

BIBLIOGRAFSKI ZAPISI IZ COBIB.SI V WORLDCAT

Gordana Budimir

Institut informacijskih znanosti
Maribor

Kontaktni naslov:
gordana.budimir@izum.si

Izvleček

Prispevek opisuje vključevanje bibliografskih zapisov iz vzajemne baze podatkov COBIB.SI in podatkov o zalogi nekaterih slovenskih knjižnic v svetovni katalog WorldCat. V letu 2009 je bil izveden inicialni prenos zapisov in nato redni četrtletni prenosi. Zaradi nekaterih pomanjkljivosti v bibliografskih zapisih v slovenski vzajemni bibliografski bazi COBIB.SI se v WorldCat še niso prenesli vsi zapisi. Na koncu prispevka so podane statistike vključenih zapisov v letu 2009 po različnih kriterijih.

Ključne besede

COBISS.SI, WorldCat, preslikava formatov, MARC 21, format COMARC/B, COBIB.SI

Abstract

The paper describes the upload of bibliographic records from the COBIB.SI shared database as well as holdings data of certain Slovenian libraries into WorldCat, the Online Union Catalogue. In 2009, the initial upload of records was made and a regular quarterly upload followed. Due to certain defects of bibliographic records in the Slovenian COBIB.SI shared database, not all records have been uploaded into WorldCat. The statistics of the records included according to various criteria in 2009 are provided at the end of the paper.

Keywords

COBISS.SI, WorldCat, format mapping, MARC 21 format, COMARC/B format, COBIB.SI

UVOD

Avgusta leta 2009 je OCLC objavil novico o vključevanju približno 3 milijonov zapisov COBIB.SI v svetovni katalog WorldCat do konca leta 2009 [1]. Novica je bila objavljena na osnovi pogodbe o sodelovanju, ki so jo pred tem sklenili predstavniki OCLC [2] in IZUM [3].

Glavni razlog za vključevanje bibliografskih zapisov iz vzajemne bibliografske baze COBIB.SI je povečanje prepoznavnosti del slovenskih avtorjev v svetovnem merilu. Prenos pripadajočih podatkov o zalogi za 10 slovenskih knjižnic in njihova predstavitev na spletnih straneh WorldCat [4] naj bi uporabnikom omogočila lažje iskanje knjižničnega gradiva v slovenskih knjižnicah preko svetovnega spleta ter lažji dostop do slovenskega vzajemnega kataloga in lokalnih katalogov slovenskih knjižnic, ki so prispevale svoje podatke o zalogi. Z vključevanjem knjižničnega fonda slovenskih knjižnic v WorldCat bi se povečala tudi prepoznavnost teh knjižnic, slovenskega vzajemnega kataloga in pripadajočih servisov. Razen tega bi bili uporabnikom na voljo različni

servisi WorldCat, ki bi zajemali tudi vključene zapise iz slovenskega kataloga, npr. kreiranje uporabniških seznamov ali izdelava bibliografij. Omogočeno je tudi iskanje izvodov gradiva po knjižnicah, ki so uporabniku geografsko najbližje (na osnovi izbrane države). Zapisi iz baze podatkov WorldCat so dostopni tudi preko servisa OCLC FirstSearch [5], ki je vključen v programsko opremo COBISS/OPAC [6]. Za knjižnice in institucije, ki prispevajo svoje podatke v bazo podatkov WorldCat, pa OCLC omogoča uporabo različnih spletnih servisov WorldCat za sestavljene spletne aplikacije ter integracijo s spletnimi stranmi knjižnic in drugih institucij.

Tako je IZUM po dogovoru z OCLC septembra 2009 pripravil za inicialni prenos v katalog WorldCat nekaj več kot 3 milijone zapisov v formatu MARC 21 [7] z več kot 4 milijoni podatkov o zalogi [8]. Priprava podatkov za inicialni prenos je potekala v naslednjih fazah:

1. analiza in kontrola zapisov iz vzajemne bibliografske baze podatkov COBIB.SI,
2. konverzija zapisov v format MARC 21,
3. prenos konvertiranih zapisov na strežnik OCLC.

Po inicialnem prenosu zapisov je IZUM začel četrletno ažurirati podatke o zalogi že vključenih zapisov. Za zdaj se prenašajo podatki o zalogi naslednjih knjižnic:

- Narodna in univerzitetna knjižnica, Ljubljana,
- Centralna medicinska knjižnica, Ljubljana,
- Centralna tehniška knjižnica Univerze v Ljubljani,
- Centralna biotehniška knjižnica, Ljubljana,
- Centralna ekonomska knjižnica, Ljubljana,
- Osrednja družboslovna knjižnica Jožeta Goričarja, Ljubljana,
- Osrednja humanistična knjižnica, Ljubljana,
- Mariborska knjižnica,
- Knjižnica Univerze v Novi Gorici,
- Znanstveno-raziskovalno središče Koper.

V OCLC se dejansko prenašajo samo kode teh knjižnic, ki se pri izpisu na spletnih straneh WorldCat razrešijo v nazive knjižnic. Vsak zapis v bazi podatkov WorldCat vsebuje tudi kodo SILIS za sistem COBIB.SI. Podatke, ki se prikazujejo v WorldCat (povezave na kataloge knjižnic in servise ter druge pomembne podatke), lahko knjižnice same pripravljajo preko aplikacije WorldCat Registry [9].

