



Združevanje bibliografskih zapisov v COBISS+: začetek novih katalogov v sistemih COBISS

Clustering bibliographic records in COBISS+: the beginning of new catalogues in the COBISS system

Andreja Krajnc Vobovnik, Gordana Mazić

Oddano: 11. 4. 2017 – Sprejeto: 28. 5. 2017

1.04 Strokovni članek

1.04 Professional article

UDK 004.78:025.4.036(497.4)

Izvleček

Skupaj z zasnovo COBISS+ smo želeli začeti postopek prestrukturiranja katalogov v sistemu COBISS po modelu FZBZ, vendar smo se po daljši analizi sedanjih odnosov med zapisi v katalogu in katalogizacijskih pravil omejili na tri odnose; to so prevodi, revizije in reprodukcije. Po njih smo oblikovali skupke bibliografskih zapisov (angl. *cluster*) in jih poimenovali »Vse izdaje in prevodi«. Pri konceptualizaciji skupkov smo imeli več težav. Prva dilema se je pojavila pri obravnavanju intelektualnih in umetniških stvaritev, ki so samostojni deli v okviru širših celot, in istih stvaritev, ki so celota zase. Glede na velikost katalogov ter na obseg domače založniške produkcije smo se odločili, da samostojne dele širših celot obravnavamo enako kot celote. Druga dilema je bila povezana z določanjem dela. V bibliografskih zapisih, ki so zgrajeni po principih opisne katalogizacije, namreč ni dovolj podatkov, ki bi omogočali zanesljivo identifikacijo istih intelektualnih ali umetniških stvaritev oziroma jasno ločitev dela od dela. Zato smo za združevanje zapisov za posamezne vrste virov pripravili celo vrsto različnih algoritmov. Uporaba teh algoritmov nam je omogočila povezovanje dejanske strukture bibliografskih zapisov z našo okvirno shemo, za katero predvidevamo, da bo zanimiva za uporabnike.

Analiza skupkov »Vse izdaje in prevodi« bo prispevala k izboljšanju algoritmov, s katerimi identificiramo bibliografske odnose v katalogih. Nadaljnji razvoj strukturiranega kataloga COBISS+ vidimo v prehodu na sodobna katalogizacijska pravila, ki upoštevajo konceptualni model FZBZ oziroma IFLIN knjižnični referenčni model (angl. *IFLA Library Reference Model* ali *IFLA LRM*). V tem vidimo prostor za prestrukturiranje sedanjih zapisov v zahtevnem postopku določanja entitet in odnosov med njimi, na katerih želimo utemeljiti strukturirane kataloge COBISS v prihodnosti.

Ključne besede: *COBISS+, odnosi med bibliografskimi zapisi, skupki bibliografskih zapisov, odnosi FZBZ*

Abstract

Along with the conception of COBISS+ the wish arose to start the process of restructuring the COBISS catalogues following the FRBR model. After an extended analysis of the existing relationships between records in the catalogue and the cataloguing rules, the procedure was limited to three relationships: translations, revisions and reproductions. Bibliographic records were clustered on the base of these relationships and the clusters were named "All editions and translations". When conceptualising the clusters we had several problems. The first dilemma was about treating intellectual and artistic works that are independent parts within wider entities, and treating these same creations that are independent entities. Depending on the catalogue size and the extent of the domestic publishing production it was decided that independent parts of larger entities should be treated the same as entities. The other dilemma was related to defining the work. Bibliographic records that are built only according to the principles of descriptive cataloguing do not contain enough data to ensure reliable identification of the same intellectual or artistic creation or allow for a clear distinction between works. This is why several different algorithms for individual types of resources were created for merging the records. The use of these algorithms allowed for the linking of the actual structures of bibliographic records with the framework scheme, for which it is assumed that it might be interesting to the users.

The analysis of the clusters "All editions and translations" will contribute towards the improvement of algorithms used to identify bibliographic relationships in the catalogue. The further development of the COBISS+ structured catalogue is seen in the transition to modern cataloguing rules that take into account the FRBR conceptual model or the IFLA Library Reference Model (IFLA LRM). These should provide options for restructuring the existing records in the demanding procedure of defining entities and the relationships between them, on which the COBISS structured catalogue shall be based in the future.

