

Razmjena bibliografskih podataka: programska sučelja i formati

Cota, Marta

Undergraduate thesis / Završni rad

2020

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zadar / Sveučilište u Zadru**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:162:472822>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-08-16**



Sveučilište u Zadru
Universitas Studiorum
Jadertina | 1396 | 2002 |

Repository / Repozitorij:

[University of Zadar Institutional Repository](#)



Sveučilište u Zadru

Odjel za informacijske znanosti

Preddiplomski sveučilišni studij informacijske znanosti (jednopedmetni)

Marta Cota

**Razmjena bibliografskih podataka: programska
sučelja i formati**

Završni rad

Zadar, 2020.

Sveučilište u Zadru

Odjel za informacijske znanosti

Preddiplomski sveučilišni studij informacijske znanosti (jednopedmetni)

Razmjena bibliografskih podataka: programska sučelja i formati

Završni rad

Student/ica:

Marta Cota

Mentor/ica:

dr. sc. Krešimir Zauder

Zadar, 2020.



Izjava o akademskoj čestitosti

Ja, **Marta Cota**, ovime izjavljujem da je moj **završni** rad pod naslovom **Razmjena bibliografskih podataka: programska sučelja i formati** rezultat mogega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na izvore i radove navedene u bilješkama i popisu literature. Ni jedan dio mogega rada nije napisan na nedopušten način, odnosno nije prepisan iz necitiranih radova i ne krši bilo čija autorska prava.

Izjavljujem da ni jedan dio ovoga rada nije iskorišten u kojem drugom radu pri bilo kojoj drugoj visokoškolskoj, znanstvenoj, obrazovnoj ili inoj ustanovi.

Sadržaj mogega rada u potpunosti odgovara sadržaju obranjenoga i nakon obrane uređenoga rada.

Zadar, 14. listopada 2020.

Sadržaj

1. UVOD.....	1
2. Povijest razmjene bibliografskih zapisa među knjižnicama	2
2.1 MARC format.....	2
3. Značaj web tehnologija i standarada za razmjenu podataka	4
3.1 World Wide Web.....	4
3.2 Softver za upravljanje referencama	6
3.2.1 Softveri Zotero i Mendeley	6
3.3 Formati za razmjenu bibliografskih metapodataka za krajnjeg korisnika	8
3.3.1 Format BibTeX.....	8
4. Usluge i formati za programsku razmjenu bibliografskih metapodataka na webu	9
4.1 Odabrane usluge	9
4.1.1. Crossref.....	9
4.1.2 Open Library.....	10
4.2 Formati	11
4.2.1 XML.....	11
4.2.2 JSON	13
4.3 Programska razmjena podataka	14
4.3.1 API sučelja.....	14
5. Praktičan projekt: programski dohvat bibliografskih metapodataka putem programskog jezika Python.....	15
6. Zaključak.....	18
7. Popis literature.....	20
8. Popis ispisa.....	23
9. Prilozi.....	24

Sažetak

U ovome radu riječ je o tome što bibliografska razmjena podataka podrazumijeva, od kakve je važnosti za knjižnice i krajnje korisnike te kontekst u kojem se pojavljuje. Govori se također o značaju MARC-a kao prvog formata koji se koristio u svrhe razmjene bibliografskih podataka, ali ističe se i koje su njegove mane. Cilj rada je prikazati različite suvremene usluge za razmjenu bibliografskih podataka gdje je glavni fokus stavljen na različite vrste korisnika radije no na isključivo stručnu upotrebu tj. knjižnice. Rad se dotiče usluga za razmjenu bibliografskih podataka Crossref i OpenLibrary. Pritom je naglasak stavljen na API sučelja koje ove usluge pružaju, ali i formate (XML, JSON) u kojima se podaci dostavljaju krajnjem korisniku. Uz navedeno, obrađuje se i softver za upravljanje referencama s naglaskom na Zotero i Mendeley te format za razmjenu podataka među ovakvim aplikacijama (BibTeX) kao drugi vid razmjene bibliografskih metapodataka među krajnjim korisnicima. Rad je zamišljen kao stručan pregled navedenih tema. U praktičnom projektu prikazat će se se programski dohvat bibliografskih metapodataka putem Pythona.

Ključne riječi: razmjena podataka, bibliografski podaci, metapodaci, XML, JSON, BibTeX, Crossref, Open Library, Zotero, Mendeley.

1.UVOD

Osnova organizacije i upravljanja zbirkama su podaci o predmetima koje te zbirke sadrže. Na primjer podaci o knjigama i ostalim publikacijama u knjižničnim zbirkama. Ovakve podatke često nazivamo metapodacima. U digitalnom umreženom svijetu, metapodaci su osnova svakog informacijskog sustava i dolaze u raznim oblicima. Ljudi svaki dan postavljaju slike na društvene mreže kao što su Facebook, Instagram i Twitter, slušaju glazbu putem glazbenih platformi kao što su Spotify, Deezer, Youtube, komuniciraju putem email-a, chatova i društvenih mreža itd. Zbog metapodataka moguća je organizacija i upravljanje zbog čega se smatraju podatkovnom jezgrom svih sustava i aplikacija koje koristimo svaki dan. Metapodaci su ključ za uspješno funkcioniranje sustava jer sadrže i skupljaju informacije koje su bitne za pružanje najbolje moguće usluge krajnjem korisniku. „Metapodaci - podaci koji pomažu u identifikaciji, opisu i lociranju nekih podataka, napose umreženih elektroničkih izvora.“¹

Budući da su metapodaci važni za ostale informacijske sustave, smatraju se i osnovnim dijelom sustava knjižnica. Tada govorimo o bibliografskim metapodacima. Važno je za napomenuti da pojava i upotreba metapodataka u knjižničarstvu nije nešto novo. Naime, knjižničarstvo se bavi tim idejama organizirano već barem 150 godina a početci sežu još od antičkog svijeta. No, da ne idemo previše daleko u prošlost, za potrebe ovog rada dovoljno je opisati kako je izgledala razmjena podataka prije pojave weba. Istaknut će se kakvi formati su bili razvijeni prije pojave novih informacijskih tehnologija te koja je njihova važnost za knjižnice. Jedan od takvih formata je i MARC format koji se smatra prvim specijaliziranim formatom za bibliografsku razmjenu metapodataka.

MARC je popularan format koji koriste mnoge knjižnice zato što se pokazao kao dostatna zamjena za dotadašnji način katalogiziranja uz pomoć kataložnih listića. Također, razvojem UNIMARC-a po prvi put su ustanovljena pravila za katalogizaciju što je dovelo do međunarodne razmjene kataložnih zapisa. No, MARC se danas smatra zastarjelim standardom jer je on dizajniran na način da sadrži sintaksu i pravila koja su ljudskom korisniku zahtjevna za učenje i korištenje. Samim time, korisnici su ograničeni iz razloga što gube razne mogućnosti i prednosti koje im programsko dohvaćanje i razmjena podataka pruža.

¹ Leksikografski zavod Miroslav Krleža. URL: <https://www.enciklopedija.hr/Natuknica.aspx?ID=40384> (2020-9-23)

Međutim, razvojem weba i Interneta razvijaju se i novi aktualniji formati i softveri koji su prilagođeniji ljudskom korisniku za razliku od prethodno navedenog MARC-a. Oni bi mogli i zamijeniti korištenje MARC standarda iako nisu nužno stvoreni za razmjenu bibliografskih metapodataka ali su se pokazali kao bolje rješenje. Samim time razvijaju se i razne online usluge koje koriste bibliografske metapodatke za razvoj proizvoda namijenjenih krajnjim korisnicima. Primjer takvih formata su XML i JSON, a primjer usluga su Crosreff i Open Library koji će biti detaljnije opisani u radu. Osim ovih formata i usluga razvili su se i softveri za upravljanje referencama poput Zotero i Mendeley. Potreba za razvojem ovih softvera pojavila se iz razloga što knjižničari nisu jedini korisnici bibliografskih metapodataka. Svi oni koji se bave znanstvenom komunikacijom (znanstvenici, studenti, novinari itd.) su također potencijalni korisnici bibliografskih metapodataka. Rad će također prikazati i praktični projekt koji će se sastojati od dohvaćanja bibliografskih metapodataka za neki DOI identifikator.

