

KONVERZIJA KATALOGOV NACIONALNE KNJIŽNICE BOLGARIJE V FORMAT COMARC

**Nadja Dimitrova
Karačodžukova**

Narodna knjižnica Sv. Cirila in
Metoda, Sofija

Simona Tušek

Institut informacijskih znanosti
Maribor

Kontakti naslov:
nadia_dk@nationallibrary.bg
simona.tusek@izum.si

Izvleček

V prispevku poročamo o konverziji katalogov Nacionalne knjižnice Bolgarije (NBKM) v format COMARC/B, COMARC/A in COMARC/H. Konverzije baz podatkov so bile izvedene v IZUM-u v sodelovanju s kolegi iz Nacionalne knjižnice. Predstavljeni so rezultati, ki smo jih dosegli po približno dveh letih skupnega dela. V tem času smo konvertirali 7 bibliografskih baz podatkov in 2 normativni bazi podatkov. S konverzijo v formate COMARC smo ustvarili lokalno bazo podatkov NBKM in vzajemno bazo podatkov COBIB.BG z 865.066 bibliografskimi in normativnimi zapisi.

Ključne besede

preslikava formatov, formati COMARC, format NBKM/ISIS, konverzija zapisov

Abstract

This paper addresses the conversion of the catalogues of the National Library of Bulgaria (NBKM) into COMARC/B, COMARC/A and COMARC/H formats. The databases were converted by IZUM, in cooperation with NBKM. The results of two years of cooperation are presented. In this time seven bibliographic databases and two authority databases were converted. With the conversion into the COMARC formats, the NBKM local database and the COBIB.BG shared database with 865,066 bibliographic and authority records were created.

Keywords

format mapping, COMARC formats, NBKM/ISIS format, conversion of records

UVOD

Nacionalna knjižnica sv. Cirila in Metoda (St. St. Cyril and Methodius National Library) ima bibliografske podatke razdeljene v več ločenih baz podatkov (skupaj 740.514 bibliografskih zapisov). Poleg bibliografskih zapisov kreirajo tudi normativne zapise za osebna imena in predmetne oznake (skupaj 124.033 normativnih zapisov). Podatke imajo v svojem lokalnem formatu – NBKM/ISIS, ki jih je bilo treba pretvoriti v format COMARC/B in COMARC/A ter vpisati v vzajemno in lokalno bazo podatkov. Nacionalna knjižnica Bolgarije ima v svoji zalogi okoli milijon računalniško obdelanih enot gradiva. Tudi podatke o zalogi je bilo treba s konverzijo pretvoriti v format COMARC/H.

NBKM – LOKALNI KNJIŽNIČNI SISTEM

Ideja o vključitvi Nacionalne knjižnice Bolgarije v sistem COBISS in dodatnih šest knjižnic v prvi fazi se je porodila pred petnajstimi leti v bolgarski knjižnični skupnosti.

V tem času se je skupina knjižnic odločila, da bo sodelovala v skupnem knjižničnem sistemu. V letu 1993 je bil zasnovan projekt za Nacionalno avtomatizirano knjižnično-informacijsko mrežo (NABIM), ki ga je finančno podprla fundacija Odrpta družba [1].

Avtomatizirana je bila Nacionalna bibliografija, ki je vir bibliografskih zapisov v strojno berljivem formatu za vso bolgarsko tiskano produkcijo. Knjižnični procesi se avtomatizirajo v mnogih bolgarskih knjižnicah, vzpostavljene pa so tudi komunikacijske povezave med knjižnicami za prenos bibliografskih podatkov. Koordinacija med posameznimi dejavnostmi v knjižnicah se vzpostavlja na nove načine s projekti dopolnjevanja ali uvajanja novih informacijskih tehnologij.

Z vključitvijo v sistem COBISS je Bolgarija dobila priložnost, da to idejo uresniči.

Začetek sega v november 2006, ko je profesorica dr. Borjana Hristova podpisala sporazum o prostem pretoku bibliografskih zapisov v mreži COBISS.Net.