V naslednjih iteracijah je predviden tudi prenos podatkov o zalogi za večje število knjižnic (Slovenska akademija znanosti in umetnosti, Institut Jožef Stefan, Univerzitetna knjižnica Maribor, Mestna knjižnica Ljubljana in druge).

Ko bo OCLC zaključil vključevanje zapisov iz inicialnega prenosa, se bodo četrletno vključevali tudi novi zapisi iz vzajemne bibliografske baze podatkov COBIB.SI.

ANALIZA ZAPISOV

Baza podatkov WorldCat danes vsebuje več kot 180 milijonov zapisov in milijardo in pol podatkov o zalogi za več kot 72.000 knjižnic iz 171 držav. Pri vključevanju novih zapisov knjižnic je pomembna predvsem kakovost zapisov, ki mora ustrezati mednarodnim standardom in strogim kriterijem, ki jih je definiral OCLC.

Nekakovostne zapise OCLC pri vključevanju v bazo podatkov WorldCat zavrne in jih mora knjižnica ali institucija pred ponovnim pošiljanjem ustrezno popraviti. Da neustreznih oz. nekakovostnih zapisov sploh ne bi prenašali, smo analizirali zapise v vzajemni bibliografski bazi podatkov COBIB.SI.

Za prenos zapisov v bazo podatkov WorldCat je bilo možno izbirati med bibliografskima formatoma UNIMARC [10] in MARC 21. V IZUM-u smo se odločili za slednjega, ker je konverzija zapisov iz formata COMARC/B [11] v format MARC 21, ki smo jo razvili

za potrebe protokola Z39.50 in prenosa podatkov v druge sisteme v programski opremi COBISS/OPAC, popolnejša od konverzije v format UNIMARC. Razen tega v OCLC uporabljajo bibliografski format, ki izhaja iz formata MARC 21. Če bi prenašali zapise v formatu UNIMARC, bi zapise najprej konvertirali v format MARC 21 in nato še v njihov format, pri čemer bi se lahko izgubila znatna količina podatkov. Da bi preverili ustreznost konvertiranih zapisov, smo poleg analize zapisov v formatu COMARC/B naredili še analizo zapisov v formatu MARC 21. Za ta namen je bil razvit preverjalnik zapisov MARC 21, s katerim so bile posredno ugotovljene tudi pomanjkljivosti zapisov v formatu COMARC/B. Preverjalnik zapisov MARC 21 preverja strukturo zapisov po specifikaciji MARC 21:

- identifikatorje polj in identifikatorje podpolj,
- indikatorje v posameznih poljih,
- ponovljivost polj in podpolj,
- izpolnjenost polj in podpolj,
- dolžino uvodnika zapisa (24 znakov),
- dolžine kontrolnih polj 006, 007 in 008 glede na vrsto gradiva,
- kode v uvodniku in kontrolnih poljih 006, 007 in 008 glede na vrsto gradiva,
- podatke, po katerih iščemo pare med zapisi iz vzajemne bibliografske baze COBIB.SI in zapisi iz baze podatkov WorldCat (algoritem je opisan v nadaljevanju pri opisu inicialnega prenosa zapisov).

Na osnovi analize zapisov po obeh bibliografskih formatih (COMARC/B in MARC 21) smo iz vzajemne baze podatkov COBIB.SI najprej izločili naslednje zapise:

- dogodke (001b = u) in raritete (001a = r),
- zapise, označene za brisanje (001a = d), prve vnose zapisov (001a = i) in predhodne (CIP) zapise (001a = p),
- stare zapise z napačnim bibliografskim nivojem (001c = e), zapise s staro hierarhijo (maska M2) in zapise na gornjem nivoju hierarhično obdelanih zapisov s staro hierarhijo (povezane preko polj 461, 462 in 463),
- zapise, prevzete iz baze podatkov ISSN [12],
- zapise za neknjižno gradivo.

Na ta način se v WorldCat niso prenesli zapisi, ki ne ustrezajo strukturi formata MARC 21, niso popolni, bodo izbrisani iz vzajemne bibliografske baze, niso bili popravljani po ukinitvi stare hierarhije ali so bili prevzeti iz baze podatkov ISSN in se zaradi licenčnih pogojev ne smejo distribuirati. Iz inicialnega prenosa smo izločili tudi zapise za neknjižno gradivo, ker je konverzija le-teh v format MARC 21 zelo kompleksna. Skupaj je tako bilo iz prenosa izločenih 385.263 zapisov.

KONTROLA ZAPISOV

Po izločitvi zapisov po izbranih kriterijih smo preostale zapise iz vzajemne bibliografske baze podatkov COBIB.SI konvertirali v format MARC 21, nato pa se je s preverjalnikom preverila njihova usklajenost s specifikacijo formata MARC 21. Glede na napake v zapisih, ki jih je program našel, so bili analizirani zapisi tudi v formatu COMARC/B. Za vsak tip napake je tako določena programska kontrola, s katero se lahko iz prenosa izločijo neustrezni zapisi v formatu COMARC/B. Vse te kontrole lahko razdelimo v tri skupine:

1. kontrole obveznosti in ponovljivosti polj in podpolj,
2. programske kontrole iz segmenta COBISS2/ Katalogizacija in
3. vsebinske kontrole.

Ker pa zapisi v formatu MARC 21 vsebujejo tudi podatke o virih člankov in drugih sestavnih delih v istem zapisu, se navedene kontrole nanašajo tudi na zapise iz vzajemne bibliografske baze COBIB.SI, ki predstavljajo te vire.