2. Povijest razmjene bibliografskih zapisa među knjižnicama

Zahvaljujući računalnoj tehnologiji knjižnice su se uskoro upoznale s novim načinima rada i međusobnom razmjenom bibliografskih podataka. Razvoj računalne tehnologije bio je osnovni preduvjet za organizaciju upravljanja podacima, te su ga knjižnice prihvatile 60-ih godina razvojem MARC standarda.

2.1 MARC format

Kako se navodi na samoj stranici Kongresne Knjižnice, MARC se počeo razvijati sredinom 1960-ih godina od strane nacionalne knjižnice SAD-a (The Library of Congress)² te je prvi format koji je napravljen da bude specijalizirani format za razmjenu među knjižnicama. Razvio se iz razloga što se pojavila potreba za dijeljenjem bibliografskih podataka preko formata koji će imati mogućnost da budu strojno čitljivi. Upravo to i jest akronim za MARC (Machine-Readable Cataloging record)³. Strojno čitljivo se u ovome slučaju odnosi na formate koji služe kako bi računala mogla razumjeti i procesuirati podatke prema određenim pravilima, atributima i upotrebom određene sintakse.

Katharine D. Morton u članku pod nazivom *The MARC Formats: An Overview* kaže kako se prva verzija MARC-a zvala USMARC, a nakon ove verzija razvijena je nova pod nazivom MARC 21 Format for Bibliographic Data. ⁴Ovaj format je prvotno zamišljen kao

² The Library of Congress. URL: <https://www.loc.gov/marc/> (2020-10-12)

³ Ibid.

⁴ Morton, Katharine. The Marc MARC Formats: An Overview. // The American Archivist (1986) 49 (1), 21-30.

format kojim će se opisivati sve vrste bibliografskih podataka zajedno, tj. na isti način. Kasnije je ustanovljeno da zbog vremenskog i kadrovskog ograničenja nije moguće identificirati i analizirati elemente podataka za cijeli skup građe na isti način, pa je stoga odlučeno da će MARC služiti kao format za svaku vrstu materijala zasebno, počevši od knjižne građe. MARC format koji služi za opisivanje knjižne građe prvi put se pojavio 1968. godine pod naslovom *Subscriber's Guide to the MARC Distribution Service*⁵. MARC koji se koristi za opis mapa, filmova, serijskih publikacija, rukopisa i glazbe objavljen je 1976. godine⁶. Najkorištenija verzija MARC-a razvijena je od strane IFLA-e 1977. godine i zove se UNIMARC koji je poseban po tome što su u njemu definirana pravila katalogiziranja koja su omogućila identičnost kataložnih zapisa.⁷

U knjizi autorice Betty Furrrie pod naslovom *Understanding MARC bibliographic: machine-readable cataloging* piše kako je jedno od najvažnijih mogućnosti koje MARC pruža ispravna katalogizacija knjižnične građe iz razloga što je dostupan neograničen broj polja s neograničenom dužinom polja⁸. To je važno jer se svakoj katalogizaciji jedinice knjižnične građe treba pristupiti na poseban način. Jedinice knjižnične građe međusobno se razlikuju po dužini naslova, materijalnom opisu, a neke su dio nakladničke cjeline zbog čega im je potrebno dodatno polje za opis i tako dalje. Ono što se dobilo razvojem takvog formata jest razmjena bibliografskih metapodataka između većeg broja knjižnica ali i međusobno korištenje tih metapodataka kako bi se postigao zajednički cilj; brža, jednostavnija i efikasnija razmjena metapodataka. Osim toga, ovime se postiže jedinstvenost i identičnost kataložnih zapisa što je uvelike olakšalo posao knjižničarima. Važnost i doprinos MARC formata za knjižnice je od velikog značaja jer je njegov nastanak i razvoj poslužio kao prva stepenica koja je dovela do međusobne knjižnične razmjene podataka. Također, od svog nastanka pa sve do danas je i dalje korišten u knjižničarskoj struci.

Usprkos gore navedenom, postoje određene mane ovog formata koje su jako važne za istaknuti. MARC je i dalje format koji ima ograničeno djelovanje i nije dovoljno moderniziran da bi se njegov puni potencijal mogao iskoristiti u knjižničnoj struci kakvu poznajemo danas. MARC nije namijenjen za korištenje u programskim jezicima i sličnim alatima za obradu

⁵ Ibid.str 23.

⁶ Ibid.

⁷ Leksikografski zavod Miroslav Krleža. URL: <https://www.enciklopedija.hr/natuknica.aspx?id=63215> [2020-9-23]

⁸ Furrrie, Betty. Why is a MARC record necessity. // *Understanding MARC bibliographic: machine readable cataloging* / Betty Furrrie. Washington, DC: Follett Software Co., 1988. 3-4.str

podataka. Ekstrahirati podatke iz MARC-a kao i služenje njima nije jednostavno te zahtjeva veliku razinu poznavanja ovog formata kao i vremena. Ovaj format nije jednostavan za razumijevanje i čitanje kako ljudima tako i računalima stoga nije idealno rješenje za razmjenu podataka. Kao jednu od najvažnijih negativnih strana koje autor Roy Tennant izdvaja u članku „*MARC must die*“ vezano za korištenje ovog formata je to što polja nisu naznačena na jednostavan način već se koriste brojčane oznake što ga čini nerazumljivim onima koji ne poznaju njegovu kompliciranu sintaksu.⁹ Također, jedna od mana MARC-a je to što nije namijenjen za korištenje na webu iz razloga što je nastao prije weba kao i to što kasnije nije prilagođen korištenju na novim informacijskim tehnologijama. Iz prethodno navedenih razloga razvili su se novi softveri, formati i web tehnologije za razmjenu podataka prilagođeni digitalnom dobu o čemu će se više pričati u nastavku.

3. Značaj web tehnologija i standarada za razmjenu podataka

3.1 World Wide Web

WWW je nastao iz potrebe organiziranja i jednostavnije razmjene podataka i dokumenata. „WWW (akronim od engl. World Wide Web: svjetska mreža), najčešće korišten internetski servis, koji korisnicima omogućava pregledavanje mnoštva digitalnih dokumenata danih na raspolaganje preko umreženih računala diljem svijeta; vrlo često samo web ili mreža.“¹⁰ WWW je najpoznatiji i najkorišteniji dio Interneta čiji je cilj da bude sustav međusobno povezanih dokumenata. Najvažniji dijelovi na kojima se zasniva djelovanje WWW-a su hiperveze (linkovi) i adrese dokumenata (URL).¹¹ Linkovi nas vode do dokumenta tj. web stranice a URL adresa se sastoji od imena protokola (http://) i adrese servera na koji se povezujemo.¹² Za početak nastanka WWW-a se veže godina 1989. a njegovi začetnici su Tim Berners Lee i Robert Cailliau.¹³ Prva web stranica bila je dostupna 1990. godine, a 1991. WWW postao je dostupan širem broju korisnika, dok je 1993. godine web postao javno dostupan za korištenje

⁹ Tennant, Roy. MARC must die. URL: <http://soiscompsfall2007.pbworks.com/f/marc+must+die.pdf> (2020-10-11)

¹⁰ Leksikografski zavod Miroslav Krleža. URL: <https://www.enciklopedija.hr/Natuknica.aspx?ID=66413> (2020-10-11)

¹¹ Ledinek, Sanja. Internet i World Wide Web dva su različita pojma – ne miješajte ih, 29.04.2014. URL: <https://www.racunalo.com/internet-i-world-wide-web-dva-su-razlicita-pojma-ne-mijesajte-ih/>. [2020-10-11]

¹² Ibid.