Glavni del zapisov bodo v skupnem katalogu COBISS. BG predstavljali zapisi Nacionalne knjižnice. To je knjižnica z najstarejšo tradicijo in najbolj bogato zbirko dokumentov v Bolgariji, ki zbira, obdeluje, hrani in posreduje temeljno nacionalno zbirko vsega knjižničnega gradiva v bolgarskem jeziku. Njene glavne naloge so [2]:

- evidentiranje nacionalne tiskane produkcije in vse druge dokumentacije skladno z Zakonom o obveznem izvodu od leta 1897, zaradi česar je knjižnica glavni center za zagotavljanje informacij mnogim avtomatiziranim knjižnicam v Bolgariji, ki posredujejo te informacije svojim uporabnikom v obliki baz podatkov;
- varovanje (arhivske) zbirke bolgarske književnosti;
- zbiranje tuje literature z nakupi ali zameno knjig z več kot 2.000 knjižnicami po vsem svetu;
- zbiranje, obdelava in varovanje zbirk rokopisov, arhivskega gradiva, zemljevidov, slik in grafik, zvočnih zapisov, partitur itd.;
- zagotavljanje službe za dostavo dokumentov, medknjižnične izposoje na lokalni in mednarodni ravni;
- zagotavljanje širokega spektra referenčnih in informacijskih storitev;
- v prihodnosti zagotavljanje podobnih storitev, kot jih izvajata centra ISBN in ISSN v Bolgariji za glasbeni tisk in elektronske vire;
- koordiniranje dejavnosti na področju knjižnične statistike, standardov in vseh drugih dejavnosti, povezanih s knjižnicami, kot raziskovalni in svetovalni center na področju knjižničarstva.

Bibliografske baze podatkov v NBKM

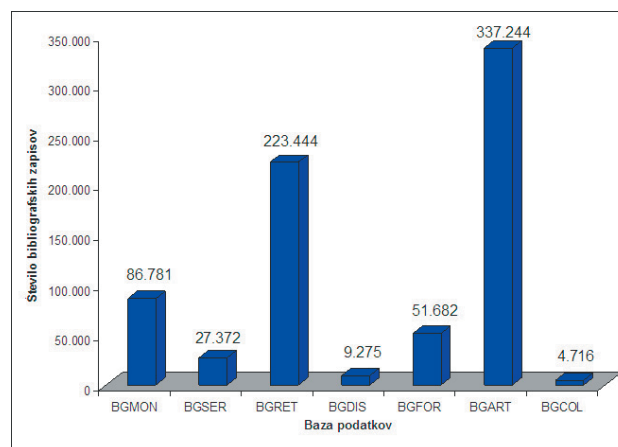
V Nacionalni knjižnici Bolgarije so bibliografski podatki razdeljeni v več posameznih baz podatkov.

Pri konverziji smo obravnavali sedem baz podatkov, in sicer:

- Bolgarske knjige iz obdobja 1992–2006 (**BGMON**),
- Kontinuirani viri (**BGSER**),
- Bolgarske knjige – retrospektiva iz obdobja 1290–1991 (**BGRET**),
- Bolgarske knjige za leto 2007 in Bolgarske disertacije iz obdobja 2006–2007 (**BGDIS**),
- Tuje knjige iz obdobja 1992–2006 (**BGFOR**),
- Članki (**BGART**),
- Zbirke (**BGCOL**).

Zaradi razlik v strukturi zapisov ni bilo mogoče teh baz podatkov konvertirati naenkrat. Konvertirali smo jih eno za drugo.

Graf v sliki 1 prikazuje število bibliografskih zapisov v posamezni bazi podatkov pred konverzijo. Vidimo, da je največje število zapisov v bazi člankov in v retrospektivni bazi bolgarskih knjig.



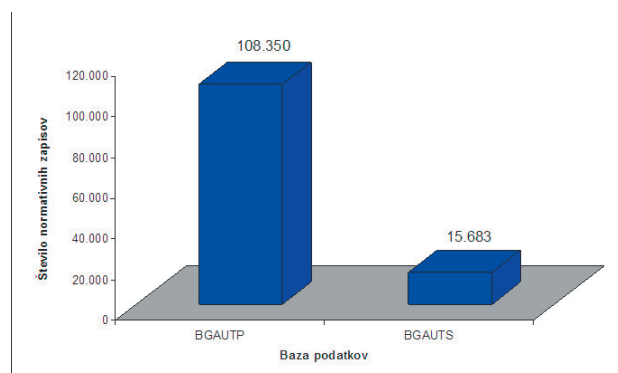
Slika 1: Število bibliografskih zapisov v posameznih bazah podatkov

Normativne baze podatkov v NBKM

Normativni zapisi so v Nacionalni knjižnici razdeljeni v dve bazi podatkov, in sicer:

- Normativno bazo za osebna imena (**BGAUTP**),
- Normativno bazo za predmetne oznake (**BGAUTS**).