Na ta način je bilo najprej izločenih preko 80.000 zapisov. Da bi zagotovili njihovo vključevanje v bazo podatkov WorldCat, smo se odločili, da analiziramo izločene zapise in preverimo možnosti programskega odpravljanja napak, saj toliko zapisov ni bilo mogoče ročno popraviti. V ta namen so bili pripravljene sezname izločenih zapisov po različnih napakah, skupaj z vsemi podatki, ki se preverjajo pri določeni kontroli, in s podatki, ki bi lahko bili v pomoč pri razreševanju napak. Na primer, za programsko kontrolo 28, s katero se preverja, ali je v podpolju 001b (*Vrsta zapisa*) koda "g" ali "m", če je v zapisu izpolnjeno polje 115, se razen polj 115 in podpolja 001b izpiše še vsebina podpolja 200b (*Splošna oznaka gradiva*). Iz tega podatka se v nekaterih primerih dá ugotoviti, kakšno kodo je treba vnesti v podpolje 001b, ali pa sklepati, da ta zapis sploh ne bi smel vsebovati polja 115.

Z analizo seznamov izločenih zapisov je bilo ugotovljeno, da je možno programsko popravljati samo dve vrsti napak:

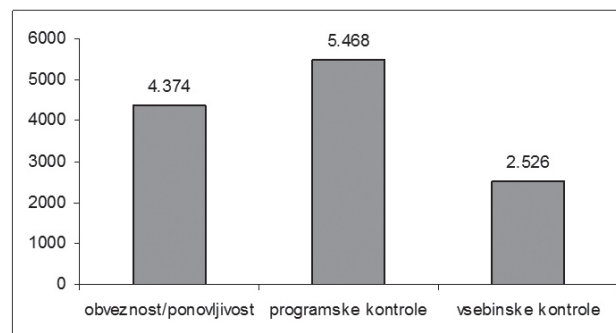
- zaporedje podpolj za podatke o sestankih v poljih 601, 710, 711 in 712 mora biti dfe (preko 8.000 zapisov),
- napačen prvi indikator v polju 101 (npr. znak "|" pri zapisih, prevzetih iz OCLC) (preko 60.000 zapisov).

Ker pa bi programsko popravljjanje napak v zapisih zahtevalo konverzijo vzajemne bibliografske baze podatkov in vseh lokalnih baz podatkov, smo se odločili, da te napake izjemoma popravimo pri konverziji zapisov v format MARC 21. Tako smo prvo napako, ki vpliva tudi na izpis kataložnih listkov v programski opremi COBISS, odpravili s spremembo vrstnega reda podpolj

dfe, drugo pa tako, da smo vrednost prvega indikatorja v polju 040 formata MARC 21 določili glede na to, ali obstaja podpolje 101b (*Jezik posrednega besedila*) ali 101c (*Jezik izvornika*) – takrat je vrednost prvega indikatorja 1, drugače pa presledek.

Drugih napak ni mogoče programsko popraviti, ker bi s tem tvegali nastanek drugih napak v zapisih. Zato je bilo odločeno, da bomo napake v IZUM-u popravljali ročno na osnovi izdelanih seznamov napak, če bo možno zapise redigirati brez gradiva. Algoritem izločanja neustreznih zapisov je nastavljen tako, da se kontrole izvajajo sekvenčno in se zapis izloči takoj pri prvi kontroli, ki najde pomanjkljivost v zapisu, kar ne pomeni, da drugih napak v zapisu ni. Pri redakciji zapisov pa je treba odpraviti tudi preostale napake v določenem zapisu, kar delo redaktorjev dodatno oteži.

Po programskih popravkih in redakciji zapisov je bilo na dan priprave podatkov za prenos v bazo podatkov WorldCat zaradi kontrol izločenih 12.368 zapisov. Posamezne skupine kontrol in statistike zapisov po različnih tipih napak (slika 1) so podrobneje opisane v nadaljevanju.



Slika 1: Statistika napak v zapisih pri inicialnem prenosu

Obveznost in ponovljivost polj in podpolj

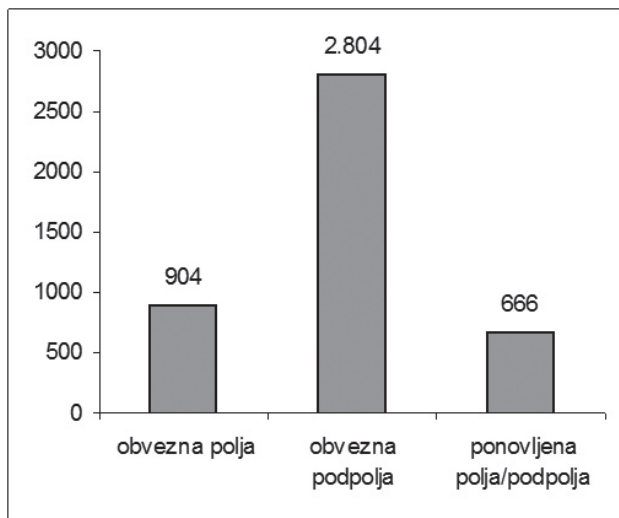
Glede na to, da se je bibliografski format COMARC/B z leti dopolnjeval in spreminjal, bibliografski zapisi pa se niso sistematično usklajevali s temi spremembami (razen izjemoma pri kakšni konverziji lokalnih bibliografskih baz in vzajemne bibliografske baze), je bilo treba preveriti, ali zapis vsebuje vsa potrebna polja in podpolja za pravilno pretvorbo zapisa v format MARC 21. Z analizo formata MARC 21 so bile določene pomanjkljivosti obveznih polj in podpolj in njihova ponovljivost v zapisih v formatu COMARC/B glede na vrsto gradiva.