¹³ Leksikografski zavod Miroslav Krleža. URL: <https://www.enciklopedija.hr/Natuknica.aspx?ID=66413> (2020-10-11)

svima.¹⁴ W3C je važna organizacija koju je osnovao Tim Berners Lee 1994. godine čija je osnovna zadaća postavljanje novih standarda u izradi web stranice.¹⁵

Pojavom i razvojem weba dogodio se niz promjena u mnogim dijelovima ljudskog života, uključujući slanje, primanje, pronalazak i služenje informacijama. Informacije su postale lakše dostupne te se razvijaju mnogi elektronički mediji i usluge namijenjene širenju informacija.

Ovakve promjene dogodile su se i u knjižnicama, pa su i knjižničari bili primorani prilagoditi svoje usluge i razviti suvremenije sustave i usluge. Da bi knjižnice uspješno izvršile svoju glavnu zadaću a to je da „omoguće jednaku dostupnost svih izvora kulturnih i znanstvenih informacija i svih vrsta građe te osiguravanje mogućnosti za cjeloživotno učenje kao uvjete za neovisno odlučivanje i kulturni razvitak pojedinaca i društvenih skupina“¹⁶, moraju svoj način poslovanja, usluge i načine na koje pružaju informacije prilagoditi i mijenjati sukladno novom okruženju suvremenog informacijskog društva. Mijenjaju se načini razmjene podataka, knjižnice se počinju umrežavati te se razvijaju online usluge knjižnica, čime postaju dostupnije korisnicima. Počeli su se razvijati online katalozi te je tako knjižnicama omogućeno da na jednom mjestu okupe svu knjižničnu građu i podatke o njoj. Na taj način je rad knjižnice postao bliži korisnicima i uključio ih u samo djelovanje knjižnice.

Najvažnija uloga koju je web imao što se tiče knjižnica je razvoj usluga i formata koje će omogućiti suvremen pristup strukturiranim bibliografskim podacima i interoperabilnost među njima. Kako je ranije navedeno dosadašnji softveri i formati za razmjenu bibliografskih metapodataka nisu bili standardizirani i namijenjeni globalnoj razmjeni podataka. Bili su specijalizirani za razmjenu određenih metapodataka i napravljeni su kao ciljani formati koji mogu razmjenjivati bibliografske metapodatke među korisnicima, a ne za potrebe stručnog korištenja. “Sukladno razvojem različitih informacijskih sustava, svaki novi sustav je razvio svoj informacijski sistem i bazu podataka. Pošto su arhitekture sustava, operativni sustavi i baze podataka bile različite za svaki sistem nisu bili u mogućnosti dijeliti informacije i međusobno komunicirati.”¹⁷ Najvažnija zadaća formata i softvera koji služe za razmjenu podataka je da oni

¹⁴ Stopić, Dražen. World Wide Web proslavio je 30. rođendan., 14.3.2019. URL: <https://www.xn--ina-usluge-ukb.hr/world-wide-web-proslavio-je-30-rodendan/>. [11-10-2020]

¹⁵ Ibid.

¹⁶ Kvesić Grašić, Tina. Pravila Knjižnice grada Zagreba, 22.04.2020. URL: <http://www.kgz.hr/hr/informacije/upisi-i-posudba-196/pravila-knjiznica-grada-zagreba-40217/40217>. [2020-10-11]

¹⁷ Zhang, Cui-xiao, et al. Data Exchange based on Web Services. // International Journal of Computer Science and Network Security 6, 5A (2006.). Str. 210.

budu razumljivi za čitanje i korištenje kako računalima tako i osobama, tj. krajnjim korisnicima. U nastavku se detaljnije opisuje što su to softveri za upravljanje referencama i pruža se uvid u 2 primjera takvih softvera koji su prilagođeni radu na webu - Zotero i Mendeley.

3.2 Softver za upravljanje referencama

Knjižničari nisu jedini korisnici bibliografskih metapodataka. Korisnici koji često barataju metapodacima su, na primjer i znanstvenici koji imaju potrebu prikupljanja velike količine znanstvenih članaka i navođenja istih u znanstvenih publikacijama u čemu im razmjena bibliografskih podataka uvelike pomaže. Za potrebe ovih korisnika nastaje nova vrsta softvera: softver za upravljanje referencama usmjeren na znanstvene radove. Dotadašnji formati poput MARC-a su im više problem nego korist iz razloga što sa svojom kompleksnom sintaksom nisu lako uporabljivi za obradu podataka. Znanstvenicima i ostalim skupinama korisnika kojima su potrebni metapodaci je otežano njihovo korištenje i dijeljenje te je bio nužan razvoj suvremenijih formata i softvera koji će im biti razumljivi i čitljivi te iz kojih će znati dohvatiti potrebne metapodatke. Primjeri ovakvih softvera su Zotero, Mendeley, EndNote i tako dalje. S obzirom da su Zotero i Mendeley slobodno dostupni softveri, slijedi kratak opis ova dva rješenja. Također, primjer ovakvog formata za razmjenu je BibTex, koji će biti detaljnije opisan kasnije u tekstu.

3.2.1 Softveri Zotero i Mendeley

U knjizi pod naslovom *Zotero: A Guide for Librarians, Researchers, and Educators* autor Jason Puckett piše da je Zotero nastao 2006. godine od strane istraživača na George Mason Sveučilišnom Centru za Povijest i Nove Medije.¹⁸ Stvoren je zato što su istraživači na ovom Sveučilištu bili frustrirani s ograničenošću koju su im pružali komercijalni softveri za upravljanje referencama koje je bilo teško za koristiti zbog velikog broja kataloga i baza podataka¹⁹. Trebalo im je jednostavnije rješenje koje će na organiziran i smislen način poslužiti kao softver za upravljanje referencama.

Mendeley je nastao po uzoru na popularnu glazbenu uslugu Last.fm na kojoj su korisnici međusobno dijelili podatke o svojim glazbenim preferencijama te na taj način dolazili do nove

¹⁸ Puckett, Jason. About Zotero. // *A Guide for Librarians, Researchers, and Educators* / Jason Puckett. Chicago, Illinois: Association of College and Research Libraries, A division of the American Library Association, 2011. 1-19 str.

¹⁹ Ibid.

glazbe i glazbenika.²⁰ Po ovome principu djeluje i Mendeley. Ovaj program sakuplja anonimne podatke korisnika o tome što čitaju i kako organiziraju sačuvane radove i dokumente te takve podatke koristi kako bi stvorio interdisciplinarnu bazu podataka s otvorenim pristupom.²¹ Ova baza podataka sastoji se od referenci, preporuka i alata namijenjenih za suradnju koji su se razvili zahvaljujući prikupljanju podataka o učestalosti posjećenosti članaka.²²

Zotero i Mendeley su softveri koji su po svojim mogućnostima i značajkama veoma slični, stoga će se u radu opisivati zajedno. Ovi softveri služe za organizaciju vlastite baze bibliografskih metapodataka te se mogu koristiti kao umetak za preglednik koji omogućuje lako pobiranje bibliografskih metapodataka. Autor Jason Puckett u knjizi po imenu *A Guide for Librarians, Researchers, and Educators* kaže kako je zadaća Zotera da omogući korisnicima spremanje referenci iz knjižničnih kataloga, baza podataka, digitalnih repozitorija i ostalih web stranica jednim klikom na vlastito računalo. Podaci su organizirano pohranjeni te kasnije lako pretraživi i dostupni. Glavna odlika ovog programa je da na jako jednostavan način omogućuju spremanje referenci u "knjižnicu" programa te ih pretvaraju u bibliografije.²³ Namijenjeni su znanstvenicima i istraživačima za organizaciju istraživanja, suradnju s drugima i pretraživanje najnovijih istraživanja.²⁴ Ukratko, omogućuju automatsko generiranje bibliografija, laku suradnju s drugim istraživačima online, laku pohranu radova drugih istraživača, pronalazak relevantnih radova na temelju onoga što čitamo, pristup vlastitim radovima bilo gdje online²⁵.

