MARC standard i suvremeni formati za razmjenu podataka

Pjanić, Valentina

Undergraduate thesis / Završni rad

2019

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

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:162:886242

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-29**

Image not found or type unknown

Repository / Repozitorij:

University of Zadar Institutional Repository

zir.nsk.hr

Image not found or type unknown

Sveučilište u Zadru

Odjel za informacijske znanosti
Preddiplomski sveučilišni studij informacijske znanosti (jednopredmetni)

Valentina Pjanić

MARC standard i suvremeni formati za razmjenu podataka

Završni rad

Zadar, 2019.

Sveučilište u Zadru

Odjel za informacijske znanosti
Preddiplomski sveučilišni studij informacijske znanosti (jednopredmetni)

MARC standard i suvremeni formati za razmjenu podataka

Završni rad

Student/ica:
Valentina Pjanić

Mentor/ica:
dr. sc. Krešimir Zauder

Zadar, 2019.



Izjava o akademskoj čestitosti

Ja, Valentina Pjanić**,** ovime izjavljujem da je moj završni rad pod naslovom MARC standard i suvremeni formati za razmjenu podataka rezultat mojega 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 mojega 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 mojega rada u potpunosti odgovara sadržaju obranjenoga i nakon obrane uređenoga rada.

Zadar, 2019.

Sadržaj:

1. UVOD 1
2. MARC standard 1
3. SUVREMENI FORMATI ZA RAZMJENU PODATAKA 4
3.1. XML 4
3.2. JSON 5
4. PRIKAZ TEHNOLOGIJA 7
4.1. Nedostaci MARC formata 7
4.2. Prednosti suvremenih formata 11
4.3. Python 13
5. IMPLEMENTACIJA 14
5.1. Cilj i svrha 14
5.2. Metodologija 15
5.3. Rezultati 17
5.4. Rasprava 18
6. ZAKLJUČAK 19
7. LITERATURA 20
8. POPIS SLIKA 23
9. POPIS ISPISA 23
10. PRILOZI 23

Sažetak

U ovom se radu govori o MARC standardu i suvremenim podacima pri opisu bibliografske građe. Svrha rada je utvrditi i usporediti mogućnosti bibliografskog opisivanja građe koristeći MARC standard i suvremene formate, odnosno XML i JSON, te predložiti prikladniji format, kao i proces pretvaranja iz MARC formata u XML i JSON. Cilj rada je istražiti prednosti i mane MARC, XML i JSON formata, usporediti način zapisa podataka u svakom od formata te naglasiti važnost i prednosti korištenja suvremenih formata. Kroz rad se prikazuju pojedinosti o svakom formatu, njihove prednosti i nedostaci i prikazani su primjeri zapisa svakog od formata. Posebno su prikazani nedostaci MARC formata, a naglašene prednosti formata XML i JSON. U istraživanju je prikazan model programa, izrađenih u programskom jeziku Pyton uz korištenje modula Pymarc, pomoću kojih se MARC zapisi pretvaraju u XML ili JSON format. Istraživanje je pokazalo da su suvremeni formati u prednosti nad MARC formatom koji je već pomalo zastario. Kao glavne prednosti suvremenih formata navodi se jednostavnost čitanja i pisanja zapisa, a izradom programa prikazan je proces pretvorbe formata. Pomoću izrađenog programa utvrđeno je da proces konverzije formata nije kompliciran i dugotrajan proces, a i za izradu programa dovoljno je imati osnovno znanje o programiranju.

Ključne riječi: MARC, XML, JSON, suvremeni formati

1. UVOD

Knjižnice, kao informacijske ustanove imaju potrebu za čuvanjem i distribucijom informacija. Jedan od njihovih zadataka je opis građe kako bi se ista mogla katalogizirati. Dapače, formalan opis građe je temelj organizacije znanja. Prije nastanka i razvitka informacijskih tehnologija, za svrhe katalogizacije koristili su se katalozi na listićima. Nove tehnologije omogućile su efikasnije načine pohrane, organizacije i korištenja bibliografskih informacija kao i njihovu razmjenu. Pojavljuje se i sve više građe koju je potrebno katalogizirati pa katalozi na listićima više nisu bili dostatni. Prema katalozima na listićima, došlo je do stvaranja strojnočitljivog formata, nazvanog MARC, a čija je svrha opis bibliografske i druge građe. MARC standard se uvelike razvio i koristi se u gotovo cijelom svijetu u raznim verzijama tog formata, a razvojem UNIMARC-a, došlo je do mogućnosti međunarodnog dijeljenja prethodno stvorenih zapisa. Međutim, daljnji razvoj tehnologije, posebno razvoj interneta i weba te sve veći fokus na strukturirane podatke, doveo je do nastanka brojnih formata za pohranjivanje i distribuciju podataka, što je dovelo do pitanja je li MARC kao format i dalje dostatan za opis bibliografskih informacija. Kao primjer se uzimaju formati XML i JSON, formati za označavanje i razmjenu podataka, koji iako nisu stvoreni za opis bibliografskih podataka, potencijalno mogu poslužiti kao bolja zamjena za MARC format. Važno je naglasiti da je podatke prema MARC standardu moguće direktno zapisati i u ove formate bez gubitaka informacija. Drugo je pitanje da li sam MARC standard treba osuvremeniti budući da je isti nastao kao računalni zapis kataložnog listića, ali to je izvan opsega ovog rada.

U ovom radu predstavit će se prednosti i mane formata MARC, XML i JSON, usporediti na koji se način podaci zapisuju i prikazuju u svakom od formata te naglasiti važnost korištenja suvremenih formata u današnje vrijeme. Rad će također demonstrirati i način zapisa MARC podataka u XML i JSON.

2. MARC standard

Od nastanka knjižnica, postojala je potreba za katalogiziranjem građe kako bi se olakšalo njeno čuvanje, organizacija i distribucija. Prije automatizacije sustava korišteni su katalozi na listićima, no razvojem tehnologije došlo je do ideje stvaranja strojno čitljivih kataloga. Tako je nastao standard MARC, kojemu je cilj bio prenijeti kataloge na listićima u strojno čitljiv oblik. MARC je akronim za Machine readable cataloguing, međunarodni standard za strojno čitljivo katalogiziranje, kao i prikaz i razmjenu bibliografskih podataka[1]. „Prema definiciji Alana Hopkinsona i Ellen Gredley, MARC se može opisati kao skupina formata koji primjenjuju određeni skup konvencija za identifikaciju i uređivanje bibliografskih podataka

[1] Leksikografski zavod Miroslav Krleža. URL: http://www.enciklopedija.hr/natuknica.aspx?ID=38783 [2019-08-16]

radi obrade računalom.[2]“ Stvoren je kako bi se katalogiziranje bibliografskih jedinica olakšalo te kako bi se omogućio jednostavan unos bibliografskih podataka u kataloge. Strojevi, odnosno računala, ne mogu razumjeti podatke koji se unose, već je potrebno postaviti određene atribute ili pravila prema kojima će se podaci unositi, stoga strojno čitljivo u nazivu MARC-a označava mogućnost stroja da prepozna kako formatirati koji podatak. Katalogiziranje kao pojam označava sami unos bibliografskih podataka za određenu bibliografsku jedinicu te pohranu u katalog.

MARC standard stvorila je Kongresna knjižnica šezdesetih godina dvadesetog stoljeća, nakon čega su druge knjižnice u svijetu stvarale svoje verzije[3]. MARC stvoren u Sjedinjenim Američkim Državama nazvan je USMARC, dok je britanska verzija nazvana UKMARC. Prema USMARC-u nastao je format MARC 21, koji koristi ISO standard 2709 za opis bibliografskih podataka. Uz UNIMARC, MARC 21 smatramo najkorištenijim MARC inačicama. Njegova struktura uključuje korištenje blokova od 1XX do 9XX te oznaka za povezivanje, a blokovi su[4]:

0XX Kontrolne informacije, brojevi, kodovi

1XX Glavna kataložna jedinica

2XX Stvarni naslovi, izdanje, imprint

3XX Materijalni opis itd.

4XX Nakladnička cjelina (kako je otisnuta u knjizi)

5XX Napomene

6XX Sporedne kataložne jedinice za predmetne odrednice

7XX Druge sporedne kataložne jedinice osim za predmetne odrednice i nakladničku cjelinu; polja za povezivanje

8XX Sporedne kataložne jedinice za nakladničku cjelinu (drugi valjani oblici); podaci o fondu; alternativna pisma

9XX Lokalne jedinice podatka

Za razvoj i održavanje formata zaslužan je program Univerzalne bibliografske kontrole (UBC), a temelji se na idućim načelima[5]:

1. Svaka zemlja treba preuzeti odgovornost za bibliografsku kontrolu svojih publikacija
2. Zapisi već izrađeni u jednoj ustanovi trebali bi biti dostupni svim drugim ustanovama

Postupno su nastajale i druge verzije tog standarda, sve dok IFLA 1977. godine nije stvorila UNIMARC, osmišljen za korištenje u cijelom svijetu s jedinstvenim pravilima katalogiziranja[6].

