<?xml version="1.0" encoding="UTF-8"?>

<rdf:RDF xmlns="https://oziz.ffos.hr/knjiznice-posjecenost#"

<!--

xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"

xmlns:skos="http://www.w3.org/2004/02/skos/core#"

xmlns:isbd="http://iflastandards.info/ns/isbd/elements/"

xmlns:dc="http://purl.org/dc/terms/"

xmlns:foaf="http://xmlns.com/foaf/0.1/"

-->

xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">


<!-- For Row Template -->

<posjecenost_knjiznice
rdf:about="{{jsonize(cells["Posjecenost"].value).replace('"', '')}}">

      <naziv>{{jsonize(cells["Naziv_knjiznice"].value).replace('"',
'')}}</naziv>

      <akronim>{{jsonize(cells["Akronim_knjiznice"].value).replace('"',
'')}}</akronim>

      <uri
rdf:resource="{{jsonize(cells["URI_knjiznice"].value).replace('"', '')}}" />

      <posjecenost rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">

      {{jsonize(cells["Posjecenost_knjiznice_po_danu"].value).replace('"',
'')}}</posjecenost>

      <datum
rdf:datatype="http://www.w3.org/2001/XMLSchema#date">{{jsonize(cells["Datum
_posjecenosti"].value).replace('"', '')}}</datum>

</posjecenost_knjiznice>
```

```
<!-- Don't put anything in Row Separator -->


<!-- For Suffix -->

</rdf:RDF>
```