Sastoje se od mnogih zanimljivih značajki kao što su automatsko dohvaćanje informacija o citiranju s web stranica, spremanje dokumenata u raznim formatima, spremanje fotografija, linkova kao i cijelih web stranica uključujući i spremljena pretraživanja i tagove. Zotero također dolazi s mnogim formatima citiranja koji u sebi imaju integrirane stilove citiranja. Ukratko, ovi softveri su osobna, lako dostupna online i offline digitalna knjižnica koja je prilagođena individualnim potrebama korisnika. Također, besplatni su programi namijenjeni svima neovisno o iskustvu i čije je korištenje lako za savladati.

²⁰ Reisswig, Jennifer. Mendeley. // *Journal of the Medical Library Association*. v.98(2); 2010. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2859264/#>. [2020-10-12]

²¹ Ibid.

²² Ibid.

²³ Puckett, Jason. *About Zotero*. // *A Guide for Librarians, Researchers, and Educators* / Jason Puckett. Chicago, Illinois: Association of College and Research Libraries, A division of the American Library Association, 2011. 1-19 str.

²⁴ Elsevier. URL: <https://www.elsevier.com/solutions/mendeley>. [2020-10-12]

²⁵ Puckett, Jason. *About Zotero*. // *A Guide for Librarians, Researchers, and Educators* / Jason Puckett. Chicago, Illinois: Association of College and Research Libraries, A division of the American Library Association, 2011. 1-19 str.

3.3 Formati za razmjenu bibliografskih metapodataka za krajnjeg korisnika

I krajnji korisnici danas imaju potrebu za bibliografskim metapodacima, ali MARC nije pogodan format za to. Kako bi razmjena podataka za upravljanje referencama bila moguća i uspješna potrebni su i posebni formati koji su razvijeni za te svrhe. Formati koji služe za razmjenu podataka sastoje se od određene sintakse i modela. To se odnosi na određeni vokabular i izraze koji se koriste te ih je nužno poznavati za korištenje programa. Računala će razumjeti naredbu koju tražimo od njih isključivo ako tu naredbu napišemo njihovim jezikom, tj. upotrebom točno određene sintakse. Jedan od takvih formata je BibTeX.

3.3.1 Format BibTeX

BibTeX format se koristi za potrebe razmjene podataka za upravljanje referencama te je vrlo jednostavan za programski rad na njemu. Također, ljudima je razumljiv i sadržajno nije preobuhvatan, nastao je za krajnje korisnike koji se bave znanstvenom komunikacijom i kojima je potrebna znanstvena literatura. BibTeX format omogućuje razmjenu bibliografskih metapodataka među korisnicima ali nije namijenjen za profesionalni rad ustanova. Ono najvažnije vezano za BibTeX je da je to format kojega će softveri poput Zotera i Mendeley-a koristiti kao formatirani navod u bazi podataka, a ne kao običan tekst. Softver koji je predviđen za to će iz njega pročitati podatke. Na idućem primjeru možemo vidjeti prikaz i objašnjenje zapisa u BibTeX formatu.

Ispis 1: Primjer BibTeX zapisa

```
@article{thomale2010interpreting,  
  title={Interpreting MARC: where's the bibliographic data?},  
  author={Thomale, Jason},  
  journal={Code4Lib Journal},  
  number={11},  
  year={2010}  
}
```

Svaki zapis započinje s @ nakon čega slijedi vrsta rada koja je u ovom slučaju article. Nakon vrste rada u vitičastim zagradama {} zapisuje se ključ, odnosno šifra članka. Zatim se u idući red zapisuju polja od kojih se sastoji rad. Polja su uvijek zapisana na način da prvo ide naziv polja, zatim =, kao znak koji nas upućuje na vrijednost polja koja je zapisana u {}. Na ovome primjeru navedena su polja *title* koja opisuje vrijednost naslova članka, *author* koji opisuje vrijednost za autora članka, *journal* opisuje naslov časopisa, *number* za broj članka i

year kako bi se navela godina izdanja članka. Također, važno je da su pri navođenju više polja oni na kraju odvojeni znakom zarez (,) i da čitavi zapis zatvorimo sa znakom za zatvarajuću vitičastu zagradu (}).

Iako je BibTex format koji je za svoju svrhu razmjene bibliografskih metapodataka u akademske svrhe odličan, i dalje ima neke mane. Jedna od tih mana je što je on specijalizirani format za razmjenu bibliografskih metapodataka među točno određenim softverom, odnosno da ima ograničeno djelovanje. Također, nije namijenjen za rad u Pythonu i za web jer je nastao prije weba.

4. Usluge i formati za programsku razmjenu bibliografskih metapodataka na webu

Vrijedi ponoviti da su formati o kojima se do sada pričalo, poput MARC-a i BibTeXa, razvijeni prije pojave weba. Web je naravno promijenio i uniformirao *online* razmjenu bilo kakvih strukturiranih podataka, ne samo bibliografskih. Najvažniji od ovih formata su XML i JSON. Uz nove formate za podatke razvile su se i nove online usluge kao što su Crossref i Open Library. U nastavku slijedi opis ovih usluga i formata.

4.1 Odabrane usluge

4.1.1. Crossref

Crossref je neprofitna organizacija osnovana 2000. godine s pravom na članstvo čija je zadaća cirkuliranje i razmjena akademskih članaka, literature, radova i sl. Svoju misiju ispunjavaju tako što označavaju i dijele metapodatke stvaranjem usluga i alata namijenjenih upravo tome te vode otvorenu infrastrukturu.²⁶

Crossref je osnovan od strane grupe izdavača kojima je trebala učinkovita metoda za povezivanje svojih članaka iz časopisa i to su ostvarili koristeći DOI-e na način da bi im služili za povezivanje referenci između članaka te tako olakšali korisniku da pronađe citirane predmete.²⁷ Citirano prema mrežnoj stranici DOI, on je skraćenica za Digital Object Identifier (digitalni identifikator objekta). DOI je u principu link koji nas vodi do važeće URL adrese izvora. Ako se npr. URL adresa promijeni, DOI će nam i dalje moći pružiti važeću i aktualnu

²⁶ Crossref. URL: <https://www.crossref.org/education/why-crossref/> [2020-10-12]

²⁷ Collins, Susan. Introducing Crossref, the basics, 2018. URL: <https://jasolutions.com.co/crossref/introducing-crossref-basics/>. [2020-10-12]

adresu i lokaciju dokumenta. ²⁸Nadalje, principi po kojima Crossref djeluje su podijeljeni na 6 načela: 1) Crossref je globalna zajednica članova sa sadržajem iz mnogobrojnih disciplina, u mnogim formatima s raznim poslovnim modelima, 2) svaki član sudjeluje u stvaranju odbora koji zastupa mišljenja i želje svakog od članova, 3) suradnja je ključ svega čime se bave, 4) stvaranje otvorenih alata i API-a kako bi se omogućila i poboljšala razmjena metapodataka, 5) otvorenost i transparentnost su glavni principi po kojima se djeluje, 6) nužan je stalni napredak i razvoj. ²⁹