[2] Willer, Mirna. UNIMARC u teoriji i praksi. Što je MARC. Rijeka: Naklada Benja, 1996. Str 13.
[3] Gorman, Michael. Technical services: Today and tomorrow. Str. 88.
[4] Taylor, Arlene G., Joudrey, Daniel N. The organisation of information. Str. 135.-136.
[5] Ibid, str. 13-14
[6] Ibid, str. 16.

UNIMARC je MARC standard koji se i danas najčešće koristi u knjižnicama diljem svijeta, a tako i u Hrvatskoj. Struktura UNIMARC formata također se sastoji od blokova i oznaka za povezivanje, a od MARC-a 21 se razlikuje po tome što više ne koristi terminologiju kataloga na listićima već blokove podataka strukturirane prema funkciji.

Blokovi u UNIMARC-u su[7]:

0— blok za identifikaciju

1— blok kodiranih informacija

2— blok glavnog opisa

3— blok napomena

4— blok za povezivanje kataložnih jedinica

5— blok srodnih naslova

6— blok sadržajne analize

7— blok podataka o odgovornosti

8— blok za međunarodnu upotrebu

Svaki blok sadrži potpolja, u koje se ulazi pomoću njihovih kodova i upisuju unaprijed zadani podaci, ovisno o vrsti jedinice građe koja se opisuje.

LDR	00000cam a2200385 ir4500
001	000173516
003	HR-ZaNSK
005	20070917154708.0
008	961004s1996 ci 001 0 hrv
015	|a HK96-1539 |2 HR
020	|a 9536003252
035	|9 (HR-ZaNSK)173718
035	|9 (HR-ZaNSK)961004079
035	|a (HR-ZaNSK)000173516
040	|a HR-ZaNSK |b hrv |c HR-ZaNSK |e ppiak
0410	|a hrv
042	|a croatica
044	|a ci |c hr
080	|a 025.31:004](035)
080	|a 004:02](035)
080	|a 006:02](035)
1001	|a Willer, Mirna
24510	|a UNIMARC u teoriji i praksi / |c Mirna Willer.
260	|a Rijeka : |b Naklada Benja, |c 1996. |e (Rijeka : |f Imprimo)
300	|a 287 str. ; |c 24 cm.
4900	|a Priručnici iz knjižničarstva ; |v knj. 3
500	|a Nakl. 700 primj
504	|a Bibliografija: str. 261-274 i uz tekst
504	|a Predmetno i imensko kazalo
650 7	|a UNIMARC |x Teorijska načela |2 nskps
650 7	|a UNIMARC |x Primjena |2 nskps
7600	|t Priručnici iz knjižničarstva |w (HR-ZaNSK)000094363
981	|a A20/96 |p CRO |r HRB1996
997	|a 287 str. |b 700 primj
LKR	|a UP |b 000094363 |l NSK01 |m UNIMARC u teoriji i praksi |n Priručnici iz knjižničarstva |r 7600
8524	|j 665.275
876	|e CM |a 7759/97
8860	|2 unimarc |b 01074nam2 2200361 450
SYS	000173516

[7] Unimarc skraćeni bibliografski format. Willer, Mirna.

Slika 1[8]. Primjer MARC zapisa iz kataloga Nacionalne i sveučilišne knjižnice

3. SUVREMENI FORMATI ZA RAZMJENU PODATAKA

Informacijske se tehnologije svakim danom poboljšavaju, nastaju nove, što dovodi do sve veće raširenosti istih. Razvija se i svijest o dosezima i novim mogućnostima tih tehnologija. Te tehnologije su nam danas potrebne u gotovo svakom segmentu života i potrebno je pratiti njihov razvoj. Sve se više ljudi služi informacijskim tehnologijama i potrebno ih je pojednostaviti što je više moguće kako bi bile dostupne svima, što uključuje i standarde i formate za katalogiziranje građe. Mnogi formati stvoreni u prošlosti i dalje su u uporabi i obavljaju svoje zadaće, ali se često smatraju zastarjelima. Odabir formata najčešće je subjektivan te uvelike ovisi o potrebama, radilo se o osobnoj upotrebi ili u sklopu neke institucije. Neki od suvremenih formata stvoreni za obradu podataka su XML i JSON. Ti formati imaju svoje prednosti i nedostatke, ali počinju se sve više koristiti za zapis i razmjenu različitih vrsta strukturiranih podataka.

Neke od bitnih karakteristika formata za razmjenu podataka trebale bi biti: jednostavnost korištenja, preglednost podataka, kompatibilnost, kao i mogućnost jednostavnog dijeljenja zapisa s drugim ustanovama. Prethodno navedeni XML i JSON, kao formati za razmjenu podataka zadovoljavaju gore navedene karakteristike što ih čini pogodnima za korištenje pri čuvanju i razmjeni bibliografskih podataka, a o njima će biti više u idućim poglavljima. Vrlo je važno i napomenuti da ovi formati propisuju strukturu podataka i sintaksu zapisa, ali ne i semantiku podataka. Za razliku od MARC porodice, u ove je formate moguće zapisati podatke o bilo čemu, a ne samo bibliografske informacije.

3.1. XML

XML, odnosno eXtensible Markup Language, jedan je od suvremenih formata stvorenih za označavanje podataka. U prijevodu proširivi jezik za označavanje, nastao je prema SGML-u (Standard Generalized Markup Language) s ciljem preglednijeg opisa podataka korištenih u elektroničkom izdavaštvu[9]. Godine 1996., nekoliko je stručnjaka započelo s pojednostavljivanjem SGML-a kako bi stvorili označiteljski jezik jednako funkcionalan kao SGML, ali jednostavniji za korištenje pa je tako u veljači 1998. godine nastao XML 1.0, kao prva verzija XML-a[10]. Zbog jednostavnosti i mogućnosti prilagodbe, XML se danas koristi u razne svrhe, odnosno može se koristiti za pohranu i dijeljenje raznih vrsta informacija. Važno je napomenuti da se XML, za razliku od velikog broja prijašnje razvijenih formata, može pokrenuti na bilo kojem operativnom sustavu i bez puno problema svladati, ukoliko smo upoznati s označiteljskim jezicima,

[8] Katalog NSK. URL: http://katalog.nsk.hr/F/RRU3VC2GY6IAS5XYG7CL8464VJQ6CPR6G8K3FNCIA4YIUT2IF6-08346?func=full-set-set&set_number=006111&set_entry=000001&format=001 [2019-08-16]

[9] W3C. URL: https://www.w3.org/XML/ [2019-08-17]

[10] Elliote, Harold. XML in a nutshell. URL:https://docstore.mik.ua/orelly/xml/xmlnut/ch01_04.htm [2019-08-17]

kao što je HTML (Hyper Text Markup Language)[11]. HTML je format za definiciju prikaza sadržaja, a ne razmjenu podataka, što je dovelo do potrebe za stvaranjem XML formata. Jedna od glavnih razlika ta dva označiteljska jezika je ta što je XML jezik za metaoznačivanje te podržava hijerarhijsku strukturu podataka, dok je HTML specijalizirani jezik za označavanje dijelova dokumenata za hijerarhijski prikaz i nepraktičan je za pohranu i razmjenu strukturiranih podataka[12]. Kako bilo, svaki od jezika ima svoju svrhu i koristi se u skladu s njom.

U XML-u se za označavanje koristimo takozvanim elementima koje samostalno imenujemo. Element se sastoji od dvije oznake i sadržaja, a oznake se pišu u izlomljenim zagradama. Kod oznake koja označava kraj, prije naziva oznake unosi se kosa crta ('/')[13]. Oznake može činiti bilo koja riječ, na bilo kojem jeziku, što doprinosi njegovoj uporabljivosti, za razliku od HTML-a u kojem su unaprijed određene oznake i atributi.