Graf na sliki 2 prikazuje število normativnih zapisov v posamezni bazi podatkov pred konverzijo.



Slika 2: Število normativnih zapisov v bazah podatkov

Ti dve bazi podatkov nista povezani z bibliografskimi bazami podatkov.

Normativni zapisi v bolgarski bazi podatkov za osebna imena vsebujejo eno enotno značnico, dve enotni značnici ali tri enotne značnice, in sicer v cirilici, v latinici ali v obeh pisavah.

Primer:

Достоевский, Федор Михайлович,
1821-1881
= Dostoevskij, Fedor Mihajlovic,
1821-1881
= Достоевски, Фьодор Михайлович,
1821-1881

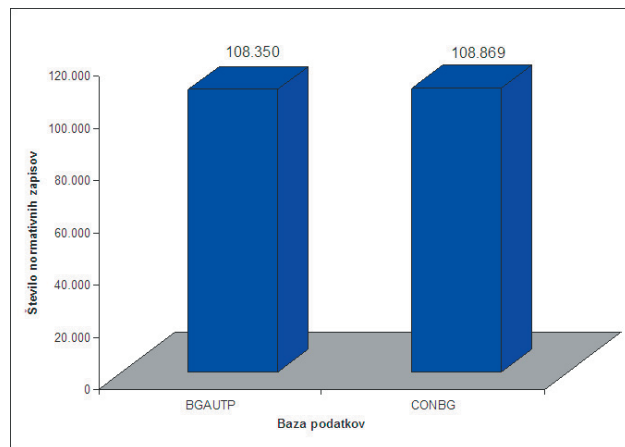
Primer prikazuje nestandardno obliko normativnega zapisa s tremi značnicami, od katerih sta dve v cirilici, kar je posledica tretjega kataloga, to je kataloga knjig v bolgarščini.

Pri konverziji značnic za osebna imena smo našli rešitev v povezanih zapisih, in sicer:

- prvi zapis, ki vsebuje paralelno originalno cirilično obliko imena in transliterirano obliko, npr.:
Достоевский, Федор Михайлович, 1821-1881
= Dostoevskij, Fedor Mihajlovic, 1821-1881
- drugi zapis, ki vsebuje cirilično različico, namenjeno katalogu v bolgarščini, npr.:
Достоевски, Фьодор Михайлович, 1821-1881

Oba zapisa sta med seboj povezana in ustvarjata mrežo podatkov.

Slika 3 prikazuje število zapisov pred konverzijo normativne baze za osebna imena in po konverziji. Število zapisov je po konverziji naraslo za približno 500 zapisov.



Slika 3: Število zapisov v bazi podatkov za osebna imena pred konverzijo in po njej

Normativni zapisi v bolgarski bazi podatkov za predmetne oznake vsebujejo 5 polj za različne vrste napotil. S konverzijo so se ta polja združila v dve polji, v katerih se vrste napotil razlikujejo po kodah.

PRIPRAVA NA KONVERZIJO

Priprava na konverzijo baz podatkov je v Bolgariji potekala v več korakih.

Najprej smo izbrali baze podatkov za konverzijo, nato smo začeli prevajati osnovne dokumente o sistemu, in sicer:

- priročnike COMARC/B, COMARC/A, COMARC/H ter
- priročnike za uporabo programske opreme COBISS3.

Sledilo je pošiljanje izvlečkov zapisov iz baz podatkov za analizo v IZUM-u in redakcija zapisov v bazah podatkov.

AKTIVNOSTI V POSTOPKU KONVERZIJE

Prvi korak skupnega dela smo začeli v Nacionalni knjižnici s pripravo predloga za preslikavo podatkov iz formata ISIS v format COMARC/B. Zapisi v bolgarskih bazah podatkov so kreirani v lokalno razvitem formatu ISIS, ki se bistveno razlikuje od formata COMARC in tudi od drugih formatov MARC. Glavne razlike so prikazane v tabeli 1.