Prema Susan Collins koja je zaposlenica Crossrefa, prioritet njihovih članova je mogućnost da sadržaj koji stvaraju bude pronađen i vidljiv onima koji pretražuju. Kada član registrira sadržaj na Crossref, automatski predaje i bibliografske metapodatke o tom sadržaju kao što su naslov, suradnici, informacije o objavljivanju, online lokacija i DOI. Ovakvi metapodaci su javno dostupni koristeći se raznim alatima za pretraživanje. Također, neki od alata koje Crossref nudi su Cited by koji pruža uvid u to tko je sve i gdje citirao rad koji korisnik čita u tom trenutku te Similarity Check uz pomoć čega se otkriva plagijarizam na način da se uspoređuje sličnost predanih rukopisa. ³⁰

4.1.2 Open Library

Kako je navedeno na web stranici Open Library, ova usluga služi kao otvoreni knjižnični katalog koji se može uređivati od strane korisnika te služi kao web stranica za svaku knjigu ikada objavljenu. Open Library je neprofitna inicijativa Internet Archive-a te izrađuje digitalnu knjižnicu kako internetskih stranica tako i drugih kulturalnih artefakata u digitalnom obliku. Krajnji cilj koji Open Library ima je omogućavanje pristupa svakom tiskanom djelu bilo kome na svijetu njihovom digitalizacijom. Iako je ovo kompleksan zadatak za njegovo ostvarenje nužna je međusobna suradnja između informacijskih i tehnoloških stručnjaka, knjižnica i autora. ³¹

Ono što je zanimljivo kod korištenja ove usluge je da svaka osoba može pomoći u širenju pristupa znanju, bez obzira radi li se o knjižničaru ili običnom čitatelju i to dodavajući nove informacije, ispravljajući pogreške, novčanim donacijama, volontirajući i stvarajući zapise o knjigama. Nadalje, registracija na stranicu je besplatna te svatko može biti član. Ono

²⁸ DOI. URL: <https://www.doi.org/>. [2020-10-12]

²⁹ Crossref. URL: <https://www.crossref.org/truths/>. [2020-10-12]

³⁰ Collins, Susan. Introducing Crossref, the basics, 2018. URL: <https://jasolutions.com.co/crossref/introducing-crossref-basics/>. [2020-10-12]

³¹ Open Library. URL: <https://openlibrary.org/>. [2020-10-12]

što se dobije registriranjem na stranicu je mogućnost online posuđivanja knjiga koje se mogu posuditi minimalno na 2 sata a maksimalno na 14 dana. Također, knjige su dostupne i u audio verziji, a stranice mogu vizualno izgledati kao da se listaju. Postoji klasično pretraživanje po autoru, naslovu, dijelu teksta, temi te listi. Liste su kreirane od strane korisnika i temelje se na njihovom izboru. Također, postoji i napredno pretraživanje po naslovu, autoru, ISBN-u, temi, mjestu, osobi i nakladniku. Što se tiče pretraživanja po temama, one su sveobuhvatne i količinski ih je puno (umjetnost, fantazija, biografije, znanost, recepti, ljubavni romani, povijesni romani, dječji romani, jezik, religija, self-help literatura, glazba itd.) Osim ovog pretraživanja postoji još i nasumično pretraživanje knjiga. Ono što je bitno za učenike je da postoji knjižnica namijenjena isključivo njima s potrebnom im literaturom. Ovdje se literatura može pretraživati prema predškolskom, vrtićkom uzrastu i razredima kao i prema razini čitanja.

4.2 Formati

4.2.1 XML

XML odnosno Extensible Markup Language (proširivi jezik za označivanje) je jednostavan, fleksibilan format nastao iz SGML (Standard Generalized Markup Language).³² XML je suvremen format koji je predviđen za strukturiranje, pohranu i razmjenu podataka, te je dizajniran da bude i ljudski i strojno čitljiv. Prvotno je nastao kako bi poslužio kao rješenje za probleme s opisom podataka u elektroničkom izdavaštvu, kasnije se ustanovilo da XML ima važnu ulogu u razmjeni različitih podataka kako na webu tako i na drugim servisima.³³

Kao što je ranije spomenuto, XML je nastao kao dodatak SGML-u. Kako se navodi u knjizi *XML in a Nutshell* autora Elliothe Rusty Harold i W. Scott Means, iako je SGML semantički i strukturalni označiteljski jezik za tekstualne dokumente, te je bio jako korišten za potrebe američke vojske i vlade kao i u mnoge druge svrhe, i dalje je ljudima jako težak za jasnijom i više dosljednom strukturom. Stoga su 1996. godine Jon Bosak, Tim Bray, James Clark, C.M. Sperberg-McQueen i ostali započeli s razvojem jednostavnije verzije SGML-a koja će i dalje sadržavati ono što je kod ovog označiteljskog jezika kvalitetno a odbacit će se one značajke koje su dokazano višak zbog svoje kompliciranosti, nečitljivosti i koje se u prethodnih 20 godina korištenja SGML-a nisu pokazale kao korisne. Prva verzija nazvana XML 1.0, kreirana je 1998. godine i pokazala se uspješnom. Nakon ove verzije nastao je standard pod nazivom *Namespaces in XML* kojim se omogućilo da npr. web stranica o knjigama koja sadrži

³² W3C. URL: <https://www.w3.org/XML/>. [2020-10-12]

³³ Ibid.

dva elementa s istim nazivom a različitim vrijednostima se mogu pisati u isti dokument bez zabune na što se koji element odnosi. Nakon njega se razvio Extensible Stylesheet Language (XSL) koji je omogućio da se dokumenti zapisani XML formatom mogu prikazati na webu. Uskoro se XSL podijelio na XSL Transformations (XSLT) i XSL Formatting Objects (XSL-FO). XSLT služi za pretvorbu jednog XML dokumenta u drugi, dok XSL-FO služi kao dodatak za opisivanje izgleda ispisane i web stranice.³⁴

Bitno obilježje XML-a je da njegove oznake nisu unaprijed određene te ih korisnik može sam smisliti, mijenjati i dijeliti. XML je kao i HTML označiteljski jezik ali on ne služi kao format za prikazivanje podataka, već za njihovo dohvaćanje i parsiranje. Ovaj format se može pokrenuti na bilo kojem operativnom sustavu što ga čini jako praktičnim i rasprostranjenim. Danas je XML, uz JSON jedan od najčešće korištenih formata za razmjenu podataka, što je i sasvim opravdano jer korisnicima pružaju razmjenu podataka bez obzira na njihovu vrstu, jednostavni su za učenje i baratanje, iskoristivost povučenih podataka je maksimalna te su interoperabilni. Ovi formati prilagođeni su za programsku iskoristivost u programskim jezicima poput JavaScript i Python. U nastavku slijedi prikaz i objašnjenje XML zapisa.

Prikaz 2: Primjer XML zapisa

```
<film>
  <naziv_filma>
  The Truman Show
  </naziv_filma>

  <godina_filma>
  1998
  </godina_filma>

  <glavni_glumac>
  Jim Carrey
  </glavni_glumac>

</glazba>
```

Podaci se u XML-u zapisuju uz pomoć elemenata koji se sastoji od 2 oznake i sadržaja. Na ovome primjeru vidimo kako prva oznaka ima vrijednost <naziv_filma>, sadržaj se odnosi na The Truman Show a druga oznaka je </naziv_filma>. Oznake se uvijek pišu unutar < >, a

³⁴ Harold Rusty, Elliotte, Means Scott W. Introducing XML. // XML in a Nutshell, 2nd edition / Elliotte Rusty Harold, W. Scott Means. SAD: O'Reilly & Associates, Inc, 2002. 3-8 str.

ako se sastoje od dvije riječi između ide znak za donju crtu (`_`). Druga oznaka je zatvarajuća stoga se za zatvaranje koristi znak kose crte (`/`). Sadržaj je ono što se nalazi između dviju oznaka.