1. Obvezna polja in podpolja:
 - za vse vrste gradiva – 200a, 001abcd, 100ch,
 - za kontinuirane vire – 100b, 101a, 110a, 210c,
 - za monografije, članke in druge sestavne dele – 101a.
2. Neponovljiva polja:
 - 011, 120, 700.

3. Neponovljiva podpolja:

- 011a, 6XXa (razen 610a in 686a), 30Xa, 31Xa, 32Xa (razen 327a), 33Xa, 215c, 675v.

S kontrolami teh polj in podpolj, ali so obvezna in ponovljiva, je izločenih 4.374 zapisov (slika 2).



Slika 2: Statistika napak v zapisih glede na obveznost in ponovljivost polj in podpolj

Statistika napak po posameznih poljih in podpoljih pa je naslednja:

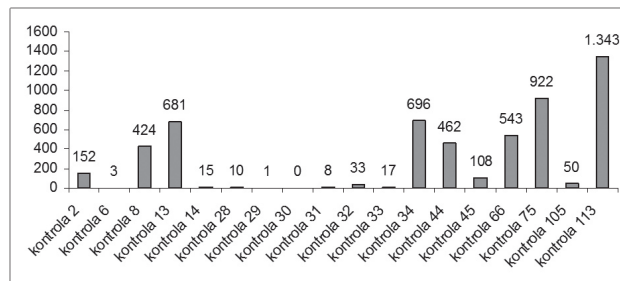
- obvezna polja 100 (3 zapisi), 101 (219 zapisov), 110 (52 zapisov), 200 (488 zapisov), 210 (142 zapisov),
- obvezna podpolja 001d (104 zapisi), 100b (32 zapisov), 100c (2.531 zapisov), 101a (12 zapisov), 110a (1 zapis), 200a (57 zapisov), 210c (67 zapisov),
- ponovljena polja/podpolja 011 (1 zapis), 011a (28 zapisov), 300a (9 zapisov), 301a (1 zapis), 320a (1 zapis), 330a (357 zapisov), 606a (3 zapisi), 627a (4 zapisi), 675a (187 zapisov), 215c (63 zapisov), 675v (12 zapisov).

Največje število zapisov je izločenih zaradi manjkajoče letnice v podpolju 100c (*Leto izida 1*) ali manjkajočega naslova v polju 200. Večina zapisov s temi napakami je bila v sistem COBISS.SI prenesena s konverzijo iz drugih sistemov. Napake pa so lahko tudi posledica kreiranja kratkih zapisov v posameznih knjižnicah za potrebe izposoje, kar je bilo možno v prvi verziji programske opreme za katalogizacijo.

Programske kontrole

Programske kontrole so v segment programske opreme COBISS2/Katalogizacija uvedene leta 1997 in so od takrat večkrat dopolnjene ali spremenjene. Ker se spremenjene ali dodane kontrole ne izvajajo sistematično

na zapisih, ki so že shranjeni v vzajemno bibliografsko bazo podatkov, je bilo treba iz prenosa izločiti vse zapise, ki ne ustrezajo izbranim kontrolam (slika 3). Opis posameznih kontrol je v priložniku *COBISS2/Katalogizacija, Dodatek G* [13].



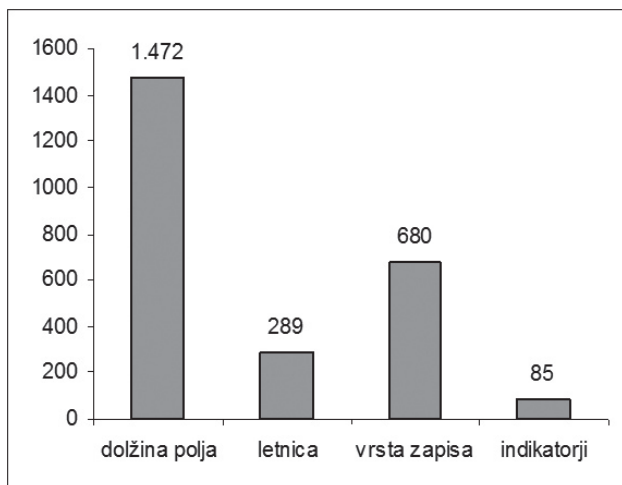
Slika 3: Statistika napak v zapisih po programskih kontrolah

S temi programskimi kontrolami je bilo izločenih 5.468 zapisov. Največ zapisov je bilo izločenih zaradi kontrole 113, ko program preverja, ali je v ponovljivem podpolju 200f (*Prva navedba odgovornosti*), ki neposredno sledi prvemu podpolju f, na prvem mestu enačaj. Pri analizi zapisov je bilo ugotovljeno, da ni možno programsko dodati enačaja v ponovljivo podpolje 200f, ker v veliko primerih obstaja napaka v oznaki podpolja – namesto podpolja 200f bi morale biti podpolje 200g (*Naslednja navedba odgovornosti*), česar pa ni mogoče programsko določiti.