Keywords: *COBISS+, bibliographic relationships, clustering bibliographic records, FRBR relationships*

1 Uvod

Pri zasnovi nove verzije spletne aplikacije COBISS/OPAC, poimenovane COBISS+, smo si zastavili cilj, da bibliografske zapise postavimo v odnose, ki bodo uporabnikom omogočili celovit vpogled v posamezne intelektualne ali umetniške stvaritve, ne glede na jezikovne transformacije (prevode), spremembe dela, ki jih prinašajo popravljene, dopolnjene ali razširjene izdaje, in tudi ne glede na spremembe, ki so posledica umetniških poustvaritev. Prav tako smo želeli, da uporabniki dobijo vpogled v ista dela na različnih nosilcih, izdana pri različnih

založnikov ali v drugi fizični obliki. Grupiranje različnih intelektualnih in fizičnih izvedb istega dela, ki jih predstavlja skupina posameznih bibliografskih zapisov, je namenjeno lažji identifikaciji dela, hkrati pa celovitemu vpogledu v knjižnični fond. Kot večina načrtovalcev javnih katalogov smo tudi mi začeli preučevati IFLIN model funkcionalnih zahtev za bibliografske zapise (v nadaljevanju FZBZ) oziroma možnost frbrizacije katalogov. Čeprav smo v literaturi našli dokaj spodbudne rezultate eksperimentalnih identifikacij FZBZ entitet in odnosov v utemeljenih formatih MARC21 in UNIMARC, se je bilo treba usmeriti na dejanske bibliografske podatke v katalogih.

V knjižnični informacijski sistem COBISS je povezanih 500 knjižnic vseh vrst in v njihovih spletnih katalogih je bilo na začetku projekta leta 2015 več kot 13 milijonov zapisov. Vse te zapise je kreiralo več kot 1000 katalogizatorjev v več kot 25 letih delovanja sistema COBISS. Bibliografski in normativni zapisi so v formatu COMARC, ki je zasnovan na formatu UNIMARC. V celotnem obdobju vodenja katalogov v sistemu COBISS smo v slovenski katalogizacijski praksi uporabljali ista katalogizacijska pravila.

Rezultati, ki smo jih dobili po prvem eksperimentu uporabe orodja Kongresne knjižnice *FRBR Display Tool* (2009), niso dali pričakovanih rezultatov (glej poglavje 2). Zato smo se usmerili na analizo vzajemnega kataloga COBIB in določenega števila knjižničnih katalogov različnih vrst knjižnic. Z analizo bibliografskih zapisov v katalogih smo želeli ugotoviti, katere elemente modela FRBR je mogoče prepoznati v dejanskih bibliografskih zapisih. Naš model je tako upošteval, da je obseg podatkov v bibliografskih zapisih v večini katalogov neenoten in ne bi omogočal prepoznavanja večine izraznih oblik dela. Število enotnih naslovov za dela, ki so izšla pod različnimi stvarnimi naslovi, je zanemarljivo, tako da je bila identifikacija dela mogoča le preko stvarnih naslovov in odnosov teh s podatki o odgovornosti.

Pri identifikaciji izraznih oblik ter odnosov med istimi in različnimi entitetami prve skupine pa smo prepoznali veliko ovir, in sicer v skopih opombah, ki bi naj opisovale odnose med bibliografskimi zapisi, pa tudi v manjšem številu polj bloka 4---. Naš model smo zato omejili le na identifikacijo dela in njegovih pojavnih oblik ter na identifikacijo prevodov in glasbenih poustvaritev kot edinima izraznima oblikama dela. Identificirali smo entitete druge skupine – osebe in korporacije, pri ugotavljanju vrste odnosa pa smo se omejili le na osebe, in sicer na samo dva odnosa – *ustvaril(a)* in *izvedel/izvedla*. Drugih odnosov nismo ugotavljali. Glede na precejšnjo poenostavitev modela FZBZ smo model opisali kot postopek oblikovanja skupkov, ki združujejo bibliografske zapise za isto delo na različnih medijih, na istem mediju v drugi fizični obliki, s spremenjenimi založniškimi ali tržnimi podatki ter prevode istega dela.

2 Teoretična izhodišča

Z definicijo dela kot tudi z navodili za razlikovanje med deli in združevanje variacij posameznega dela v skupke predstavlja model FZBZ uporabno orodje za identifikacijo, opis in primerjavo del (Bennett, Lavoie in O'Neill, 2003). Model FZBZ je odprt za različne interpretacije in implementacije; tako ga nekateri sistemi uporabljajo v celoti, nekateri pa le delno. Predvsem je problematična identifikacija izraznih oblik, zato se večina prototipov zadovolji z identifikacijo del in pojavnih oblik (Pisanski in Žumer, 2009).