Osim već opisanog i prikazanog suvremenog standarda za razmjenu podataka XML-a, postoji još jedan sličan format koji je u današnje informacijsko doba jako korišten. To je format JSON i o njemu će se više pričati u nastavku teksta.

4.2.2 JSON

JSON je akronim za JavaScript Object Notation i služi nam za organizirano i lako dostupno pohranjivanje informacija.³⁵ Kao i XML, JSON je također razumljiv i ljudima i računalima te je kompatibilan sa svim programskim jezicima, ali zasniva se na JavaScriptu. Ovo je otvoreni standard koji dohvaća i pohranjuje podatke na način da je s njima jednostavno programski raditi.

Nastao je 1999. godine, a kreirao ga je Douglas Crockford koji je i popularizirao termin 'JSON'³⁶. No, JSON nije nastao odjednom, preko noći. Ideja za formatom koji ima osobine JSON-a je stara barem od 1996. godine.³⁷ Ono što je omogućilo JSON-u da bude toliko popularan je pojava hardvera i softvera kojima je bilo potrebno nešto za izmjenu podataka kako bi mogli funkcionirati.³⁸

JSON sadrži jasnu i jednostavnu strukturu podataka a osnovna struktura podataka kojom se koristi je objekt. Objekt je skup ključ-vrijednost parova, od kojih ključ mora biti tekst a vrijednost može biti tekst, null, popisi, booleove vrijednosti, brojevi i sam objekt. JSON za razliku od XML-a ne koristi ranije opisane oznake zbog čega je još više razumljiv i jednostavan za korištenje. Korištenje JSON-a se obično događa u 2 smjera: parsiranje ili pretvaranje niza koji sadrži JSON u JavaScript objekt i obrnuto, pretvaranje JavaScript objekta u JSON niz.³⁹

Ukratko, JSON u svome djelovanju ima velik broj prednosti nad ostalim formatima i standardima a to je da je pogodan svima za korištenje, lako ga je naučiti i koristiti zbog jednostavne strukture podataka te je manje veličine od ostalih formata jer se ne služi velikim brojem dodatnih znakova. U nastavku teksta je prikazan zapis u JSON formatu.

³⁵ SQLizer. URL: <https://blog.sqlizer.io/posts/json-history/>. [2020-10-12]

³⁶ Ibid.

³⁷ Ibid.

³⁸ Ibid.

³⁹ Squeak. URL: <https://squeak.ru/hr/megafon/peremennaya-json-vvedenie-v-json-chto-znachit-json.html>. [2020-10-12]

Ispis 3: Primjer JSON zapisa

```
{
  'naziv filma': The Truman Show,
  'godina filma': 1998,
  'glavni glumac': Jim Carrey
}
```

Na ovome jednostavnom i kratkom primjeru JSON zapisa možemo uočiti osnovnu strukturu kojom se koristi – objekt. Zapis se sastoji od 3 objekta od kojeg se svaki sastoji od jedne kombinacije ključ – vrijednost parova. Dakle, "naziv filma" je ključ koji mora biti tekst, a The Truman Show je vrijednost, u ovom slučaju isto tekst. Svi objekti se moraju pisati unutar znakova za vitičaste zagrade {}. Ključ se zapisuje pod navodnicima, nakon čega slijedi znak za dvotočku (:) i posljednja se zapisuje vrijednost. Objekti se razdvajaju znakom za zarez (,) jer se radi o nabranjanju.

4.3 Programska razmjena podataka

Važno je za istaknuti da su opisani formati, kao i raniji MARC i BibTeX razvijeni za programsko korištenje. U umreženom svijetu, ovakve podatke se može programski dohvaćati kroz online usluge, takozvana API sučelja. Primjer ovakvih usluga su razna API sučelja koja nudi sustav Crossref.

4.3.1 API sučelja

Aplikacijsko programsko sučelje (API) je sučelje koje je za razliku od grafičkog sučelja namijenjenog ljudskom korisniku, stvoreno za programski pristup i iskoristivost podataka. Uz pomoć API-ja koristimo već postojeću web infrastrukturu za dobivanje i rad sa strukturiranim podacima preko URL-a, a ne HTML-a koji je namijenjen za prikaz čovjeku. „Na webu API-ji omogućuju integraciju aplikacija s raznim uslugama kao što su npr. Google Maps i Facebook.“

⁴⁰Također, važna značajka mrežnih API-ja je to što podržavaju web-preglednike, web baze podataka i mobilne aplikacije, te su podržane u programskim jezicima i operativnim sustavima.⁴¹

⁴⁰ 4meahc. URL: <https://hrv.4meahc.com/about-network-apis-25487>. [2020-10-12]

⁴¹ Ibid.

API usluge su prestale koristiti formate kao što su BibTeX jer bi to značilo da je potrebna izrada zasebnih formata za različite vrste podataka i različite potrebe korisnika. Namjera je stvaranje univerzalnog formata za razmjenu bilo kakvih podataka, neovisno o tome radi li se o bibliografskim metapodacima, podacima o prognozi vremena, podacima o kemijskim spojevima, o tekstovima pjesama i slično. Podaci dobiveni iz ovakvih sučelja su u XML i JSON formatu, što ih čini izrazito jednostavnima za korištenje kao i za samo dohvaćanje. Kako bi se dobila jasnija ideja o tome kako se vrši dohvaćanje metapodataka s aplikacijskih programskih sučelja, u praktičnom projektu koji slijedi prikazat će se dohvaćanje bibliografskih metapodataka za neki DOI identifikator.

5. Praktičan projekt: programski dohvat bibliografskih metapodataka putem programskog jezika Python

U ovome dijelu rada prikazat će se programski dohvat bibliografskih metapodataka za DOI identifikatore 10 radova koji se nalaze na webu putem programskog jezika Python. Biti će prikazani i objašnjeni pojedinačni dijelovi koda koji se smatraju najvažnijima kao i nekoliko primjera rezultata projekta, tj. prikaz onoga što program obavlja. U dodacima će se nalaziti prikaz cijelog koda i rezultata.

Ispis 4: Prikaz funkcije getUrl()

```
def getUrl(url, encoding='utf-8'):  
    response = urllib.request.urlopen(url)  
    raw_text = response.read()  
    text = raw_text.decode(encoding)  
    response.close()  
    return text
```

Funkcija getUrl služi kako bi se dohvatili podaci sa weba. Ono što u primjeru nije prikazano, ali je važno za spomenuti je da se prije korištenja ove funkcije u Python treba uvesti modul urllib.request. Dakle, getUrl() je metoda koja uz pomoć modula urllib.request šalje zahtjev za dohvat podataka koji će biti enkodirani u UTF-8. Podaci koji su dohvaćeni su sirovi podaci koji se zatim dekodiraju. Zahtjev za dohvaćanjem podataka se zatvara i podaci se vraćaju kao tekst.

Ispis 5: Prikaz funkcije saveUrl()

```
def saveUrl(url, path, encoding='utf-8'):  
    text = getUrl(url, encoding)  
    out_file = open(path, 'w', encoding=encoding)  
    out_file.write(text)  
    out_file.close()
```

Ova funkcija služi kako bi dobiveni podaci bili spremljeni. Prvo se datoteka otvori uz pomoć naredbe `open` a zatim se ispiše kao tekst uz pomoć naredbe `.write`. Na kraju se datoteka zatvori s naredbom `.close`.