	NBKM/ISIS	COMARC/B
Bloki	polja niso razporejena v bloke	polja so razporejena v bloke in se vrstijo po naraščajočem vrstnem redu
Oznaka polja	sestavljena je iz enega, dveh ali treh znakov	sestavljena je iz treh znakov
Podpolja	le določena polja vsebujejo podpolja	vsa polja vsebujejo podpolja
Indikatorji	ne obstajajo (izpise krmilijo interna polja)	obstajajo (določajo način izpisovanja)
Ponovljivost polj/podpolj	polja so ponovljiva, podpolja so neponovljiva	polja/podpolja so ponovljiva ali neponovljiva
Vnos podatkov	prost, programsko nekontroliran vnos	prost/kodiran, programsko kontroliran
Struktura polja	v različnih bazah različna struktura	enoten namen

Tabela 1: Razlike med formatoma NBKM/ISIS in COMARC/B

V formatu XML so v NBKM pripravili za konverzijo baze podatkov po posameznih letih in jih poslali v IZUM. Slika 4 prikazuje bibliografski zapis iz baze podatkov Bolgarske knjige v formatu XML.

```

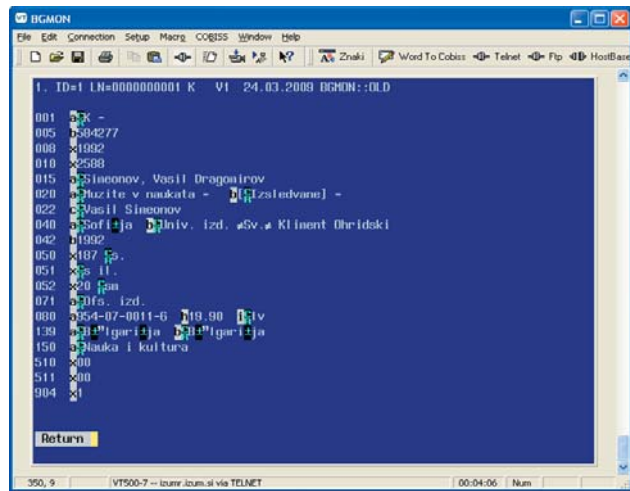
<?xml version="1.0" encoding="UTF-8" ?>
<DATABASE_S1>
  <RECORD>
    <Mfn>1</Mfn>
    <Tag_1>
      <Tag_1_a>K</Tag_1_a>
    </Tag_1>
    <Tag_5>
      <Tag_5_b>584277</Tag_5_b>
    </Tag_5>
    <Tag_8>1992</Tag_8>
    <Tag_10>1</Tag_10>
    <Tag_15>
      <Tag_15_a>Симеонов, Васил Драгомиров</Tag_15_a>
    </Tag_15>
    <Tag_20>
      <Tag_20_a>Музите в науката</Tag_20_a>
      <Tag_20_b>[Изследване]</Tag_20_b>
    </Tag_20>
    <Tag_22>
      <Tag_22_c>Васил Симеонов</Tag_22_c>
    </Tag_22>
    <Tag_40>
      <Tag_40_a>София</Tag_40_a>
      <Tag_40_b>Унив. изд. <Св.> Климент Охридски</Tag_40_b>
    </Tag_40>
    <Tag_42>
      <Tag_42_b>1992</Tag_42_b>
    </Tag_42>
    <Tag_50>187 с.</Tag_50>
    <Tag_51>с ил.</Tag_51>
    <Tag_52>20 см</Tag_52>
    <Tag_71>
      <Tag_71_a>Офс. изд</Tag_71_a>
    </Tag_71>
    <Tag_73>
      <Tag_73_a>4 065</Tag_73_a>
    </Tag_73>
    <Tag_80>
      <Tag_80_a>954-07-0011-6</Tag_80_a>
      <Tag_80_h>19.90</Tag_80_h>
      <Tag_80_i>лв</Tag_80_i>
    </Tag_80>
    <Tag_150>
      <Tag_150_a>Наука и култура</Tag_150_a>
    </Tag_150>
    <Tag_505>0</Tag_505>
    <Tag_510>00</Tag_510>
    <Tag_511>00</Tag_511>
  </RECORD>

```

Slika 4: Bibliografski zapis v formatu XML

Korenski element je DATABASE_S1, element za posamezni zapis je RECORD, element za polje ima oznako Tag_xxx (xxx – eno-do tri-številčna oznaka polja), element za podpolje pa Tag_xxx_y (y – oznaka podpolja). V vsakem zapisu je tudi element Mfn, ki predstavlja zaporedno številko zapisa pri izvozu (ni številka zapisa) [3].