Pri pregledu tuje prakse smo preverjali predvsem študije o frbrizaciji (angl. *frbrisation*), ki v knjižničarski skupnosti označuje identifikacijo entitet po modelu FZBZ (ter njihovih atributov in odnosov) iz obstoječih bibliografskih zapisov v formatu MARC (Aalberg in Žumer, 2013). V Sloveniji imamo pri prevodu izraza »frbrisation« težavo pri izbiri med angleško izpeljanko frbrizacija in FZBZ-izacija po predlogu Sekcije za terminološke slovarje Inštituta za slovenski jezik Frana Ramovša ZRC SAZU. Do sprejetja slovenskega izraza v splošno uporabo bomo zaenkrat uporabljali bolj uveljavljeni izraz frbrizacija.

Pri pregledu literature smo se osredotočili na ugotovitve, ki se nanašajo na vzajemne baze podatkov, še posebej izsledke frbrizacije slovenske nacionalne bibliografije. Aalberg in Žumer (2013) sta ugotovila, da se metode za interpretacijo enega kataloga ne morejo avtomatično uporabiti za drug katalog. Pri večini projektov frbrizacije je glavni problem, da v zapisih ni enotnega naslova, ta pa je najbolj zanesljiv podatek za identifikacijo dela, ki je izšlo pod različnimi naslovi. Ugotovila sta, da ima v slovenski nacionalni bibliografiji 21 % bibliografskih zapisov vpisan enotni naslov, kar je precej več kot v drugih katalogih, ki sta jih analizirala v svoji študiji.

V nadaljevanju smo pregledali, kako v modelu FZBZ in pri drugih projektih obravnavajo samostojna dela v okviru širših celot in istih stvaritev, ki so celota zase in so zaradi svoje kompleksne narave manj raziskane. V modelu FZBZ so poimeno-vane kot združene entitete ali agregati (angl. *aggregates entities*, v nadaljevanju agregati). V dokumentu Final Report of the Working Group on Aggregates (2011) je agregat definiran kot pojavna oblika, ki uteleša več različnih izraznih oblik, in obratno kot ena izrazna oblika, ki lahko uteleša več pojavnih oblik. Primeri so zbrana/izbrana dela, antologije, različne novele, objavljene skupaj v enem zvezku, posamezne pesmi, izdane skupaj na CD-ju, razširitve, kot so objave s predgovori, uvodi, ilustracijami, opombami, ter vzporedne objave, npr. objave za večjezična okolja, spletne objave z možnostjo izbire jezika ali objava originalnega besedila s prevodom. Aalberg in Žumer (2013) sta natančneje preučila aggregate in ugotovila, da je težko zanesljivo najti informacijo, ki označuje aggregate. Po eni

strani je za identifikacijo agregatov na voljo veliko potencialnih virov, po drugi strani pa jih je težko sistematično identificirati, ne da bi pri tem prišlo do napačne interpretacije. Zmožnost procesiranja agregatov močno vpliva na kvaliteto vsakega procesa frbrizacije.

Pri prvi zasnovi smo za identifikacijo entitet uporabili orodje Kongresne knjižnice *FRBR Display Tool* (2009), ki bibliografske podatke iz formata MARC 21 prikaže kot tri entitete – delo, izrazno obliko in pojavno obliko. Prikazi, ki smo jih dobili, so bili na majhnem vzorcu zapisov za literarna dela ustrezne kakovosti, tisti, ki smo jih dobili na večjem vzorcu zapisov za raznovrstna dela, pa niso zadoščali za uporabo v COBISS+. Nato smo pregledali še druge knjižnične kataloge, kot so OCLC-jev katalog WorldCat.org,¹ skupni katalog švedskih univerzitetnih in raziskovalnih knjižnic Libris,² belgijski katalog The Libraries of Université catholique de Louvain UCL,³ katalog nizozemskih splošnih knjižnic⁴ in še nekaj drugih. Ugotovili smo, da večina naštetih katalogov ne uporablja prikaza vseh entitet modela FZBZ in da zapise, ki so povezani z izbranim zapisom, uvrščajo v skupke, ki se imenujejo »Druge izdaje«, »Vsi formati in jeziki« in podobno. Po podrobnejših analizah smo izbrali OCLC-jev *FRBR Work-Set Algorithm* (Hickery in Toves, 2009), saj podatki v njem ustrezajo strukturi zapisov v sistemu COBISS.