Ispis 6: Prikaz koda preuzetog iz projekta

```
doi_file = open ('./dois.txt', encoding = 'utf-8')  
  
bib_file = open ('./bibliografija.txt', 'w', encoding = 'utf-8')  
  
path_data = './data.json'  
  
for doi in doi_file:  
    doi = doi.strip()  
    print('resolving doi:', doi)  
  
    doi_url = url_form.format(doi, 'mcota@student.unizd.hr')  
    saveUrl(doi_url, path_data)
```

`Doi_file` otvara tekstualnu datoteku `dois.txt` u kojoj su spremljeni DOI identifikatori, a `bib_file` otvara tekstualnu datoteku u kojoj će se pokretanjem programa ispisati podaci u formatu `.txt`. `Path_data` definira naziv i putanju datoteke u koju će se spremati podaci u `.json` formatu. Nakon ovog dijela koda prikazane su naredbe koje će ukloniti višak redaka između DOI identifikatora u `doi_file` datoteci. Na ekranu će se ispisati tekst 'resolving doi:', a nakon

teksta će se prikazati DOI identifikator. Posljednje 2 linije koda se odnose formu koja definira DOI link. Link se sastoji od DOI identifikatora i e-mail adrese za identifikaciju korisnika.

Ispis 7: Prikaz rječnika

```
with open(path_data, encoding='utf-8') as file:
    data = json.load(file)

title = data['created']['title']
journal = data['container-title']
year = data['created']['date-parts'][0]
volume = data['created']['volume']
issue = data['created']['issue']
pages = data['page']
```

Dio koda koji se odnosi na rad s DOI podacima. Ova struktura podataka se naziva rječnicima. Oni omogućuju hijerarhijsko slaganje kako bi se podaci mogli dohvatiti. Hijerarhijsko slaganje uključuje redoslijed kojim je određena lokacija dokumenta. Da bi se došlo do podataka o naslovu, časopisu, godinu, volumenu, broju i stranicama potrebno je znati kako i gdje su ti podaci stacionirani jer se dohvaćaju preko cijele putanje, a ne preko konačne lokacije. Tako ovdje uočavamo kako se svi podaci nalaze u datoteci 'data'. Do podatka o naslovu ćemo doći preko 'created', podatak o časopisu je određen s 'container-title', podatak o godini se nalazi na putanji 'created', 'date-parts' i [0]. Ova 0 se odnosi na prvi element u indeksnoj listi. Volume se nalazi na putanji 'created', 'volume'. Na isti način se dolazi i do podataka o izdanju i stranicama.

Nakon što smo dohvatili podatke, potrebno je pripremiti autore i stvoriti formu za stvaranje teksta za jednog autora. Nakon toga se stvara prazan popis u koji će se pospremati tekstovi autora. Zatim se uz pomoć metode 'for in' definira da se pospremaju tekstovi za svakog autora. Za autore se stvara string "prezime, ime" kako bi se podacima svih DOI-a prikazali istim redoslijedom. Posljednje što treba napraviti je spojiti popis autora sa ';' i definirati nazive kao što vidimo u dolje prikazanom ispisu.

Ispis 8: Prikaz koda koji služi za definiranje naziva

```
citation = cit_form.format(  
    autor=authors,  
    godina=year,  
    naslov=title,  
    casopis=journal,  
    volumen=volume,  
    broj=issue,  
    stranice=pages
```

Kao rezultat ovog projekta dohvatili smo podatke s weba preko DOI identifikatora. Podaci za svakog autora su zapisani po sljedećem redoslijedu: autor, godina, naslov, časopis, volumen, broj i stranice. Kako to izgleda možemo vidjeti na idućem primjeru:

Ispis 9: Primjer rezultata projekta

Bourne, Charles P. (2002). Frequency and impact of spelling errors in bibliographic data bases / Information Processing & Management. 13(1). str. 1-12.

Ferrara, Alfio; Salini, Silvia (2012). Ten challenges in modeling bibliographic data for bibliometric analysis / Scientometrics. 93(3). str. 765-785

Ono što je važno za naglasiti je da ovome programu ima još puno prostora za nadogradnju i poboljšanje. Jedna od mana koje bih izdvojila je to što ne javlja pogrešku ako neki od tekstova koje smo preuzeli nema dio koji je u definiran u formiranju redoslijeda podataka. Npr. ako u nekom od tekstova fali podatak o volumenu, program nam neće javiti pogrešku.

Prikaz cijelog koda i rezultata se nalazi u Prilozima.

6. Zaključak

Metapodatke smatramo osnovom svakog informacijskog sustava stoga je bitno da su oni kvalitetni i da se informacijske ustanove s njima mogu koristiti na jednostavan i razumljiv način. Računalna razmjena metapodataka je imala veliki utjecaj na rad knjižnica jer im je omogućena međunarodna razmjena bibliografskih zapisa. Prvi format koji se razvio isključivo za tu svrhu je format MARC. Neki od najvažnijih doprinosa MARC-a je to što je on strojno

čitljiv i što je zamijenio kataložne listiće. Iako je uloga MARC-a u razmjeni bibliografski podataka velika, danas se on smatra zastarjelim formatom koji nije prilagođen današnjem informacijskom društvu jer se razvio prije weba. Većini ljudi je nerazumljiv i previše kompleksan za služenje.

Pojavom weba razmjena podataka i dokumenata je postala organiziranija i jednostavnija. Do informacija se dolazi na brži i lakši način jer su postale dostupnije. Ovo naravno utječe i na knjižnice koje moraju mijenjati svoj način poslovanja. Ono najbitnije vezano za web je razvoj formata i usluga koje će omogućiti suvremen pristup strukturiranim bibliografskim metapodacima. Prije weba formati su bili specijalizirani za razmjenu određenih metapodataka te su služili za razmjenu metapodataka među korisnicima, a ne za potrebe stručnog korištenja.

Osim knjižničara, postoje i razni mnogi drugi korisnici bibliografskih metapodataka kao npr. znanstvenici, novinari, studenti itd. Ova vrsta korisnika ne zna baratati MARC formatom stoga se za njihove potrebe razvio softver namijenjen isključivo upravljanju referencama. Ti formati se nazivaju Zotero, Mendeley, EndNote itd. MARC nije format koji je namijenjen korištenju krajnjih korisnika. Za njihove potrebe razvio se BibTex koji služi kao format za razmjenu bibliografskih podataka u akademske svrhe. No, nije kreiran za rad u Pythonu i za web.

Formati nastali nakon pojave weba su XML i JSON, a neke od usluga su Crossref i Open Library. Najvažnije za ponoviti vezano za ove formate jest da su oni prilagođeni za programsku iskoristivost u JavaScriptu i Pythonu. Također, krajnjim korisnicima su razumljivi i mogu se naučiti koristiti njima. Open Library je online usluga koja izrađuje digitalnu knjižnicu te pruža pristup svakom tiskanom djelu bilo kome na svijetu. Crossref služi kao usluga za povezivanje referenci između članaka čime se olakšava korisniku pri potrazi citiranih predmeta.

Dakle, za sam kraj rada valja zaključiti kako knjižnice imaju kvalitetne podatke, ali ti podaci nisu dostupni u suvremenim strukturama te ih obični korisnici ne mogu koristiti. Dok druge usluge imaju vrlo kvalitetne podatke (Crossref) ali ih ne pružaju u standardima namijenjenim za knjižnice već web standardima. Ponekad ove usluge imaju i podatke slabije kvalitete koji bi bili kvalitetniji da ih nude knjižnice. Da bi se iskoristio sav potencijal koji bibliografska razmjena podataka pruža, potrebno je osuvremeniti ili zamijeniti neke od formate koji se koriste u knjižnicama te postići programsku razmjenu s web standardima.