Kongresna knjižnica je u suradnji s uredom MARC standarda odlučila razviti sustav koji omogućuje zapis u MARC standardu zapisati u XML formatu, a nazvan je MARC XML. Njegova je svrha olakšati zapisivanje MARC zapisa te korištenjem XML formata olakšati distribuciju zapisa[14].

```
<knjiga>
      <naslov>
      Unimarc u teoriji i praksi
      </naslov>
      <autor>
      Willer, Mirna
      </autor>
      <godina_izdavanja>
      1996.
      </godina izdavanja>
</knjiga>
```

Ispis 1. Primjer XML zapisa

3.2. JSON

Još jedan od suvremenih formata nazvan JSON, JavaScript Object Notation, općeniti je format za razmjenu podataka. Zasniva se na JavaScript programskom jeziku, a njegova je posebnost to što je ljudima

[11] Pitts, Natanya. Osnove XML. Upoznavanje XML-a, str. 7
[12] Microsoft office. XML za početnike. URL: https://support.office.com/hr-hr/article/xml-za-po%C4%8Detnike-a87d234d-4c2e-4409-9cbc-45e4eb857d44#bm6 [2019-08-27]
[13] Obdurodon.org. What is XML. URL: http://dh.obdurodon.org/what-is-xml.xhtml [2019-08-17]
[14] Library of Congress. MARC in XML. URL: https://www.loc.gov/marc/marcxml.html [2019-08-18]

jednostavan za čitanje i pisanje, ali je lako čitljiv i strojno. Kao format je potpuno zaseban standard, ali se najčešće koristi pomoću nekog od programskih jezika, kao što su na primjer JavaScript i Python[15]. JSON je upravo i nastao kao format koji zapisuje podatke na način na koji je s njima lako programirati. Primarna struktura podataka u JSON-u je objekt. JSON definira objekt kao skup ključ-vrijednost parova. Ključevi moraju biti tekst, a vrijednosti mogu biti tekst, brojevi, booleove vrijednosti, vrijednost *null*, popisi ili objekti. Iz navedenog je vidljivo da je JSON po samoj svojoj prirodi podoban za hijerarhijske podatke jer vrijednost nekog ključa može biti objekt, odnosno struktura vrijednosti. Zanimljivost je i da postoje baze podataka specijalizirane za pohranu JSON-a koje najčešće nazivamo „baze podataka orijentirane na dokumente“ (eng. *document oriented databases*), a propulzivan primjer ovakve baze je MongoDB[16]. Također, mnoge suvremene relacijske baze podataka dopuštaju pohranu JSON i XML zapisa direktno u vrijednosti atributa, kao ugniježđene strukturirane podatke.

Podaci zapisani u JSON formatu lako se dijele i prikazuju zato što korištenjem ključeva jednostavno možemo odabrati podatke koje želimo prikazati[17]. Najbolji način za baratanje podacima zapisanim u JSON-u je korištenjem nekog programskog jezika, što nam omogućuje potpuno manipuliranje željenim podacima na masovnoj razini. Iako stvoren prema JavaScript programskom jeziku, JSON je kompatibilan i s drugim programskim jezicima. Na primjer, Pythonov rječnik je strukturom jednak JSON objektu[18].

JSON format nastao je kao jednostavnija zamjena za XML. Iako je XML bogatiji format s više mogućnosti, JSON može prikazati veliki udio podataka za koje se često koristi XML, a ima prednost jednostavnosti i direktnosti korištenja kroz programske jezike. Manjak JSON-a je ista ta jednostavnost: XML ima kompletniji sustav standarada za, na primjer, validaciju i transformaciju. Također, vrijedi spomenuti da se danas se mnogu podaci „pakiraju“ u XML ili JSON iz kakvog internog rješenja po potrebi te korisnik sam može odlučiti u kojem formatu želi preuzeti podatke. Navedeno je temelj razmjene podataka na suvremenom webu. Mnoge napredni modeli podataka, poput RDF-a koji je temelj semantičkog weba, mogu se zapisati i u XML i u JSON.

[15] JSON. URL: https://www.json.org/ [2019-08-18]
[16] MongoDB. URL: https://www.mongodb.com/ [2019-09-22]
[17] Ibid.
[18] Coding networker blog. URL: https://codingnetworker.com/2015/10/python-dictionaries-json-crash-course/ [2019-08-25]

```
{
  "id": 1,
  "autor": "Willer, Mirna”,
  "naslov": "Unimarc u teoriji i praksi",
  "godina_izdavanja": 1996.
}
```

Ispis 2. Primjer JSON zapisa

4. PRIKAZ TEHNOLOGIJA

4.1. Nedostaci MARC formata

MARC je kao format za opis bibliografskih podataka nastao u vrijeme kada je tehnologija bila skupa i njegov je cilj bio što vjerodostojnije zamijeniti kataloge na listićima. Kao format koji se i dalje koristi, on odrađuje svoju zadaću, no s obzirom na činjenicu da je od njegovog nastanka stvoreno mnogo formata pristupačnijih za izradu i distribuiranje informacija, postoje prednosti kod prelaska na nove formate. Novi formati su pristupačniji za korištenje koristeći se općenitim alatima za obradu podataka i programskim jezicima. Navedeno vrijedi kako za korisnike knjižnice, uključujući i istraživače, tako i knjižničare. MARC format je tipično težak za direktan pristup podacima i ekstrakciju i transformaciju istih, čak i iskusnim programerima. XML i JSON s druge strane su sami sebi podržani od strane većine suvremenih programskih jezika i baza podataka te se po svojoj prirodi lakše uključuju u suvremene web tehnologije. Navedeno ne omogućuje samo kvalitetniju razmjenu podataka, već i eksperimentaciju s novim mogućnostima unutar struke. Obzirom da živimo u svijetu podataka koji tek počinje pokazivati kojekakve inovativne mogućnosti, ta eksperimentacija je vrlo važna. Knjižnice su mjesto distribucije informacija, stoga bi trebale ići u korak s novim tehnologijama koje im omogućavaju što kvalitetniji rad. Zapisi u MARC formatu nisu jednostavni za čitanje (ljudsko i strojno) i potrebno je imati barem osnovna znanja o formatu kako bi se mogli snaći u zapisu što ga ne čini pogodnim za prikaz osobama koje se nikada nisu susrele s nečim sličnim. Koristi se šifriranim oznakama pa čak i mnogi katalogizatori ne znaju veliki broj šifri, već samo osnovne koje se često koriste te se često za opis građe moraju služiti priručnikom. „Postoje samo dvije vrste ljudi koji vjeruju da mogu pročitati MARC zapis bez da se služe raznim priručnicima: nekoliko vrsnih katalogizatora i ljudi na teškim drogama.[19]“ Iz ove rečenice iz članka „MARC must die“, odnosno u prijevodu „MARC mora umrijeti“, možemo vidjeti kako autor Roy Tennant nije zadovoljan tim formatom i smatra da je potrebno pronaći bolje rješenje. U članku također naglašava kako je veliki problem to što se polja označavaju šiframa,

[19] Tennant, Roy. MARC must die. URL: http://soiscompsfall2007.pbworks.com/f/marc+must+die.pdf [2019-08-20]

što ga čini nečitljivim većini ljudi, odnosno svima koji nisu upoznati s njegovom sintaksom. Činjenica da je MARC kao format gotovo nemoguće koristiti bez priručnika trebala bi biti poticaj za prijelaz u jednostavniju shemu podataka, koja ne zahtijeva mnogo znanja već se uglavnom služi logikom.

Kao jedan od većih problema Tennant u svom članku naglašava i problem granularnosti te neujednačenosti pisanja zapisa[20]. Na primjer, mnogim je katalogizatorima bitno prenijeti zapis onako kako je to izgledalo na katalogu na listićima pa često neka polja izostave ili pojedini podatak upišu na za to nepredviđeno mjesto. Kod pisanja MARC zapisa u poljima bloka glavnog opisa prepisuju se podaci s naslovne stranice onako kako su tamo navedeni pa se često uz stvarne naslove dodaju podaci koji tamo ne pripadaju, kao na primjer podaci o izdanju, koji se pišu na za to predviđeno polje, odnosno polje o izdanju. U MARC formatu su mnoge vrijednosti jednostavno tekst koji se koristi kataložnom interpunkcijom, a ne strukturama podataka za raščlanjivanje. Navedeno osujećuje računalnu iskoristivost bibliografskih podataka i posebno je vidljivo prilikom transformacije u druge formate za razmjenu bibliografskih informacija gdje interpunkcija poput „/“ ili „:“ postane dio naslova ili mjesta objave.

Mnogi knjižnični katalozi korisnicima nude prikaz MARC zapisa, a neki i mogućnost preuzimanja MARC zapisa na vlastito računalo. Neki od njih su katalog Nacionalne i Sveučilišne knjižnice u Zagrebu, katalog Sveučilišne knjižnice Sveučilišta u Zadru, katalog knjižnice Filozofskog fakulteta u Zagrebu i katalog Kongresne knjižnice. Katalog Kongresne knjižnice nudi mogućnost dohvaćanja zapisa u nekoliko formata, uključujući MARC te preuzimanje takvog zapisa rezultira pohranom valjanog MARC zapisa. S druge strane, katalog NSK nudi mogućnost pohrane nevaljanog MARC zapisa zato što, osim oznaka i podataka od kojih se MARC zapis sastoji, u zapis su uneseni neki podaci koje MARC zapis inače ne sadrži. Takav zapis onemogućuje daljnji strojni rad s podacima, zato što se „višak“ ne prepoznaje kao MARC element i potrebno je „očistiti“ zapis kako bi se omogućila daljna uporaba. Navedeno poražava originalnu namjenu MARC-a odnosno strojnu čitljivost. Katalog knjižnice zadarskog sveučilišta koji se koristi Crolist integriranim knjižničnim sustavom i omogućen je samo prikaz MARC zapisa, a ne i pohrana na računalo, dok se katalog knjižnice Filozofskog fakulteta u Zagrebu koristi Koha integriranim knjižničnim sustavom koji ujedno dozvoljava i pohranu valjanih MARC zapisa u nekoliko formata.