3 Priprava koncepta

Testiranje smo začeli z manjšimi stratificiranimi vzorci zapisov, ki so nam omogočili hipotetično preverjanje modela na skupinah zapisov, ki so bili bolj ali manj primerni za strukturiranje po modelu FZBZ. Najprej smo preverili možnost identifikacije entitet iz prve skupine. Že na začetku smo se srečali s težavo, kako naj prepoznamo delo. Normativnih zapisov za dela ni, tudi enotni naslovi za dela, ki se izdajajo pod različnimi naslovi, se v bibliografske zapise vnašajo nedosledno. Zato smo za identifikacijo dela imeli na voljo le stvarne naslove in podatke o odgovornosti. Najprej smo upoštevali cel niz podatkov o glavnem stvarnem naslovu z vsemi dodatki, naslovi podrejenih del in vzporednimi naslovi. Takoj smo ugotovili, da je cel niz podatkov o naslovu pravzaprav element za izločanje, in ne za združevanje zapisov. Nato smo iz niza podatkov o naslovu izločili tiste elemente, ki nam povezovanje preprečujejo. Pri kasnejših analizah smo določili, katere manjše nize podatkov naj upoštevamo, vendar ne kot splošni kriterij, ampak kot

¹ Dostopen na <http://www.worldcat.org/>.

² Dostopen na <http://libris.kb.se/>.

³ Dostopen na <https://bib.uclouvain.be/opac/ucl/>.

⁴ Dostopen na <http://zoeken.oba.nl/>.

posamezne kriterije za različne vrste gradiva. Analize, ki smo jih naredili z namenom identifikacije izraznih oblik, prav tako niso dale neposredno dobrih rezultatov, razen pri prevodih. Drugih odnosov, kot so skrajšave, revizije ali priredbe, nismo mogli natančno identificirati. Opombe v bibliografskih zapisih, s katerimi naj bi bili ti odnosi opisani, namreč niso del katalogizacijske prakse. Ugotovili smo, da se lahko zanašamo le na podatek o izdaji, ki omogoča združevanje, ne omogoča pa natančne identifikacije vrste odnosa med izraznimi oblikami istega dela. Pri pojavnih oblikah pa je bilo za nas pomembno, da iste izrazne oblike združimo neodvisno od njihovega nosilca. Že pri pripravi koncepta smo vedeli, da bomo različne nosilce enostavno prepoznali na osnovi kodiranih podatkov. Vendar niti ta naloga ni bila tako enostavna, saj so bibliografski zapisi kreirani s primarnim dostopom do medija in intelektualne vsebine, s čimer identifikacija iste intelektualne vsebine na različnih medijih ni zgolj vprašanje izbire osnovnih kodiranih podatkov.

Pri analizi druge skupine entitet, ki zajema osebe in korporacije, smo ugotovili, da dosledna identifikacija oseb preko normativnih zapisov ne bo mogoča, kljub relativno daljšemu obdobju uporabe normativnih zapisov za osebe. V bibliografskih zapisih namreč še zmeraj zasledimo osebna imena, ki niso pod normativno kontrolo.

Ob pripravi koncepta smo posebno pozornost namenili intelektualnim in umetniškim stvaritvam, ki so samostojni deli v okviru širših celot, in istim stvaritvam, ki so celote zase: npr. Bevkovi črtici *Pestrna*, ki je večkrat izšla samostojno, enkrat v zbirki črtic ter enkrat skupaj s črtico *Lukey in njegov škorec* (glej Sliko 3). Glede na velikost katalogov ter na obseg domače založniške produkcije smo se odločili, da samostojne dele širših celot, ki so v dokumentu *Final Report of the Working Group on Aggregates* (2011, str. 3) poimenovane *združene zbirke izraznih oblik* (angl. *aggregate collection of expressions*), obravnavamo enako kot celote. Predvidevali smo, da bo uporabnikom pri iskanju določene stvaritve vseeno, ali je objavljena samostojno ali v okviru širše celote. Ob tem smo se omejili na dela, ki jih je mogoče identificirati s podatkom o naslovu, ki se v formatu COMARC zapisuje v samostojnih ali v vgrajenih poljih. Organska dela (Funkcionalne zahteve za bibliografske zapise, 2000, str. 78), ki so v bibliografskih zapisih praviloma opisana v opombah ali v poljih za fizični opis, so ostala zunaj našega koncepta.