Veliko zapisov je bilo izločenih tudi zaradi kontrole 75, ko program preverja, ali ima drugi indikator v poljih 70X in 600 vrednost 0, če polje vsebuje podpolje d (*Rimske številke*). Ta kontrola v segmentu COBISS2/Katalogizacija dopušča shranjevanje napačnega zapisa, pri prenosu v bazo podatkov WorldCat pa takšni zapisi povzročajo strukturne napake v zapisih v formatu MARC 21, zato so bili izločeni. Podobno kontrole 28–34 bistveno vplivajo na strukturo zapisa v formatu MARC 21, na zapis v formatu COMARC/B pa ne. Te kontrole se nanašajo na usklajenost kode za vrsto gradiva v podpolju 001b z izpolnjenostjo določenih polj v zapisu, preko katerih se določajo kode v kontrolnih poljih 006, 007 in 008 v formatu MARC 21. Druge uporabljene programske kontrole pa bistveno vplivajo na zapise v obeh formatih, zato so neustrezni zapisi zaradi teh kontrol izločeni iz prenosa v bazo WorldCat.

Vsebinske kontrole

Da bi izločili čim več zapisov, ki ne ustrezajo strukturi formata MARC 21, smo uvedli dodatne vsebinske kontrole, ki preverjajo vsebino določenih podpolj in indikatorjev polj (slika 4).



Slika 4: Statistika napak v zapisih po vsebinskih kontrolah

Zapisi so imeli naslednje napake:

- napačna dolžina podpolja – 100c (546 zapisov), 100d (26 zapisov), 101a (16 zapisov), 105a (883 zapisov), 115b (1 zapis), skupaj 1.472 zapisov,
- napačna letnica v podpolju 100c (289 zapisov),
- v podpolju 115b podatek ni številka,
- če je v podpolju 001c koda "d", mora biti v podpolju 001b koda "u" (671 zapisov),
- če je v podpolju 135b koda "h" ali "j", mora biti v podpolju 001b koda "l" ali "m",
- koda v podpolju 106a je lahko samo "d", "e", "f", "g" ali "h",
- če je v podpolju 001c koda "d", v zapisu ne sme biti podpolja 115a, 116a, 124b in 135b,
- za članke v serijskih publikacijah v podpolju 001b ne sme biti kode "b", "d" ali "f" (9 zapisov),
- vrednosti prvega in drugega indikatorja v poljih so lahko številke ali presledek (85 zapisov),
- polje 135 mora vsebovati vsaj eno podpolje a ali b,
- polja 115, 116, 126 in 130 morajo vsebovati podpolje a,
- polje 124 mora vsebovati podpolje b,
- če zapis vsebuje polje 130, mora biti koda v podpolju 001b "a" ali "b",
- če je v podpolju 106a koda "g", mora v zapisu obstajati polje 130,
- koda "c" ne more biti hkrati v podpoljih 124a in 124b,
- če je v podpolju 115a koda "a", v podpolju 115g ne sme biti kode "g", "h", "i", "j", "k" in "l" in polje 115 ne more vsebovati podpolja k ali l,
- če je v podpolju 115a koda "b", v podpolju 115g ne sme biti kode "a", "b", "c" ali "d".

S temi kontrolami vsebine je bilo tako izločenih 2.526 zapisov. Iz seznama je razvidno, da z večino kontrol program ni našel napačnih zapisov. To je zaradi tega,

ker so te kontrole namenjene zapisom za neknjižno gradivo, ki v inicialnem prenosu zapisov v bazo podatkov WorldCat ni vključeno, bo pa vključeno v prihodnjih četrletnih prenosih zapisov po končanem vključevanju zapisov iz inicialnega prenosa.

Največ zapisov je bilo izločenih zaradi napačne dolžine vsebine posameznih podpolj ali zaradi napačnih kod v določenih podpoljih. Med njimi izstopajo zapisi z napačnimi kodami v podpolju 105a (*Kode za ilustracijo*). Kode in njihova dolžina se je v formatu COMARC/B spremenila, konverzija teh kod pa za obstoječe zapise ni bila narejena. Veliko zapisov ima tudi napačno dolžino ali vsebino v podpolju 100c, ki na primer vsebuje letnice v oklepajih ([, < >), letnice s pomišljaji, sestavljene letnice ali okrajšave (npr. "s.a.", "s.n.", "b.l."). Največ teh napak je nastalo pri konverzijah zapisov iz drugih sistemov v sistem COBISS.

INICIALNI PRENOS ZAPISOV

Z izločitvijo neustreznih in neakovostnih zapisov glede na opisane kriterije je bilo za inicialni prenos zapisov iz vzajemne bibliografske baze podatkov COBIB.SI dne 31. 8. 2009 pripravljenih 3.129.559 zapisov s 4.309.076 podatkov o zalogi. Ker se bibliografski zapisi, vključeni v WorldCat, ne ažurirajo, pač pa se ažurirajo samo podatki o zalogi, smo zapise v formatu COMARC/B skupaj s kodami knjižnic, ki imajo zalogo, shranili v posebno bazo prenesenih zapisov. Pri ažuriranju podatkov o zalogi bo tako v OCLC vedno poslan enak bibliografski zapis, kakršen je bil v času vključevanja.

Zapise iz baze prenesenih zapisov smo nato konvertirali v format MARC 21 in pripravili paket 35 datotek (vsaka z največ 90.000 zapisi) po standardu ISO 2709 [14], ki smo jih 4. 9. 2009 prenesli na strežnik OCLC v okviru njihovega servisa za izmenjavo zapisov EDX (*Electronic Data Exchange*) [15] preko protokola FTP (*File Transfer Protocol*).