Za lažjo analizo baz so bili podatki preneseni v COBISS2/Katalogizacija v izvirnem formatu. Slika 5 prikazuje isti bibliografski zapis kot prej v formatu XML v programski opremi COBISS2/Katalogizacija. V zapise je bila dodana pred vsebino vsakega podpolja tudi pisava (latinica, cirilica), razen pred podpolja, ki so vsebovala le številke.

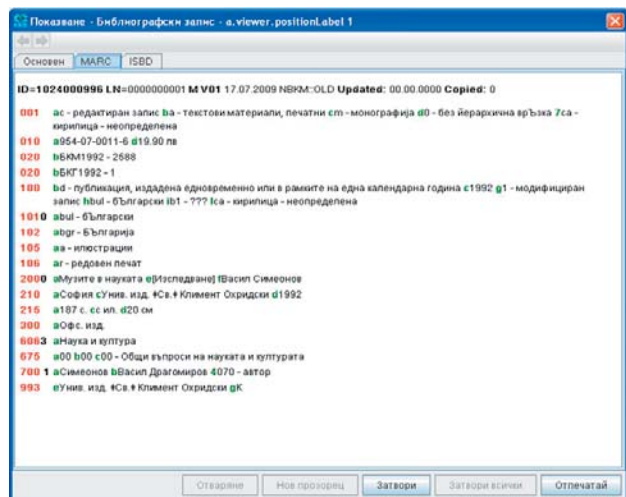


Slika 5: Bibliografski zapis v segmentu COBISS2/Katalogizacija

V IZUM-u smo izvedli natančno analizo vseh poslanih baz podatkov in algoritmov za konverzijo polj in podpolj ter pregledali specifikacije za preslikavo.

V IZUM-u smo pripravili tudi konverzijske tabele iz formata COMARC/B v format ISIS in tabele za pretvorbo prostega besedila v kodirane podatke. Napisali oz. dopolnili smo tudi dodatne algoritme, ki smo jih potrebovali za konverzijo.

Nato smo testno konvertirali baze podatkov v formate COMARC. Slika 6 prikazuje že konvertirani bibliografski zapis v programski opremi COBISS3. Zapise je bilo možno iskati in pregledovati v razvojno verzijo programske opreme COBISS3.

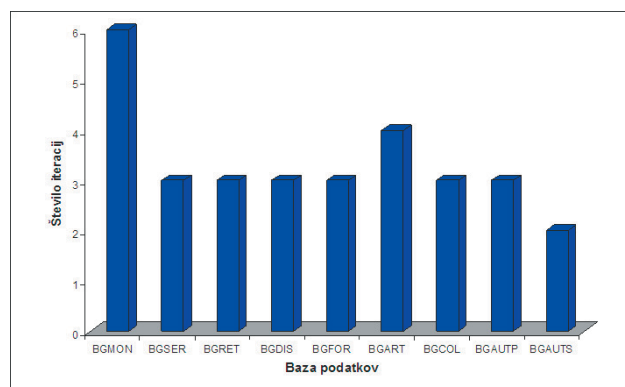


Slika 6: Bibliografski zapis v segmentu COBISS3/Katalogizacija

Vsaka konvertirana baza podatkov je bila testirana najprej v IZUM-u, nato pa še v Bolgariji. Pri testiranju smo preverjali:

- kako popolni so preneseni podatki,
- ali so pravilno razčlenjeni in uvrščeni znotraj formatov COMARC,
- ali so pravilno združeni znotraj formatov COMARC,
- ali so pravilno programsko dodani,
- ali so besedila pravilno pretvorjena v kodirane podatke,
- ali so maske vnosa pravilno izbrane,
- kako je oblikovano sistemsko polje,
- ali so kombinacije podatkov glede na programske kontrole pravilne in
- ali so podatki preneseni v ustrezni nabor znakov COBISS.

Slika 7 prikazuje število iteracij testiranja za posamezno bibliografsko in normativno bazo podatkov. Večino baz smo testirali v treh različnih iteracijah.



Slika 7: Število iteracij v postopku testiranja

Na koncu smo popravke in dopolnitve upoštevali in vključili v postopek konverzije.