Komplete monografskih publikacij s skupnim naslovom »Zbrana dela«, »Izbrana dela« in podobnimi izrazi v drugih jezikih smo iz praktičnih razlogov obravnavali kot novo delo. Prvi razlog je kvaliteta podatkov v bibliografskih zapisih, drugi pa tveganje slabo preglednih skupkov, kadar skupki združujejo več sto zapisov. Če bi jih obravnavali enako kot monografije, kjer se primerjajo tudi naslovi samostojnih del v okviru širših celot, bi v skupkih na osnovi množice naslovov iz vgrajenih

polj dobili obsežne sezname zapisov (glej poglavje 5.1). Tako smo dosegli, da smo med sabo primerjali samo zbrana in izbrana dela.

Vse opravljene analize so nam pomagale pri odločanju, s kakšnim modelom bomo začeli prestrukturiranje kataloga, ter pri iskanju prave mere med željo, da model FZBZ implementiramo v COBISS+, in dejanskim stanjem v katalogu, kjer je nekaj milijonov zapisov, ki že desetletja nastajajo z različnimi nameni in različnim odnosom do izčrpnega navajanja podatkov. Zavedali smo se, da dosledna implementacija tudi v omejenem obsegu ne bo mogoča, kar je med drugim potrdil vpogled v tuje knjižnične kataloge (glej poglavje 2).

4 Identifikacija skupkov

Zapise, ki pripadajo istemu delu, smo oblikovali v skupke in jih prikazali v zavihku »Vse izdaje in prevodi« (Slika 1). Delo smo identificirali z enotnim naslovom, pri zapisih brez enotnega naslova pa z glavnim naslovom, s primarnim avtorjem in alternativnimi avtorji, bibliografskim nivojem in vrsto zapisa. Naknadna testiranja so pokazala veliko napak pri združevanju zapisov za anonimna dela, ki nimajo podatkov o odgovornosti. Zato smo v skupke vključili le zapise z enotnim naslovom, pri zapisih brez enotnega naslova pa le tiste z nekaterimi vrstami sekundarne odgovornosti. Samo pri slikovnem in kartografskem gradivu, kjer je večina zapisov brez podatka o odgovornosti, smo delo identificirali po naslovu.

Osnovni podatki Podrobni podatki Vse izdaje in prevodi

Povest o knjigah
Capuder, Andrej
esej
2., pregledana izd. - Ljubljana : Slovenska matica, 2015
slovenski
ISBN 978-961-213-250-7 : 28,00 EUR
COBISS.SI-ID: 281851392

Išči dalje
Avtor
Capuder, Andrej |
Drugi avtorji
Capuder Grce, Neva |
Teme
Knjige | Eseji | Književnost | Branje | Življenjska obdobja |

Vse izdaje in prevodi:

Naslov	Avtor	Vrsta gradiva	Jezik	Leto
Povest o knjigah	Capuder, Andrej	esej	slovenski	2014

Slika 1: Prikaz skupka »Vse izdaje in prevodi« za Capudrovo delo *Povest o knjigah*

Pri različnih testiranjih smo ugotovili, da ne moremo uporabiti enotnega algoritma za vse vrste virov, ker ima vsaka skupina svoje posebnosti. Zato smo pri združevanju zapisov pripravili celo vrsto različnih algoritmov za posamezne vrste virov, in sicer za monografije, kartografsko in slikovno gradivo, videoposnetke, filme, notne zapise, glasbo in serijske publikacije. Iz združevanja smo izločili

članke, rokopise, integrirne vire in zbirne zapise, saj združevanje v teh primerih ni smiselno.

- **Monografije** smo ločili na leposlovje in neleposlovje. Pri leposlovju smo se odločili samo za zapise z individualnim primarnim avtorjem, iz skupine leposlovja pa smo izločili gledališke liste, programe, radijske igre, scenarije in priredbe. Pri ugotavljanju dela smo upoštevali le glavni naslov, ne pa tudi dodatkov k naslovu in drugih delov naslova. V skupke smo vključili tako samostojna dela kot samostojne dele širših celot (Slika 3).