Prenesene zapise so v OCLC preverili po standardu ISO 2709 in nam vrnilo 5 neustreznih zapisov. Ugotovljeno je bilo, da je struktura teh zapisov neustrezna zaradi predolghih polj 327 in 330. Ker je dolžina polja po standardu ISO 2709 največ 9999 znakov, v formatu COMARC/B pa takšnih omejitev ni, se bodo v prihodnje predolga podpolja izločila iz zapisa pred konverzijo v format MARC 21.

Nato je v OCLC stekel postopek iskanja parov med poslanimi zapisi iz vzajemne bibliografske baze COBIB.SI in zapisi v bazi podatkov WorldCat. Cilj tega postopka je, da se v bazi WorldCat ne kreirajo podvojeni zapisi, pač pa da se podatki o zalogi knjižnic dodajo že

obstoječim zapisom, če je le možno. Algoritem iskanja ustreznega zapisa v podatkovni bazi WorldCat poteka v štirih korakih.

1. Iskanje preko identifikatorjev:
 - OCLC iz podpolja 035a,
 - LCCN iz podpolja 010a,
 - ISBN iz podpolja 020az,
 - ISSN iz podpolja 022a,
 - CODEN iz podpolja 030a,
 - URL iz podpolja 856u,
 - številke založnika iz podpolja 028a in 262c ali
 - kakšne druge identifikacijske številke iz podpolj 016a, 024a, 027a, 088a, ali 029a.
2. Če v prvi fazi program ni našel ustreznega zapisa, se upoštevajo podatki o:
 - avtorjih iz podpolj 100a, 110ab, 130a, 700a, 710ab, 711a, 730a in 720a,
 - naslovu iz podpolj 245akbfnp,
 - založniku iz podpolj 533c, 260bf, 261abe in 262b,
 - letu izida iz polja 008 ali
 - vrsti gradiva.
3. Če je program našel več ustreznih zapisov za povezovanje, je treba dodatno preveriti njihovo ustreznost na osnovi podatkov o:
 - jeziku katalogizacije iz podpolja 040b,
 - letu izdaje iz podpolj 260c, 553d in 362a,
 - naslovih v poljih 246 in 247 in
 - fizični obliki.
4. Določanje najbolj ustreznega zapisa s točkovanjem od 1 do 15 na osnovi dodatnih internih kriterijev v OCLC.

Pred iskanjem parov med zapisi se posamezni podatki normalizirajo po določenih pravilih (npr. brisanje interpunkcij, pretvorba v male črke itn.). Če z algoritmom za iskanje parov program ne najde nobenega ustreznega zapisa, se prenesen zapis vključi v bazo podatkov WorldCat kot originalni zapis knjižnice oz. institucije, ki je zapis poslala. Tako 3.063.840 zapisov, ki so bili preneseni iz vzajemne bibliografske baze COBIB.SI z inicialnim prenosom, predstavlja originalne zapise v WorldCatu. Program je našel tudi 36.269 ustreznih zapisov, ki so jim dodani samo podatki o zalogi. Razlog za tako veliko število originalnih zapisov je jezik katalogizacije (slovenski), ki je eden od kriterijev za določanje podvojenih zapisov v podatkovni bazi WorldCat.

Na ta način je bilo z inicialnim prenosom v WorldCat vključenih 3.100.109 zapisov. Preostalih 29.445 zapisov je nerazrešenih in čakajo na nadaljnjo obdelavo. V OCLC bodo te zapise najprej poskušali programsko obdelati glede na napake, ki so jih ugotovili pri vključevanju v bazo podatkov. Zapise, ki niti po programski odpravi napak ne bodo vključeni, OCLC pošlje knjižnicam ali

drugim institucijam skupaj z opisom napak, ki naj bodo odpravljene pred ponovnim pošiljanjem zapisov v OCLC.

OCLC pripravi za vključene zapise tabele preslikav med identifikatorji zapisov OCLC-ID in COBIB.SI-ID (tako imenovane tabele XREF), kar omogoča sinhronizacijo baze prenesenih zapisov v IZUM-u z bazo podatkov WorldCat. V vsak vključen zapis IZUM doda podpolje 035a z identifikatorjem OCLC-ID, ki pri ažuriranju zaloge zapisov omogoča enoznačno povezovanje zapisov iz vzajemne bibliografske baze COBIB.SI z ustreznimi zapisi v podatkovni bazi WorldCat. Na ta način se onemogoči kreiranje podvojenih zapisov v bazi WorldCat.

ČETRTLETNI PRENOSI ZAPISOV

Po končanem inicialnem prenosu se bodo zapisi iz vzajemne bibliografske baze podatkov COBIB.SI prenašali v bazo podatkov WorldCat četrtno (vsake 3 mesece). S tem bo v bazi podatkov WorldCat omogočeno ažuriranje podatkov o zalogi knjižnic pri že vključenih zapisih in vključevanje novih zapisov iz vzajemne baze COBIB.SI. Ker OCLC še ni končal postopka inicialnega vključevanja zapisov v bazo podatkov WorldCat (počakati je treba na poročilo o napakah v nevklučenih zapisih), smo se v IZUM odločili, da do takrat v OCLC ne pošiljamo novih zapisov, pač pa samo ažuriramo podatke o zalogi že vključenih zapisov.