[20] Ibid.

```
Datum: 22/09/2019


Pretražujući katalog Nacionalne i sveučilišne knjižnice, pronađeni su sljedeći
rezultati.

Podaci o pošiljatelju:


_____

FMT                   BK
LDR                   00000cam a2200385 ir4500
001                   000173516
003                   HR-ZaNSK
005                   20070917154708.0
008                   961004s1996 ci 001 0 hrv
...
```

Ispis 3. MARC zapis preuzet iz kataloga NSK

Author: Palin, Michael.
Title: Full circle : a Pacific journey / with Michael Palin.
Published: New York : St. Martin's Press, 1997.
Description: p. cm.
Dewey No.: 919.04 21
ISBN: 0312169604
Subjects: Pacific Area -- Description and travel.
Palin, Michael -- Journeys -- Pacific Area.

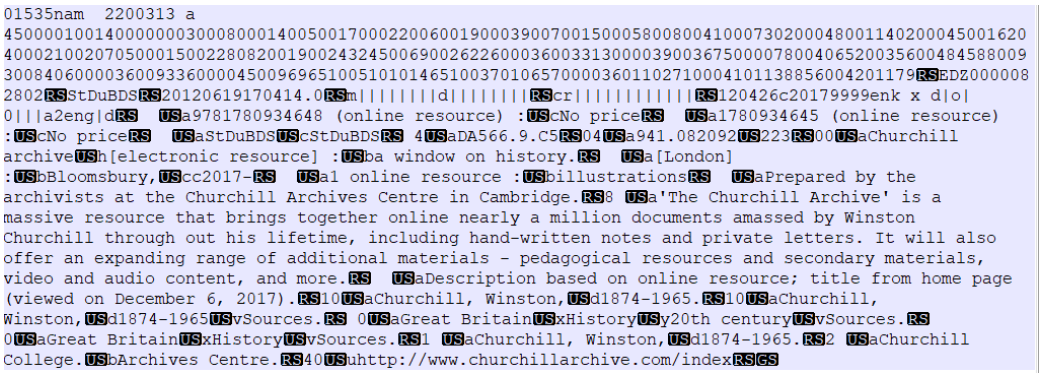
Slika 3[21]. Primjer MARC zapisa koji se prikazuje korisnicima

001 1652498
005 19970502082209.6
008 970428s1997 nyu 000 0aeng
035 $9(DLC) 97014664
906 $a7$bcbc$corignew$d1$eocip$f19$gy-gencatlg
955 $apc05 to sa00 04-28-97; lg09 to lg03 04-30-97; lg03 to sl 04-30-97; lg08 05-01-97 to DDC
010 $a 97014664
020 $a0312169604
040 $aDLC$cDLC
043 $ap------
050 00$aDU23.5$b.P35 1997
082 00$a919.04$221
100 1 $aPalin, Michael.
245 10$aFull circle :$ba Pacific journey /$cwith Michael Palin.
260 $aNew York :$bSt. Martin's Press,$c1997.
263 $a9710
300 $ap. cm.
651 0$aPacific Area$xDescription and travel.
600 10$aPalin, Michael$xJourneys$yPacific Area.

Slika 4[22]. MARC zapis kako ga knjižničari vide

[21]Libraries linking Idaho. How to Understand a MARC Record. URL: http://lili.org/forlibs/ce/able/course8/19understand1.htm [2019-08-25]
[22] Libraries linking Idaho. How to Understand a MARC Record. URL: http://lili.org/forlibs/ce/able/course8/20understand2.htm [2019-08-25]

Na slikama 3. i 4. nalaze se zapisi preuzeti s web stranice Libraries linking Idaho (lili.org)[23]. Slika 3. prikazuje zapis na način na koji se prikazuje korisnicima u katalogu. Svi su podaci jasno opisani i pregledni te je moguće jasno razumjeti o kojem se djelu radi, tko je autor, itd. Na idućoj slici, zapis je prikazan u originalnom MARC formatu, onakvom kakvim ga knjižničari, odnosno katalogizatori, vide. Možemo primjetiti kako se tu nalazi mnogo više podataka, nego što je prikazano korisniku, a među tim podacima su i podaci o katalogizatoru i ustanovi gdje je jedinica građe katalogizirana. Osim osnovnih podataka o jedinici građe, nalaze se i brojne šifre koje ljudima nisu razumljive, već samo strojevima. Iako zapis nakon formatiranja izgleda sasvim čitljivo i svi bitni podaci se prikazuju, postoji potreba za pojednostavljenjem procesa katalogizacije. Mnogi se od segmenata MARC zapisa mogu opisati na ljudima jednostavniji i razumljiviji način.



Slika 5[24]. Primjer MARC zapisa u uređivaču teksta

Na slici 5. možemo vidjeti kako zapis u MARC formatu izgleda, otvoren u uređivaču teksta, kao što je u ovom slučaju Notepad ++[25]. Podaci nisu formatirani za ljudsku čitljivost i zapis djeluje veoma nepregledno. Za razliku od zapisa na slici 2., koji je preuzet s web kataloga Nacionalne i sveučilišne knjižnice u Zagrebu, zapis u ovakvom izdanju je gotovo nemoguće razumjeti i iskusnim katalogizatorima. Da bi MARC zapisi bili čitljivi ljudima potrebno je koristiti neki od specijaliziranih programa za pisanje i čitanje MARC formata, što je jedna od njegovih većih mana, pogotovo obzirom na slabu dostupnost ovakvog softvera (u usporedbi s generičkim formatima poput XML-a i JSON-a). Podatke je moguće izmijeniti i korištenjem običnog uređivača teksta, no veoma se teško snaći u ovakvom zapisu, a pogotovo kada se radi o opsežnijem zapisu, odnosno sa više podataka. Ne začuđuje s toga što sami knjižničari često nisu u mogućnosti raditi sa „sirovim“ MARC zapisima. Također, MARC format se često koristi s MARC8

[23] Ibid.
[24] Churchill Archive. URL: http://www.churchillarchive.com/subscriber-services/marc-records [2019-08-22]
[25] Notepad ++. URL: https://notepad-plus-plus.org/ [2019-08-23]

kodnom stranicom koja je danas sve više problematična, pogotovo obzirom na prednost i rasprostranjenost UTF-8 kodne stranice.

4.2. Prednosti suvremenih formata

Kao što je već prethodno spomenuto u radu, od nastanka MARC formata do danas, stvoreno je mnogo novijih formata za razmjenu podataka i to posebno nakon razvoja WWW-a i suvremenih paradigmi organizacije, upravljanja i razmjene podataka. MARC format je specijaliziran za razmjenu bibliografskih podataka, no ne znači da drugi formati nisu pogodna zamjena. Mnogi formati za razmjenu podataka podržavaju opise raznih vrsta podataka, a imaju mogućnosti prilagoditi se ovisno o potrebama opisa. Suvremeni formati o kojima se govori u ovom radu su XML i JSON. Ta su dva formata odabrana zato što su se pokazali kao odlično rješenje pri pohrani i posebno razmjeni podataka. Oba formata imaju svoje prednosti, ali i mane, no odluka o formatu koji će određena ustanova koristiti, sasvim je individualna i subjektivna. XML i JSON su formati nastali ne tako davno, nakon što se informacijska tehnologija uvelike razvila, stoga su prilagođeni zapisivanju većih količina podataka. Vrlo je značajno i da su ovi formati nastali u sklopu razvoja weba te su utkani u način na koji strukturirani podaci kolaju internetom. Bez obzira tražili podatke od Ujedinjenih naroda, Svjetske banke, Facebooka, Twittera ili pak Crossrefa (prilikom razrješavanja DOI identifikatora), podatke ćemo dobiti u jednom od ova dva formata, a čest je slučaj da možemo i birati koji od ova dva formata želimo. Bitno je naglasiti i da je stvaranje XML i JSON zapisa mnogo jednostavnije nego što je slučaj kod MARC zapisa. Za razliku od MARC-a, XML i JSON zapisi mogu se zapisivati i čitati u običnim uređivačima teksta, na primjer u već spomenutom Notepad++-u.

XML kao format koristi se hijerarhijskom strukturom i njegovo je osnovno korištenje poprilično jednostavno ukoliko postoji osnovno znanje korištenja jezika za označavanje. Jedan od poznatijih ovakvih jezika je HTML, a kao i XML, nastao je prema SGML-u. Sva tri jezika imaju sličnu strukturu i način pisanja pa svladavanje jednog nakon što smo već upoznati s drugim ne bi trebao biti problem.

```
<crossref_result version="2.0"
                 xsi:schemaLocation="http://www.crossref.org/qrschema/2.0
https://www.crossref.org/schema/crossref_query_output2.0.xsd">
    <query_result>
        <head>
            <doi_batch_id>none</doi_batch_id>
        </head>
        <body>
            <query status="resolved" fl_count="4">
                <isbn type="print">978-1-61737-051-9</isbn>
                <isbn type="electronic">978-1-4757-3906-0</isbn>
                <volume_title>Clinical Research in Diabetes and Obesity</volume_title>
                <contributors>
                    <contributor sequence="first" contributor_role="author">
                        <given_name>Andrea</given_name>
                        <surname>Dunaif</surname>
                    </contributor>
                    <contributor sequence="additional" contributor_role="author">
                        <given_name>Carol Beth</given_name>
                        <surname>Book</surname>
                    </contributor>
                </contributors>
                <component_number>Chapter 14</component_number>
                <year media_type="print">1997</year>
                <first_page>249</first_page>
                <last_page>274</last_page>
                <doi type="book_content">10.1007/978-1-4757-3906-0_14</doi>
                <publication_type>full_text</publication_type>
                <article_title>Insulin Resistance in the Polycystic Ovary
Syndrome</article_title>
            </query>
        </body>
    </query_result>
</crossref_result>
```

Ispis 4[26]. Primjer većeg XML zapisa

Na četvrtom ispisu prikazan je opsežniji XML zapis, no iz zapisa se jasno može vidjeti da se radi o XML zapisu o knjizi, preuzetom putem Crossref stranice. Svi navedeni podaci jasno su prikazani i lako čitljivi, a oznake pružaju kvalitetan opis.