Težave pri konverziji

Pri konvertiranju bibliografskih baz podatkov Nacionalne knjižnice Bolgarije smo se soočili s precejšnim številom težav, ki so se pojavile zaradi:

- **Razlik pri definiranju ponovljivosti polj.** V formatu ISIS ponovljivost polj in podpolj ni natančno definirana, saj obstajajo tudi polja, ki so hkrati ponovljiva in neponovljiva.
- **Razlik v predstavitvi intelektualne odgovornosti.** V formatu ISIS namreč obstajajo le tri polja za vnos imen – osebnih in korporativnih. V ta polja so bile vnesene:
 - enotne značnice primarne odgovornosti za osebna imena in korporacije;
 - enotne značnice alternativne in sekundarne

odgovornosti za osebna imena, enotne značnice osebnih imen kot predmetne oznake ter variantne oblike vseh omenjenih enotnih značnic;

- enotne značnice alternativne in sekundarne odgovornosti za korporacije ter njihove variantne oblike.

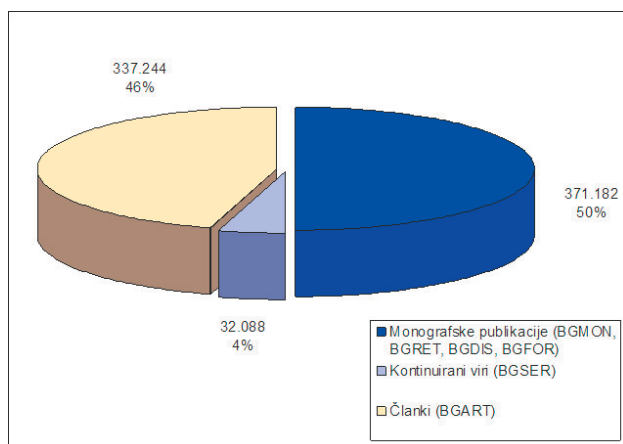
Tudi struktura značnic bistveno odstopa od standarda. V bolgarskih zapisih polja za značnice nimajo podpolj, s katerimi bi bilo mogoče identificirati dele značnice. Zato je bilo treba z različnimi manipulacijami iz 3 polj s 6 podpolji ustvariti 16 polj s skupaj 67 podpolji.

- **Nestandardnih vsebin podatkov.** Zaradi nestandardnih vsebin podatkov za registre tiskarjev in založnikov v Nacionalni bibliografiji smo morali pri konverziji dveh popolnoma nestandardnih polj, za katere ustrezne utemeljitve polj ni bilo mogoče najti niti v standardih ISBD, določiti novo polje na lokalnem nivoju.
- **Odsotnosti kodiranih podatkov.** Bolgarski zapisi ne vsebujejo kodiranih podatkov, z izjemo kode za vrsto dokumenta, zaradi česar smo razlike v zvezi s kodiranimi podatki rešili s pripravo tabel za pretvorbo prostega besedila v kodirane podatke.
- **Sistemskih napak.** V času konverzije smo posegali tudi v besedilne podatke in popravili sistematsko ustvarjene napake, da bi v določeni meri prilagodili besedilo izpisnim pravilom v bolgarskem sistemu.
- **Neenotne uporabe katalogizacijskih pravil.** Težave nam je povzročil predvsem neenoten način vnosa podatkov variantnih značnic v različnih časovnih obdobjih. V bazi bolgarskih monografij sta obstajala za isti postopek dva načina vnosa in s tem tudi dva različna algoritma. Po natančni analizi velikega števila zapisov smo se odločili za algoritem, ki je zajel večinski delež števila zapisov.

Zapise, ki se niso pravilno konvertirali, bo treba popraviti ročno po zaključeni konverziji.

Konvertirani bibliografski zapisi

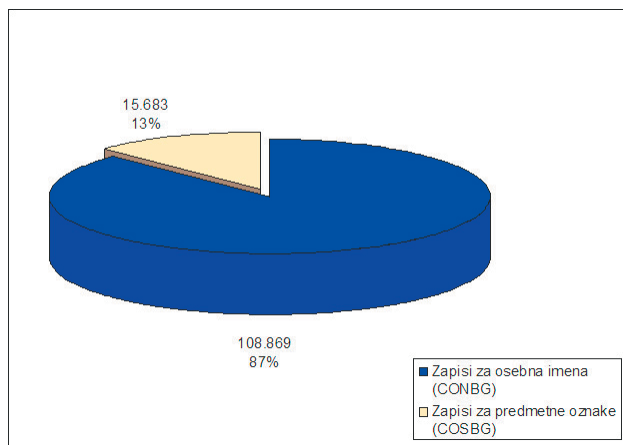
V lokalno in vzajemno bazo podatkov je bilo konvertiranih skupaj 740.514 bibliografskih zapisov. Glede na tip gradiva je bilo med njimi 50 odstotkov zapisov za monografske publikacije, 46 odstotkov zapisov za članke in 4 odstotki zapisov za kontinuirane vire (slika 8).