Načeloma se bibliografski zapisi iz podatkovne baze WorldCat ne brišejo. Zato se za zapise, ki so brisani iz vzajemne baze COBIB.SI in so vključeni v bazo podatkov WorldCat, pobrišejo samo podatki o stanju zaloge skupaj s kodo SILIS za sistem COBISS.SI.

Podatki o stanju zaloge se ažurirajo tako, da se v OCLC prenesejo skupaj z bibliografskim zapisom, ki je bil poslan v OCLC ob vključitvi le-tega v bazo podatkov WorldCat. Natančneje, pošlje se en zapis s kodami knjižnic, ki nimajo več zaloge, in še dodaten zapis s kodami knjižnic, ki so v tem času pridobile zalogo. Razlika v teh dveh bibliografskih zapisih je samo v statusu zapisa v podpolju 001a: "d" – *zapis s kodami za brisanje*, "n" – *zapis z dodanimi kodami*.

Novi zapisi, ki se bodo prenašali v bazo podatkov WorldCat, se bodo pripravljali na enak način kot v postopku inicialnega prenosa. Pred prenosom bo program s prej opisanimi kontrolami preveril zapise v formatu COMARC/B in izločil neustrezne in nekakovostne zapise. Skupaj z zapisi, ki so pripravljani za ažuriranje podatkov o zalogi, se bodo nato konvertirali v format MARC 21. Zapisi, ki bodo ustrezali strukturi formata MARC 21, bodo preneseni na strežnik OCLC in uvrščeni v vrsto za vključevanje v bazo podatkov WorldCat. Rok za obdelavo

poslanih zapisov v OCLC je 90 dni, lahko pa se to obdobje podaljša, kar pa ne zadrži naslednjega četrletnega prenosa zapisov, ki ga opravi knjižnica ali druga institucija. Po vsaki vključitvi zapisov pošlje OCLC tabele XREF, na osnovi katerih se sinhronizirata baza prenesenih zapisov v IZUM-u in baza podatkov WorldCat.

STATISTIKE VKLJUČENIH ZAPISOV

V letu 2009 je bilo v bazo podatkov WorldCat vključenih 3.099.821 zapisov iz vzajemne bibliografske baze COBIB.SI in 4.387.116 podatkov o zalogi knjižnic.

V tabeli 1 je prikazana statistika vključenih zapisov po kreatorjih (sistemih v mreži COBISS.Net, v katerih so zapisi kreirani) in bibliografskem nivoju. Iz statistike je razvidno, da je največ vključenih zapisov za monografske publikacije, od prevzetih zapisov pa največ iz sistema COBISS.SR.

Sistem	Mono-grafije	Serijske	Integrirni viri	Članki	Zbirni zapisi	Skupaj
COBIB.SI	1.832.021	28.883	123	1.229.972	1.840	3.092.839
COBIB.SR	5.802	110	0	25	0	5.937
COBIB.MK	578	10	0	6	0	594
COBIB.BH	305	10	0	0	0	315
COBIB.CG	129	7	0	0	0	136
Skupaj	1.838.835	29.020	123	1.230.003	1.840	3.099.821

Tabela 1: Statistika vključenih zapisov po kreatorjih in bibliografskem nivoju

V tabeli 2 je prikazana statistika vključenih zapisov po jeziku in bibliografskem nivoju. Iz statistike je razvidno, da je največ bibliografskih enot v slovenskem jeziku, od monografskih publikacij pa tistih v angleškem jeziku.

Jezik	Monografije	Serijske	Integrirni viri	Članki	Zbirni zapisi	Skupaj
slv	519.397	12.017	59	907.026	672	1.439.171
eng	525.436	5.016	25	179.404	170	710.051
ger	280.520	2.757	33	19.176	271	302.757
scc	107.307	2.250	3	48.238	226	158.024
scr	108.648	2.508	0	38.658	240	150.054
ita	73.000	978	0	9.422	75	83.475
fre	72.645	531	0	4.695	57	77.928
hun	29.284	233	0	3.980	7	33.504
rus	27.197	512	0	1.053	3	28.765
mac	11.324	304	0	11.154	17	22.799
bos	1.296	22	3	346	2	1.669
drugi	82.781	1.892	0	6.851	100	91.624
Skupaj	1.838.835	29.020	123	1.230.003	1.840	3.099.821

Tabela 2: Statistika vključenih zapisov po jeziku in vrsti gradiva

V tabeli 3 je prikazana statistika podatkov o zalogi vključenih zapisov po kreatorjih (sistemih v mreži COBISS.Net, v katerih so zapisi kreirani) in knjižnicah. Iz statistike je razvidno, da je največ podatkov o zalogi preneseno za NUK in da je največ podatkov o zalogi za prevzete zapise iz COBISS.SR.

Knjižnica	COBIB.SI	COBIB.SR	COBIB.MK	COBIB.BH	COBIB.CG	Skupaj
COBIB.SI	3.092.839	5.937	594	315	136	3.099.821
NUK	710.518	833	15	19	18	711.403
FFLJ	209.544	1.005	395	54	46	211.044
CTK	115.934	185	0	1	0	116.120
SIKMB	89.138	20	0	6	0	89.164
ODKLJ	74.080	13	8	6	2	74.109
CEKLJ	45.038	26	1	5	2	45.072
CMK	15.632	2	0	1	0	15.635
PNG	10.298	7	0	2	0	10.307
ZRSKP	9.988	6	0	0	0	9.994
BFCBK	4.446	1	0	0	0	4.447
Skupaj	4.377.455	8.035	1.013	409	204	4.387.116

Tabela 3: Statistika podatkov o zalogi po kreatorjih in knjižnicah

V tabeli 4 je prikazana statistika podatkov o zalogi vključenih zapisov po kreatorjih (sistemih v mreži COBISS.Net, v katerih so zapisi kreirani) in jezikih. Iz statistike je razvidno, da je največ podatkov o zalogi preneseno za bibliografske enote v slovenskem jeziku in da največ podatkov o zalogi za prevzete zapise posredujejo iz COBISS.SR za bibliografske enote v srbskem jeziku.