Osim XML-a, kao primjer dobre prakse pokazao se i format JSON. Iako stvoren u programskom jeziku JavaScript, dovoljno je jednostavan i prilagođen da ga je moguće koristiti i pomoću drugih programskih jezika te je s vremenom postao standard posve odvojen od JavaScripta. Koristeći se strukturom ključeva i vrijednosti, JSON-om je također moguće opisati razne vrste formata, uključujući i bibliografske. Obzirom da po prirodi podržava hijerarhijske podatke i strukture podataka kao vrijednost, bibliografske podatke je vrlo lako prikazati u JSON formatu.

[26] http://www.crossref.org/openurl/

```
{
  "issued": {"date-parts": [1997]},
  "link": [
    {"intended-application": "crawler-based",
      "content-version": "vor",
      "content-type": "unspecified",
      "URL": "http://www.springerlink.com/index/pdf/10.1007/978-1-4757-3906-0_14"}],
  "score": 1,
  "prefix": "http://id.crossref.org/10.1007",
  "author": [{"family": "Dunaif", "given": "Andrea"},
             {"family": "Book", "given": "Carol Beth"}],
  "container-title": "Clinical Research in Diabetes and Obesity",
  "citedby-count": 4,
  "reference-count": 110,
  "page": "249-274",
  "deposited": {"date-parts": [2019, 5, 10], "timestamp": 1557458365000},
  "created": {
    "date-parts": [2012, 12, 31],
    "issue": "",
    "title": "Insulin Resistance in the Polycystic Ovary Syndrome",
    "subtitle": "",
    "type": "book_content",
    "DOI": "10.1007/978-1-4757-3906-0_14",
    "ISBN": ["9781617370519", "9781475739060"],
    "URL": "https://doi.org/10.1007/978-1-4757-3906-0_14",
    "source": "Crossref",
    "publisher": "Springer Science and Business Media LLC",
    "indexed": {"date-parts": [2019, 5, 9], "timestamp": 1557458365000},
    "volume": "", "member": "http://id.crossref.org/member/297"}}
```


Ispis 5[27]. Primjer JSON zapisa s više podataka

Na ispisu 5. možemo vidjeti primjer JSON zapisa koji se ujedno koristi ugniježđenim podacima. Iz ključeva možemo jasno isčitati o kakvim se podacima radi te kakve vrijednosti očekivati. Jedna od pozitivnih strana JSON-a je što po samoj svojoj prirodi podržava hijerarhijske podatke. Zapis mogu čitati i ljudi i strojevi što ga čini jednim od boljih formata za opisivanje danas. Također jedna od prednosti korištenja JSON formata je mogućnost jednostavnog dohvaćanja podataka korištenjem ključeva. Mnogi se programeri sve češće odlučuju za korištenje JSON-a umjesto XML-a, a neki od razloga su jednostavnost i korištenje struktura podataka bliskih programiranju, koja omogućuje preglednije i čišće zapise[28].

4.3. Python

Modernizacijom informacijskih tehnologija došlo je do sve veće korištenosti iste. Danas gotovo svaka osoba ima osobno računalo i koristi se raznim programima i aplikacijama koje mogu biti jednostavne, ali i kompleksne. Svaku je pojedinu aplikaciju bilo potrebno izraditi, a za njihovu izradu korišteni su programski jezici, odnosno formalna vrsta jezika koja, kao i svaki drugi jezik, ima određenu sintaksu i rječnik[29]. Programski se jezici dijele na strojne i simboličke jezike, mogu biti više i niže razine i objektno orijentirani

[27] http://www.crossref.org/openurl/
[28] Ibid.
[29] Nikola Tesla: Nacionalni portal za učenje na daljinu. URL: https://tesla.carnet.hr/mod/resource/view.php?id=5066 [2019-08-22]

ili procedurno orijentirani[30]. Među najpoznatije i korištenije jezike svrstavamo C, C++, Java, , JavaScript i Python[31].

Python je jedan od novijih programskih jezika, nastao 1991. godine, a stvorio ga je Guido von Rossum[32]. Smatra se jezikom visoke razine i opće namjene, što ga čini podobnim za brzo dolaženje do rezultata i posebno za tzv. skriptno programiranje gdje programi nisu softver pripremljen za diseminaciju već programske skripte koje rješavaju neki lokalan problem poput ekstrakcije i pripreme podataka iz knjižničnog kataloga. Koristi se tzv. „modulima“, odnosno proširenjima mogućnosti programskog jezika koje je moguće uvesti u program po potrebi, a jedan od tih modula je Pymarc, korišten za svrhu ovog istraživanja[33]. Pymarc je modul stvoren za rad s bibliografskim opisima zapisanima u formatu MARC21, a neke od mogućnosti su uređivanje MARC zapisa i konverzija iz MARC formata u drugi format[34]. Neki od formata u koje je moguće konvertirati zapis korištenjem Pymarc-a su XML i JSON, a kako su to suvremeni formati namijenjeni za opis podataka, odabrani su kao formati u koje će se MARC zapisi konvertirati za potrebe ovog istraživanja.

Jedna od zanimljivosti kod službenog opisa Pymarc modula jest da kaže „da iako se često poziva na smrt MARC formata, isti uspješno nastavlja sretan život kao zombi.“[35]

5. IMPLEMENTACIJA

Razvoj informacijskih tehnologija doveo je do pojednostavljivanja dijelova ljudskog života, a time i potrebu i mogućnost dodatnog olakšavanja pojedinih aspekata. Mnogi su knjižničari uvidjeli kako MARC format, iako odrađuje svoju funkciju, nije dostatan za budući opis građe. U članku MARC must die, koji je napisao Roy Tennant, možemo vidjeti mnoge razloge za prelazak na neki suvremeniji format, o čemu je bila riječ u jednom od prethodnih poglavlja[36].

5.1. Cilj i svrha

Svrha ovoga rada je utvrditi i usporediti mogućnosti opisa bibliografske građe koristeći MARC ili XML i JSON formate te predložiti prikladniji format, kao i način konvertiranja iz MARC formata u drugi. Cilj ovog istraživanja je istražiti pozitivne i negativne strane svakog od formata, usporediti način na koji se podaci prikazuju i opisuju i naglasiti važnost korištenja suvremenih formata.

Ciljevi:

1. Istražiti način opisa građe u formatima MARC, XML i JSON

[30] Leksikografski zavod Miroslav Krleža. URL: http://www.enciklopedija.hr/natuknica.aspx?ID=50558 [2019-08-22]
[31] ICT business. URL: https://www.ictbusiness.info/vijesti/top-15-najpopularnijih-programskih-jezika.phtml [2019-08-22]
[32] W3 Schools. URL: https://www.w3schools.com/python/python_intro.asp [2019-08-19]
[33] Python. URL: https://docs.python.org/3/ [2019-08-22]
[34] Pypi. URL: https://pypi.org/project/pymarc/ [2019-08-22]
[35] Ibid.
[36] Tennant, Roy. MARC must die. URL: http://soiscompsfall2007.pbworks.com/f/marc+must+die.pdf [2019-08-20]

2. Programom izrađenim u programskom jeziku Python prikazati pretvorbu MARC zapisa u XML i JSON formate
3. Usporediti zapise MARC, XML i JSON formata i opisati kvalitete i nedostatke