Slika 8: Število konvertiranih bibliografskih zapisov

Konvertirani normativni zapisi

V normativnih bazah podatkov smo konvertirali skupaj 124.552 normativnih zapisov, med katerimi je bilo 87 odstotkov zapisov za osebna imena in 13 odstotkov zapisov za predmetne oznake (slika 9).



Slika 9: Število konvertiranih normativnih zapisov

Konvertirani zapisi s polji za zalogo

Po izvedeni konverziji bibliografskih podatkov je začela potekati konverzija podatkov o zalogi.

Doslej smo v zapise za bolgarske in tuje knjige ter doktorske disertacije dodali 373.267 polj za zalogo. Nadaljujemo s postopkom dodajanja polj z zalogo pri zapisih za kontinuirane vire.

Pri konverziji podatkov o zalogi smo naleteli na težave pri konverziji signatur. V Nacionalni knjižnici Bolgarije vnašamo signature v dve polji, ob tem pa ni pomembno, kje so bili podatki vneseni, temveč, kako so bili podatki izpisani v tiskanem katalogu in OPAC-u. Že tako zapleten algoritem za konverzijo je še dodatno otežil neenoten vnos.

Elementi signature so bili pri konverziji ločeni. Podatki iz devetih polj in štirinajstih podpolj so bili v formatu COMARC/H konvertirani v eno polje in sedem podpolj.

ZAKLJUČEK

S konverzijo v formate COMARC smo ustvarili lokalno bazo podatkov Nacionalne knjižnice Bolgarije in vzajemno bazo podatkov COBIB.BG z 865.066 bibliografskimi in normativnimi zapisi.

Postopek testnih konverzij za vse bibliografske in normativne baze podatkov je trajal leto in pol. Decembra 2007 smo določili, da bo baza podatkov Bolgarske knjige iz obdobja 1992–2006 prva baza, ki bo konvertirana v format COMARC/B. Avgusta 2008 je bila izvedena prva iteracija testiranja konverzije te baze podatkov v Izumu, decembra 2008 pa še v Bolgariji. Julija 2009 smo zaključili testne konverzije sedmih bibliografskih in dveh normativnih baz podatkov.

Vzpostavitev vzajemne baze podatkov COBIB.BG daje priložnost preostalim bolgarskim knjižnicam, da se priključijo in skupno sooblikujejo nacionalni avtomatizirani knjižnični sistem COBISS.BG.

Reference

- [1] Karačodžukova, N. (2009). Bolgarske knjižnice od srednjega veka do danes. Organizacija znanja, Letn. 14, Zv. 3, 58–61.
- [2] Vasileva, M. (2005). Kratko poročilo o stanju avtomatizacije v nacionalni knjižnici "Sv. Ciril in Metod", Sofija. Organizacija znanja, Letn. 10, Zv. 4, 246–247.
- [3] Budimir, G. (2008). Postopek konverzije baze bolgarskih knjig 1992–2006. (interni dokument), 4 str.
- [4] COMARC/B Format za bibliografske podatke [Elektronski vir]: priročnik za uporabnike. Maribor: IZUM, 2008. Dostopno na: http://izobrazevanje.izum.si/EntryFormDesktopDefault.aspx?tabid=38&type=manual&manual=1_Comarc_B_svn (21. 12. 2009, dostop z uporabniškim imenom in geslom).
- [5] COMARC/A Format za normativne podatke [Elektronski vir]: priročnik za uporabnike. Maribor: IZUM, 2008. Dostopno na: http://izobrazevanje.izum.si/EntryFormDesktopDefault.aspx?tabid=38&type=manual&manual=Comarc_A_svn.
- [6] COMARC/H Format za podatke o stanju zaloge [Elektronski vir]: priročnik za uporabnike. Maribor: IZUM, 2008. Dostopno na: http://izobrazevanje.izum.si/EntryFormDesktopDefault.aspx?tabid=38&type=manual&manual=Comarc_H_svn.