Jezik	COBIB.SI	COBIB.MK	COBIB.SR	COBIB.BH	COBIB.CG	Skupaj
slv	2.191.359	1	110	1	0	2.191.471
eng	937.974	23	214	8	10	938.229
ger	408.369	0	29	2	2	408.402
scr	201.871	2	762	83	0	202.718
scc	193.812	11	6.300	46	191	200.360
ita	103.395	0	5	0	0	103.400
fre	100.355	1	12	1	0	100.369
rus	40.466	0	14	0	0	40.480
hun	34.906	0	129	0	0	35.035
mac	26.033	967	357	1	0	27.358
bos	2.105	0	48	251	0	2.404
drugi	136.810	8	55	16	1	136.890
Skupaj	4.377.455	1.013	8.035	409	204	4.387.116

Tabela 4: Statistika podatkov o zalogi po jeziku in kreatorjih

ZAKLJUČEK

Zaradi večje razpoznavnosti slovenskih knjižnic in lažjega dostopa do njihovih podatkov o zalogi je IZUM v letu 2009 začel postopek vključevanja bibliografskih zapisov iz vzajemne bibliografske baze podatkov COBIB.SI v svetovni katalog WorldCat in prenosa podatkov o zalogi za 10 slovenskih knjižnic. Do sedaj je vključenih preko 3 milijone zapisov in preko 4 milijone podatkov o zalogi in že opažamo povečanje zahtevkov za medknjižnično izposojlo bibliografskih enot, ki so vključene v WorldCat in imajo oznako slovenskega vzajemnega kataloga SILIS.

Zaradi kompleksnosti postopka prenosa in vključevanja zapisov ter neustreznosti določenih bibliografskih enot iz vzajemne bibliografske baze podatkov COBIB.SI po kriterijih, ki jih predpisuje OCLC, je pred popolno avtomatizacijo postopka potreben še dodatni razvoj in usklajevanje s postavljenimi zahtevami. V naslednjih iteracijah je najprej predviden prenos podatkov o zalogi že vključenih zapisov za večje število knjižnic ter bibliografskih zapisov za neknjižno gradivo. Sproti se bodo redigirali tudi bibliografski zapisi z napakami glede na kontrole, ki se izvajajo pred prenosom zapisov v vsaki iteraciji. Te kontrole bodo tudi dopolnjene glede na poročilo OCLC-ja o neustreznih zapisih, ki so jih zavrnili.

Glede na to, da se je takoj po vključitvi zapisov povečalo zanimanje za gradivo iz slovenskega vzajemnega kataloga, lahko rečemo, da je cilj projekta dosežen, to je povečana prepoznavnost vzajemnega kataloga in knjižnic, ki so prispevale podatke o zalogi. Naslednji korak je uporaba spletnih servisov WorldCat, ki bodo vključevali tudi naše zapise. Že v naslednjo verzijo programske opreme COBISS/OPAC, V6, bodo vključene povezave do zapisov v WorldCat.

Ker je zanimanje knjižnic za vključitev v WorldCat vedno večje, lahko IZUM s svojimi izkušnjami in modelom vključevanja pomaga knjižnicam oz. sistemom v mreži COBISS.Net pri prenosu zapisov iz njihovih vzajemnih bibliografskih baz v bazo podatkov WorldCat.

Reference

- [1] OCLC announces agreements in Europe to extend coverage in WorldCat. OCLC. Dosegljivo na: <http://www.oclc.org/us/en/news/releases/200946.htm> (20. 3. 2010).
- [2] OCLC, <http://www.oclc.org>.
- [3] IZUM, <http://www.izum.si>.
- [4] WorldCat.org, <http://www.worldcat.org>.
- [5] OCLC FirstSearch, <http://www.oclc.org/firstsearch/>.
- [6] COBISS/OPAC, <http://www.cobiss.si>.
- [7] MARC 21 Format for Bibliographica data. Library of Congress, Network Development and MARC Standards Office. Dosegljivo na: <http://www.loc.gov/marc/bibliographic/ecbdhome.html> (20. 3. 2010).
- [8] Letno poročilo za leto 2009. Maribor: IZUM, 25. 2. 2010.
- [9] WorldCat Registry, <http://www.worldcat.org/registry/Institutions>.
- [10] UNIMARC manual, Bibliographic format. München, K. G. Saur, 2008.
- [11] COMARC/B format za bibliografske podatke. Maribor: IZUM, 2008.
- [12] ISSN International Centre, <http://www.issn.org/>.
- [13] COBISS2/Katalogizacija : priročnik za uporabnike. Maribor, IZUM, 2008.
- [14] ISO 2709:1996(E). Information and documentation – Format for information exchange. International Organization for Standardization, Geneva, 1996.
- [15] OCLC Electronic Data Exchange, <http://www.oclc.org/edx/>.