5.2. Metodologija

Za izradu programa korišten je programski jezik Python, uz korištenje modula Pymarc. Modul Pymarc također omogućava stvaranje MARC zapisa stoga je izrađen i poseban program pomoću kojeg se izrađuju MARC zapisi. Izrada programa uključivala je proučavanje dokumentacije o modulu Pymarc i korištenje primjera pronađenih na GitHub-u[37]. Programi su izrađeni u PyCharmu, alatu za uređivanje programskog kôda, specijaliziranim za rad s Python programskim jezikom[38]. Za provedbu ovog istraživanja izrađena su dva programa, program za stvaranje MARC zapisa i program za pretvorbu MARC zapisa u JSON i XML formate. Prvi korak bio je izrada programa pomoću kojeg ćemo stvoriti MARC zapis. Izrađeni program najjednostavnija je moguća verzija programa, izrađena samo za svrhu usporedbe prikaza zapisa te postoji mogućnost daljnje nadogradnje programa.

```
# uvođenje modula pymarc u program
from pymarc import *


# stvori novi marc zapis
record = Record()

# dodaj polje u zapis
record.add_field(
    Field(
        # odabir korištenih oznaka i unos željenih podataka
        tag='245',
        indicators=['0', '1'],
        subfields=[
            'a', 'Technical Services: ',
            'b', 'today and tomorrow/',
            'c', 'Michael Gorman.']))

# zapiši zapis u binarni marc format
writer = MARCWriter(open('./data_gen/file.dat', 'wb'))
writer.write(record)
writer.close()

# zapiši zapis u tekstualni marc format radi vizualno
preglednijeg ispisa rezultata
writer = TextWriter(open('./data_gen/file.txt', 'wt'))
writer.write(record)
writer.close()
```

Ispis 6. Izgled programa za izradu MARC zapisa

[37]GitHub. URL: https://github.com/edsu/pymarc [2019-08-23]
[38] JetBrains. Pycharm Community. URL: https://www.jetbrains.com/pycharm/download/#section=windows [2019-08-23]

Kao prvi korak provedbe implementacije, izrađen je program za izradu MARC zapisa. Na Ispisu 6. prikazan je program pomoću kojeg se definiraju oznake koje se koriste u MARC zapisu te podaci koji pripadaju tim oznakama. Nakon što se zapis izradi sprema se u binarnu i tekstualnu datoteku.

```
import os
import glob
from pymarc import *

path = './data_marc/*.mrc'

# za svaku datoteku s 'utf8' ekstenzijom
for in_path in glob.glob(path):
    print(in_path)

    # dohvati naziv datoteke
    file_name = os.path.split(in_path)[1]
    # odvoji ekstenziju od naziva datoteke
    file_name = os.path.splitext(file_name)[0]

    # pristupi datoteci
    with open(in_path, "rb") as in_file:
        # čitaj marc format ne sirovi tekst
        reader = MARCReader(in_file)
        # usnimi sve zapise iz datoteke u memoriju
        records = list(reader)

    # za svaki zapis u usnimljenim zapisima
    for record in records:
        # zapiši zapis u json format
        out_path = './data_out/' + file_name + '.json'
        with open(out_path, 'w', encoding='utf-8') as out_file:
            writer = JSONWriter(out_file)
            writer.write(record)
            writer.close()

        # zapiši zapis u xml format
        out_path = './data_out/' + file_name + '.xml'
        with open(out_path, 'wb') as out_file:
            writer = XMLWriter(out_file)
            writer.write(record)
            writer.close()
```

Ispis 7. Program za konverziju MARC zapisa u JSON i XML format

Drugi dio implementacije uključio je izradu programa koji će zapis u MARC formatu pretvoriti u zapis u XML i JSON formatu. Kako bi se to provelo, potrebno je definirati koje će se datoteke koristiti i bitno je poznavati koju ekstenziju imaju. Na ispisu 7. prikazana je pretvorba MARC zapisa s ekstenzijom '.mrc'. Nakon što program pronađe sve datoteke sa zadanom ekstenzijom, pokreće čitanje datoteke i snima ih u memoriju. Kako bi sačuvali zapis iz memorije, u programu je definirano u koju datoteku će se zapis pohraniti te pod kojim imenom. Zapis će se pohraniti u formatima XML i JSON. Nije nužno da MARC zapis ima ekstenziju '.mrc', ali je bitno da je stuktura MARC zapisa u MARC standardu, inače neće biti moguće izvršiti konverziju zato što program neće biti u mogućnosti prepoznati „višak“.

5.3. Rezultati

```
00101     2200037   4500245006300000-01aTechnical Services: btoday and tomorrow/cMichael
Gorman.-
```

Ispis 8. Ispis MARC zapisa u binarnu datoteku

Nakon što pokrenemo program možemo vidjeti kako se zapis ispisao. Na ispisu 8. prikazan je ispis pohranjen u binarnu datoteku, ispisan u jednom retku te, iako smo pri unosu podataka u program unijeli separatore, ispisani je tekst gotovo nečitljiv i nerazumljiv, a neki znakovi nisu prepoznati.

```
=LDR            22        4500
=245  01$aTechnical Services: $btoday and tomorrow/$cMichael Gorman.
```

Ispis 9. Ispis MARC zapisa u .txt datoteku

Izrađenim programom omogućeno je i zapisivanje MARC zapisa u običnu tekstualnu datoteku (.txt), a na ispisu 9. vidimo da se na taj način zapis vizualno preglednije formatirao i lakši je za čitanje ljudima.

```
...

 {"100": {"subfields": [{"a": "Orwell, George,"},{"d": "1903-1950."}],
    "ind1": "1","ind2": " "}},
{"240": {"subfields": [{"a": "Nineteen eighty-four"}],
    "ind1": "1","ind2": "0"}},
{"245": {"subfields": [{"a": "1984 :"},{"b": "a novel /"},
      {"c": "by George Orwell ; with a special preface by Walter Cronkite ;
and an afterword by Erich Fromm."}],"ind1": "1", "ind2": "0"}},
{"260": {"subfields": [
      {"a": "New York, N.Y. :"},{"b": "New American Library,"},
      {"c": "[1983] c1977."}],
    "ind1": " ", "ind2": " "}},

...
```

Ispis 10. Ispis zapisa pretvorenog u JSON format

Nakon što program obavi sve što smo mu zadali, možemo vidjeti ispis rezultata. Na ispisu 10. vidljiv je odabrani dio zapisa u JSON formatu ispisan pomoću prethodno izrađenog programa. Podaci su prikazani točnom strukturom JSON formata i veoma jasno možemo vidjeti što koji podatak predstavlja. Podaci su prikazani hijerarhijski, prema odgovarajućim potpoljima. Oznake polja i potpolja korišteni su kao ključevi,

kojima su vrijednosti podaci o publikaciji. Prikazana polja su 100, 240, 245 i 260 koja prikazuju osnovne podatke o publikaciji, a cjeloviti ispis nalazi se u prilozima.

```
...
<datafield ind1="1" ind2=" " tag="100">
    <subfield code="a">Orwell, George,</subfield>
    <subfield code="d">1903-1950.</subfield>
</datafield>
<datafield ind1="1" ind2="0" tag="240">
    <subfield code="a">Nineteen eighty-four</subfield>
</datafield>
<datafield ind1="1" ind2="0" tag="245">
    <subfield code="a">1984 :</subfield>
    <subfield code="b">a novel /</subfield>
    <subfield code="c">by George Orwell ; with a special preface by Walter Cronkite ;
and an afterword by Erich
        Fromm.
    </subfield>
</datafield>
<datafield ind1=" " ind2=" " tag="260">
    <subfield code="a">New York, N.Y. :</subfield>
    <subfield code="b">New American Library,</subfield>
    <subfield code="c">[1983] c1977.</subfield>
...
```

Ispis 11. Ispis XML zapisa

Rezultat pretvorbe MARC zapisa u XML format vidljiv je u ispisu 11. Radi usporedbe prikazan je isti dio rezultata rada programa kao i na ispisu 10. Kao i u JSON zapisu, potpolja su jasno uređena od glavnih oznaka, a svaki od podataka jasno je čitljiv.

Izvršavanje programa za konverziju MARC datoteke preuzete s kataloga NSK rezultira pogreškom i pretvorbu iz MARC formata u XML i JSON nije moguće provesti. Razlog tomu je višak teksta koji se ne koristi MARC oznakama (vidljivo na ispisu 3.) što program ne prepoznaje i ne može provesti zadane funkcije. Ovakav se slučaj dogodio samo pri radu sa zapisima preuzetima iz kataloga NSK, dok su zapisi iz Kongresne knjižnice i knjižnice Filozofskog fakulteta u Zagrebu valjani, a konverzija je provedena s lakoćom.

5.4. Rasprava

Kroz istraživanje je prikazan proces izrade programa za pretvaranje zapisa stvorenih MARC standardom u zapise XML i JSON formata. Rezultati prikazani u prethodnom poglavlju prikazuju da, iako JSON i XML nisu idealni formati za opis podataka, svejedno prednjače nad MARC formatom. MARC format je i dalje funkcionalan i obavlja svoju funkciju, no mnogi njegovi elementi su zastarjeli i njihova se efikasnost

smanjila. Jedan od većih problema prelaska iz MARC formata na drugi format je to što je velik broj zapisa, u gotovo svakom katalogu, koji se nalaze u MARC formatu potrebno konvertirati u drugi format. Prikazom modela programa izrađenog u programskom jeziku Python vidljivo je kako je proces konverzije zapravo jednostavan. Programski jezik Python omogućava korištenje modula nazvanog Pymarc, kojim se u nekoliko linija koda kreira program koji će prebaciti MARC zapise u XML ili JSON. U istraživanju je prikazan ispis cijelog MARC zapisa u drugi format, a program automatski sam detektira MARC oznake i prikazuje ih prema XML, odnosno JSON standardu prikazivanja podataka.