7. Popis literature

1. Collins, Susan. Introducing Crossref, the basics, 2018. URL: <https://jasolutions.com.co/crossref/introducing-crossref-basics/>. [2020-10-12]
2. Crossref. URL: <https://www.crossref.org/truths/>
3. DOI. URL: <https://www.doi.org/>.
4. Elsevier. URL: <https://www.elsevier.com/solutions/mendeley>
5. Furrie, Betty. Why is a MARC record necessary. // Understanding MARC bibliographic: machine readable cataloging / Betty Furrie. Washington, DC: Follett Software Co., 1988. 3-4.str
6. Harold Rusty, Elliotte, Means Scott W. Introducing XML. // XML in a Nutshell, 2nd edition / Elliotte Rusty Harold, W. Scott Means. SAD: O'Reilly & Associates, Inc, 2002.
7. Kvesić Grašić, Tina. Pravila Knjižnice grada Zagreba, 22.04.2020. URL: <http://www.kgz.hr/hr/informacije/upisi-i-posudba-196/pravila-knjiznica-grada-zagreba-40217/40217>
8. ¹ Ledinek, Sanja. Internet i World Wide Web dva su različita pojma – ne miješajte ih, 29.04.2014. URL: <https://www.racunalo.com/internet-i-world-wide-web-dva-su-razlicita-pojma-ne-mijesajte-ih/>.
9. Leksikografski zavod Miroslav Krleža. URL: <https://www.enciklopedija.hr/natuknica.aspx?id=63215>
10. Leksikografski zavod Miroslav Krleža. URL: <https://www.enciklopedija.hr/Natuknica.aspx?ID=40384> (2020-9-23)
11. Leksikografski zavod Miroslav Krleža. URL: <https://www.enciklopedija.hr/Natuknica.aspx?ID=66413>
12. Morton, Katharine. The Marc MARC Formats: An Overview. // The American Archivist (1986) 49 (1)
13. Open Library. URL: <https://openlibrary.org/>
14. : <https://www.xn--ina-usluge-ukb.hr/world-wide-web-proslavio-je-30-rodendan/>.
15. SQLizer. URL: <https://blog.sqlizer.io/posts/json-history/>.

16. Squeak. URL: <https://squeak.ru/hr/megafon/peremennaya-json-vvedenie-v-json-chno-znachit-json.html>.
17. Tennant, Roy. MARC must die. URL: <http://soiscompsfall2007.pbworks.com/f/marc+must+die.pdf> (2020-10-11)
18. The Library of Congress. URL: <https://www.loc.gov/marc/>
19. Zhang, Cui-xiao, et al. Data Exchange based on Web Services. // International Journal of Computer Science
20. and Network Security 6, 5A (2006.)
21. W3C. URL: <https://www.w3.org/XML/>
22. 4meahc. URL: <https://hrv.4meahc.com/about-network-apis-25487>

Bibliographic data exchange: programming interfaces and formats

Main topic of this paper is bibliographic data exchange. Also, context in which is developed, importance of it for libraries and for final users are described in this paper. It is also about significance of MARC format since it is first one of its kind. Main goal of this paper is to show different contemporary services, formats and software that are used for bibliographic data exchange. Main focus is on the types of users rather than professional use in libraries. Some of the bibliographic data exchange services that are described in the paper are Crossref and Open Library. Emphasis is on API interfaces but also formats such as XML and JSON. With the already mentioned, this paper also covers reference management software like Zotero and Mendeley but also format like BibTex. A project about bibliographic metadata storage and retrieval is shown as well. Paper is organized as professional review of previously mentioned topics.

Key words: data exchange, bibliographic data, metadata, XML, JSON, BibTex, Crossref, Open Library, Zotero, Mendeley.

8. Popis ispisa

Ispis 1. Primjer BibTeX zapisa

Ispis 2. Primjer XML zapisa

Ispis 3. Primjer JSON zapisa

Ispis 4. Prikaz funkcije getUrl()

Ispis 5. Prikaz funkcije saveUrl()

Ispis 6. Prikaz koda preuzetog iz projekta

Ispis 7. Prikaz rječnika

Ispis 8. Prikaz koda koji služi za definiranje naziva

Ispis 9. Primjer rezultata projekta

9. Prilozi

Prilog 1. Cjeloviti prikaz koda

```
import json
import urllib.request

url_form = "http://www.crossref.org/openurl/?id=doi:{}&noredirect=true&pid={}&format=json"
cit_form = "{autor} ({}godina). {naslov} / {casopis}. {volumen}({broj}). str. {stranice}.\n"

def getUrl(url, encoding='utf-8'):
    response = urllib.request.urlopen(url)
    raw_text = response.read()
    text = raw_text.decode(encoding)
    response.close()
    return text

def saveUrl(url, path, encoding='utf-8'):
    text = getUrl(url, encoding)
    out_file = open(path, 'w', encoding=encoding)
    out_file.write(text)
    out_file.close()

doi_file = open ('./dois.txt', encoding = 'utf-8')
bib_file = open ('./bibliografija.txt', 'w', encoding = 'utf-8')
path_data = './data.json'

for doi in doi_file:
    doi = doi.strip()
    print('resolving doi:', doi)

    doi_url = url_form.format(doi, 'mcota@student.unizd.hr')
    saveUrl(doi_url, path_data)

    with open(path_data, encoding='utf-8') as file:
        data = json.load(file)
```

Slika 1. Kod preuzet iz projekta

```

title = data['created']['title']
journal = data['container-title']
year = data['created']['date-parts'][0]
volume = data['created']['volume']
issue = data['created']['issue']
pages = data['page']

aut_form = "{prezime}, {ime}"

author_list = []

for author_data in data["author"]:

    author_string = aut_form.format(
        prezime=author_data['family'],
        ime=author_data['given']
    )
    author_list.append(author_string)

authors = '; '.join(author_list)

citation = cit_form.format(
    autor=authors,
    godina=year,
    naslov=title,
    casopis=journal,
    volumen=volume,
    broj=issue,
    stranice=pages
)

bib_file.write(citation)

doi_file.close()
bib_file.close()

```

Slika 2. Kod preuzet iz projekta

Prilog 2. Cjeloviti ispis rezultata

Bourne, Charles P. (2002). Frequency and impact of spelling errors in bibliographic data bases / *Information Processing & Management*. 13(1). str. 1-12.

Ferrara, Alfio; Salini, Silvia (2012). Ten challenges in modeling bibliographic data for bibliometric analysis / *Scientometrics*. 93(3). str. 765-785.

Kazienko, Przemysaw; Brodka, Piotr; Musial, Katarzyna; Gaworecki, Jarosaw (2010). Multi-Layered Social Network Creation Based on Bibliographic Data / 2010 IEEE Second International Conference on Social Computing. (). str. 407-412.

Stefaniak, Barbara (2005). Use of bibliographic data bases for scientometric studies / *Scientometrics*. 12(3-4). str. 149-161.

Borman, Lorraine; Mittman, Benjamin (2007). Interactive search of bibliographic data bases in an academic environment / *Journal of the American Society for Information Science*. 23(3). str. 164-171.

Hennemann, S.; Wang, T.; Liefner, I. (2011). Measuring regional science networks in China: a comparison of international and domestic bibliographic data sources / *Scientometrics*. 88(2). str. 535-554.

Kessler, M. M. (2007). Bibliographic coupling between scientific papers / *American Documentation*. 14(1). str. 10-25.

Kumar, Sharma; Ujjal, Marjit; Utpal, Biswas (2013). Exposing MARC 21 Format for Bibliographic Data As Linked Data With Provenance / *Journal of Library Metadata*. 13(2-3). str. 212-229.

Kim, Jongwoo; Kanungo, Tapas; Le, Daniel X.; Thoma, George R.; Barney Smith, Elisa H.; Hu, Jianying; Kantor, Paul B. (2003). <title>Automated labeling of bibliographic data extracted from biomedical online journals</title> / *Document Recognition and Retrieval X*. 5010(). str. 47-56.