Ispisani zapisi XML-a i JSON-a odgovaraju strukturi, a njima je moguće i manipulirati. Ovakvi su zapisi lakše čitljivi ljudima, a ne gube smisao zato što ih strojevi i dalje razumiju. Umjesto brojčanih oznaka, koriste se riječima, a mogućnost osmišljavanja vlastitih oznaka, prikazuje ove formate boljim rješenjem.

Važno je naglasiti i manjak upotrebljivosti MARC zapisa ako nisu napisani točnom strukturom MARC standarda, što je stvorilo problem pri konverziji MARC zapisa preuzetog iz kataloga NSK. Koristeći se podacima koji nisu dio MARC zapisa i napisani koristeći MARC standard, programu je onemogućena konverzija u XML i JSON zato što nije u mogućnosti prepoznati „višak“, već samo javlja pogrešku.

6. ZAKLJUČAK

Katalogizacija građe veoma je bitan dio poslovanja knjižnica, stoga je bitno da zapisi budu pravilno izrađeni. Katalogizacija građe ponekad i nije jednostavan posao, no svaka se jedinica građe mora obraditi i dobiti svoje mjesto u katalogu. Korištenje specijaliziranih standarda i formata za bibliografski opis, pojednostavnilo je opis građe, kao i distribuciju zapisa među ustanovama s ciljem ujednačavanja zapisa i formata u svijetu. Razvoj tehnologije doveo je do otvaranja brojnih mogućnosti u mnogim segmentima ljudskog života pa je tako i informacijskim ustanovama omogućio bolje poslovanje. Prvi među takvim formatima, nastao je MARC, standard za opis bibliografske građe u strojno čitljivom formatu. Za opis građe se koristi dugi niz godina, no zadnjih se godina otvorila mogućnost korištenja drugih formata, koji iako nisu stvoreni za bibliografski opis, nude više mogućnosti prilagodbe i olakšavaju distribuciju zapisa.

Kao što se iz rada može zaključiti, MARC format je zastario i potrebna je promjena u opisu bibliografskih podataka. Iako i dalje obavlja svoju zadaću, za razliku od suvremenih formata, MARC format se može stvarati i čitati samo uz pomoć specijaliziranih alata, a zapisi nisu čitki ljudima koji se tek prvi put susreću s takvim formatom. Samo nekolicina ljudi s razumijevanjem može pročitati zapis napisan MARC formatom, a u većinu tih ljudi podrazumijevamo katalogizatore koji se svakodnevno susreću s takvim zapisima. Glavni cilj kod stvaranja MARC standarda bio je kataloge na listićima prikazati u obliku koji stroj može čitati, stoga su korištene oznake koje imaju određeno značenje i funkciju, a manje se vodilo računa o tome hoće li ljudi s razumijevanjem moći razumjeti takav zapis.

Mnogi se ljudi koji razmjenjuju podatke danas odlučuju na formate poput JSON-a i XML-a. Neki od glavnih razloga uključuju jednostavnost strukture kod oba formata, olakšanu čitljivost i pohranjivanje podataka pomoću njih te mogućnost jednostavnog stvaranja i otvaranje takvih zapisa bez potrebe za specijaliziranim programom. Ovim je radom cilj bio prikazati prednosti korištenja takvih formata u usporedbi s MARC formatom koji, za današnje informacijske potrebe i mogućnosti knjižnica, ima više negativnih strana. S obzirom na to da postoji velik broj zapisa bibliografskih opisa u MARC formatu, veliki je problem te zapise pretvoriti iz jednog formata u drugi, no danas postoje alati i za to. Kao primjer u ovom radu, uzet je programski jezik Python putem kojeg je, uz pomoć modula Pymarc, moguće izraditi jednostavan program koji bi sve zapise pretvorio iz jednog formata u drugi. Kroz izradu programa za pretvorbu, prikazano je kako je dovoljno osnovno znanje programiranja u programskom jeziku Python da bi se izradio jedan takav program, a zapisi će biti čitljivi i funkcionalni i ljudima i strojevima.

7. LITERATURA

1. Churchill Archive. URL: http://www.churchillarchive.com/subscriber-services/marc-records [2019-08-22]
2. Cloud elements. URL: https://blog.cloud-elements.com/json-better-xml [2019-08-22]
3. Coding networker blog. URL: https://codingnetworker.com/2015/10/python-dictionaries-json-crash-course/ [2019-08-25]
4. Elliote, Harold. XML in a nutshell. URL:https://docstore.mik.ua/orelly/xml/xmlnut/ch01_04.htm [2019-08-17]
5. Geeks for geeks. URL: https://www.geeksforgeeks.org/xml-basics/ [2019-08-22]
6. GitHub. URL: https://github.com/edsu/pymarc [2019-08-23]
7. Gorman, Michael. Technical services: Today and tomorrow. Englewood, Col.: Libraries Unlimited, 1998.
8. ICT business. URL: https://www.ictbusiness.info/vijesti/top-15-najpopularnijih-programskih-jezika.phtml [2019-08-22]
9. Jetbrains. Pycharm: URL: https://www.jetbrains.com/pycharm/download/#section=windows [2019-08-23]
10. JSON. URL: https://www.json.org/ [2019-08-18]
11. Katalog Nacionalne i sveučilišne knjižnice. URL: http://katalog.nsk.hr/F?RN=242120206 [2019-08-20]
12. Katalog sveučilišne knjižnice Sveučilišta u Zadru. URL: http://www.unizd.hr/sveucilisnaknjiznica/katalozi/sveucilisna-knjiznica-zadar [2019-08-23]

13. Leksikografski zavod Miroslav Krleža. MARC. URL: http://www.enciklopedija.hr/natuknica.aspx?ID=38783 [2019-08-16]
14. Leksikografski zavod Miroslav Krleža. Programski jezici. URL: http://www.enciklopedija.hr/natuknica.aspx?ID=50558 [2019-08-22]
15. Libraries linking Idaho. How to Understand a MARC Record. URL: http://lili.org/forlibs/ce/able/course8/19understand1.htm [2019-08-25]
16. Library of Congress. MARC in XML. URL: https://www.loc.gov/marc/marcxml.html [2019-08-18]
17. Microsoft office. XML za početnike. URL: https://support.office.com/hr-hr/article/xml-za-po%C4%8Detnike-a87d234d-4c2e-4409-9cbc-45e4eb857d44#bm6 [2019-08-27]
18. MongoDB. URL: https://www.mongodb.com/ [2019-09-22]
19. MSSQL Tips. URL: https://www.mssqltips.com/sqlservertip/2899/importing-and-processing-data-from-xml-files-into-sql-server-tables/ [2019-08-22]
20. Nikola Tesla: Nacionalni portal za učenje na daljinu. URL: https://tesla.carnet.hr/mod/resource/view.php?id=5066 [2019-08-22]
21. Notepad ++. URL: https://notepad-plus-plus.org/ [2019-08-23]
22. Obdurodon.org. What is XML. URL: http://dh.obdurodon.org/what-is-xml.xhtml [2019-08-17]
23. Open data Handbook. Formati datoteka. URL: http://opendatahandbook.org/guide/hr/appendices/file-formats/ [2019-08-20]
24. Pitts, Natanya. Osnove XML. Upoznavanje XML-a. Čačak: Kompjuter biblioteka, 2000.
25. Pypi. URL: https://pypi.org/project/pymarc/ [2019-08-22]
26. Python. URL: https://docs.python.org/3/ [2019-08-22]
27. Taylor, Arlene G., Joudrey, Daniel N. The organisation of information. London : Libraries Unlimited, 2009.
28. Tennant, Roy. MARC must die. URL: http://soiscompsfall2007.pbworks.com/f/marc+must+die.pdf [2019-08-20]
29. Unimarc skraćeni bibliografski format. Willer, Mirna. Zagreb: Hrvatsko knjižničarsko društvo, 2009.
30. Willer, Mirna. UNIMARC u teoriji i praksi. Što je MARC. Rijeka: Naklada Benja, 1996.
31. W3 Schools. Python introduction. URL: https://www.w3schools.com/python/python_intro.asp [2019-08-19]
32. W3C. URL: https://www.w3.org/XML/ [2019-08-17]

MARC standard and contemporary formats for data exchange

Abstract

This paper talks about MARC standard and contemporary formats for bibliographic description. The purpose of this paper is to establish and compare options of bibliographic description of library material using MARC format, apropos XML and JSON, and to propose a better format, as well as a process of converting MARC record to XML and JSON. The main aim of this paper is to see into advantages and disadvantages of each format, compare means of recording in each format and emphasize the importance and advantages of using modern formats. Particulars of each format are shown through paper as well as an example of recording in each of formats. Disadvantages of MARC format are especially shown, but advantages of modern formats are emphasized. Research included making of program in programming language Python, while using module called Pymarc, whose purpose is to convert MARC record into XML or JSON format. Research showed that modern formats have an advantage in front of MARC standard which is a bit outdated. Main advantages of modern formats are that they are easy to read and write. The process of converting MARC records into modern formats is not complicated, as it was discovered by creating a conversion program. It is sufficient to have basic knowledge about programming to create such a program and it will serve to its purpose.

Keywords: MARC, XML, JSON, contemporary formats

8. POPIS SLIKA

Slika 1. Primjer MARC zapisa iz kataloga Nacionalne i sveučilišne knjižnice

Slika 2 . Primjer prikaza osnovnih podataka u MARC zapisu

Slika 3. Primjer MARC zapisa koji se prikazuje korisnicima

Slika 4. MARC zapis kako ga knjižničari vide

Slika 5. Primjer MARC zapisa u uređivaču teksta

9. POPIS ISPISA

Ispis 1. Primjer XML zapisa

Ispis 2. Primjer JSON zapisa

Ispis 3. MARC zapis preuzet iz kataloga NSK

Ispis 4. Primjer većeg XML zapisa

Ispis 5. Primjer JSON zapisa s više podataka

Ispis 6. Izgled programa za izradu MARC zapisa

Ispis 7. Program za konverziju MARC zapisa u JSON i XML format

Ispis 8. Ispis MARC zapisa u binarnu datoteku

Ispis 9. Ispis MARC zapisa u tekstualnu datoteku

Ispis 10. Ispis zapisa pretvorenog u JSON format

Ispis 11. Ispis XML zapisa

10. PRILOZI

Prilog 1. Cjeloviti rezultat ispisa zapisa u JSON format

```
[{"leader": "01040cam a2200301 a 4500",
    "fields": [{"001": "3303902"}, {"005": "20040721135337.0"},
      {"008": "840202t19831977nyu      b    000 1 eng  "},
      {"035": {"subfields": [{"9": "(DLC)   83165326"}],
          "ind1": " ", "ind2": " "}},
      {"906": {"subfields": [{"a": "7"}, {"b": "cbc"}, {"c": "orignew"},
            {"d": "2"}, {"e": "ncip"}, {"f": "19"}, {"g": "y-gencatlg"}],
          "ind1": " ", "ind2": " "}},
      {"010": {"subfields": [{"a": "   83165326 "}], "ind1": " ", "ind2": " "}},
      {"020": {"subfields": [{"a": "0452254264 (pbk.) :"}, {"c": "$5.95"}], "ind1": " ", "ind2":
" "}},
      {"040": {"subfields": [{"a": "DLC"}, {"c": "DLC"}, {"d": "DLC"}], "ind1": " ", "ind2": "
"}},
      {"050": {"subfields": [{"a": "PR6029.R8"}, {"b": "N49 1983"}], "ind1": "0", "ind2": "0"}},
      {"082": {"subfields": [{"a": "823/.912"}, {"2": "19"}], "ind1": "0", "ind2": "0"}},
      {"100": {"subfields": [{"a": "Orwell, George,"}, {"d": "1903-1950."}], "ind1": "1",
"ind2": " "}},
      {"240": {"subfields": [{"a": "Nineteen eighty-four"}], "ind1": "1", "ind2": "0"}},
      {"245": {"subfields": [{"a": "1984 :"}, {"b": "a novel /"},
        {"c": "by George Orwell ; with a special preface by Walter Cronkite ; and an afterword
by Erich Fromm."}],
        "ind1": "1", "ind2": "0"}},
      {"260": {"subfields": [{"a": "New York, N.Y. :"}, {"b": "New American Library,"},
        {"c": "[1983] c1977."}], "ind1": " ", "ind2": " "}},
      {"300": {"subfields": [{"a": "267 p. ;"}, {"c": "21 cm."}], "ind1": " ", "ind2": " "}},
      {"504": {"subfields": [{"a": "Includes bibliography."}], "ind1": " ", "ind2": " "}},
      {"651": {"subfields": [{"a": "London (England)"}, {"v": "Fiction."}], "ind1": " ", "ind2":
"0"}},
      {"650": {"subfields": [{"a": "Totalitarianism"}, {"x": "Fiction."}], "ind1": " ", "ind2":
"0"}},
      {"655": {"subfields": [{"a": "Political fiction."}, {"2": "lcsh"}], "ind1": " ", "ind2":
"7"}},
      {"655": {"subfields": [{"a": "Science fiction."}, {"2": "gsafd"}], "ind1": " ", "ind2":
"7"}},
      {"655": {"subfields": [{"a": "Dystopias."}, {"2": "gsafd"}], "ind1": " ", "ind2": "7"}},
      {"740": {"subfields": [{"a": "Nineteen eighty-four."}], "ind1": "0", "ind2": " "}},
      {"991": {"subfields": [{"b": "c-GenColl"}, {"h": "PR6029.R8"}, {"i": "N49 1983"},
            {"t": "Copy 1"}, {"w": "BOOKS"}], "ind1": " ", "ind2": " "}]}]
```


Prilog 2. Cjeloviti rezultat ispisa zapisa u XML format

```
<?xml version="1.0" encoding="UTF-8"?>
<collection xmlns="http://www.loc.gov/MARC21/slim">
    <record>
        <leader>01040cam a2200301 a 4500</leader>
        <controlfield tag="001">3303902</controlfield>
        <controlfield tag="005">20040721135337.0</controlfield>
        <controlfield tag="008">840202t19831977nyu b 000 1 eng</controlfield>
        <datafield ind1=" " ind2=" " tag="035"> <subfield code="9">(DLC) 83165326</subfield> </datafield>
        <datafield ind1=" " ind2=" " tag="906">
            <subfield code="a">7</subfield>
            <subfield code="b">cbc</subfield>
            <subfield code="c">orignew</subfield>
            <subfield code="d">2</subfield>
            <subfield code="e">ncip</subfield>
            <subfield code="f">19</subfield>
            <subfield code="g">y-gencatlg</subfield></datafield>
        <datafield ind1=" " ind2=" " tag="010"><subfield code="a">83165326</subfield></datafield>
        <datafield ind1=" " ind2=" " tag="020">
            <subfield code="a">0452254264 (pbk.) :</subfield>
            <subfield code="c">$5.95</subfield></datafield>
        <datafield ind1=" " ind2=" " tag="040">
            <subfield code="a">DLC</subfield>
            <subfield code="c">DLC</subfield>
            <subfield code="d">DLC</subfield></datafield>
        <datafield ind1="0" ind2="0" tag="050">
            <subfield code="a">PR6029.R8</subfield>
            <subfield code="b">N49 1983</subfield></datafield>
        <datafield ind1="0" ind2="0" tag="082">
            <subfield code="a">823/.912</subfield>
            <subfield code="2">19</subfield></datafield>
        <datafield ind1="1" ind2=" " tag="100">
            <subfield code="a">Orwell, George,</subfield>
            <subfield code="d">1903-1950.</subfield></datafield>
        <datafield ind1="1" ind2="0" tag="240"> <subfield code="a">Nineteen eighty-
four</subfield></datafield>
        <datafield ind1="1" ind2="0" tag="245">
            <subfield code="a">1984 :</subfield>
            <subfield code="b">a novel /</subfield>
            <subfield code="c">by George Orwell ; with a special preface by Walter Cronkite ; and an
afterword by Erich Fromm.
            </subfield></datafield>
        <datafield ind1=" " ind2=" " tag="260">
            <subfield code="a">New York, N.Y. :</subfield>
            <subfield code="b">New American Library,</subfield>
            <subfield code="c">[1983] c1977.</subfield></datafield>
        <datafield ind1=" " ind2=" " tag="300">
            <subfield code="a">267 p. ;</subfield>
            <subfield code="c">21 cm.</subfield></datafield>
        <datafield ind1=" " ind2=" " tag="504">
            <subfield code="a">Includes bibliography.</subfield></datafield>
        <datafield ind1=" " ind2="0" tag="651">
            <subfield code="a">London (England)</subfield>
            <subfield code="v">Fiction.</subfield></datafield>
        <datafield ind1=" " ind2="0" tag="650">
            <subfield code="a">Totalitarianism</subfield>
            <subfield code="x">Fiction.</subfield></datafield>
        <datafield ind1=" " ind2="7" tag="655">
            <subfield code="a">Political fiction.</subfield>
            <subfield code="2">lcsh</subfield></datafield>
        <datafield ind1=" " ind2="7" tag="655">
            <subfield code="a">Science fiction.</subfield>
            <subfield code="2">gsafd</subfield></datafield>
        <datafield ind1=" " ind2="7" tag="655">
            <subfield code="a">Dystopias.</subfield>
            <subfield code="2">gsafd</subfield></datafield>
        <datafield ind1="0" ind2=" " tag="740"><subfield code="a">Nineteen eighty-
four.</subfield></datafield>
        <datafield ind1=" " ind2=" " tag="991">
            <subfield code="b">c-GenColl</subfield>
            <subfield code="h">PR6029.R8</subfield>
            <subfield code="i">N49 1983</subfield>
            <subfield code="t">Copy 1</subfield>
            <subfield code="w">BOOKS</subfield></datafield>
    </record>
</collection>
```