Vse preostale monografije smo obravnavali kot neleposlovje. Iz skupine smo izločili zapise brez podatka o odgovornosti ter zapise z različnimi, manj pomembnimi vrstami sekundarne odgovornosti; izvzeli smo prireditelje, zbiratelje in urednike. Pri identifikaciji dela smo upoštevali vse dele naslova razen vzporednih oblik. Previde del pa smo določili preko originalnega naslova oziroma opombe o prevodu (Slika 2).

Naslov	Avtor	Vrsta gradiva	Jezik	Leto
The galley slave	Jančar, Drago	roman	angleški	2011
Galerjik : roman	Jančar, Drago	roman	beloruski	1990
Galejnik	Jančar, Drago	roman	češki	1990
Kaškyn : roman	Jančar, Drago	roman	kazaški	1987
A gályarab	Jančar, Drago	roman	madžarski	1985
Galiot : [roman]	Jančar, Drago	roman	makedonski	1984
Der Galeot	Jančar, Drago	roman	nemški	1991
Der Galeot : Roman	Jančar, Drago	roman	nemški	2004
Der Galeerensträfling : Roman	Jančar, Drago	roman	nemški	2015
De galeislaaf	Jančar, Drago	roman	nizozemski	1995
Galerjik	Jančar, Drago	roman	poljski	1988
Dorotej : roman. Galerjik : roman. Podzemna reka : roman Vsebina: Galjot	Nenadić, Dobrilo	roman	ruski	1982
Galjot	Jančar, Drago	roman	slovenski	1984
Galjot	Jančar, Drago	roman	slovenski	2004
Galjot : roman	Jančar, Drago	roman	slovenski	1980

Slika 2: Slovenske izdaje in prevodi Jančarjevega romana Galjot

Naslov	Avtor	Vrsta gradiva	Jezik	Leto
Pestna [Zvočni posnetek]	Bevk, France	zvočna kaseta	slovenski	[med 1990 in 1993]
Pestna	Bevk, France	kratka proza	slovenski	1952
Otroci samote : [Izbor iz Bevkovih mladinskih del] Vsebina: Pestna	Bevk, France	kratka proza	slovenski	1963
Pestna	Bevk, France	kratka proza	slovenski	1965
Pestna	Bevk, France	kratka proza	slovenski	1967
Lukey in njegov škorec : Pestna	Bevk, France	kratka proza	slovenski	1974

Slika 3: Bevkova črtica Pestna je objavljena samostojno ali skupaj z drugimi deli; tiskane izdaje in zvočna izdaja so izenačene

- Skupino **zbranih in izbranih del** je sicer mogoče prepoznati glede na kodo za vrsto vsebine, se pa ta koda uporablja nedosledno. Zato smo skupino zbranih in izbranih del oblikovali na osnovi individualnega primarnega avtorja in seznama generičnih izrazov, kot so zbrana dela, izbrana dela, dela ipd. v desetih različnih jezikih. Zapisi se združujejo za vsak jezik posebej, in sicer za slovenščino, srbsščino, bosanščino, makedonščino, bolgarščino, albanščino, nemščino, angleščino, francoščino, ruščino in italijanščino. Pri določanju dela smo upoštevali samo glavni stvarni naslov in oznako podrejenega dela.
- Pri **kartografskem in slikovnem gradivu** smo pri določanju dela upoštevali samo naslove. Glavne stvarne naslove in vzporedne naslove smo obravnavali enakovredno, npr. zapis za delo z naslovom Novo mesto = Rudolfswert se je združil z zapisom za delo, ki ima v naslovu samo Novo mesto, in z zapisom za delo, ki ima v naslovu samo Rudolfswert.
- Pri **videoposnetkih in filmih** smo izločili zapise brez podatka o odgovornosti ter zapise z različnimi, manj pomembnimi vrstami sekundarne odgovornosti, nismo pa izločili režiserjev, umetnikov, bibliografskih predhodnikov, scenariстов, avtorjev snemalne knjige in skladateljev. Tudi tukaj smo glavne stvarne naslove in vzporedne naslove obravnavali enakovredno, npr. zapis za delo z naslovom Franklin = Franček se je združil z zapisom za delo, ki ima v naslovu samo Franklin, in z zapisom za delo, ki ima v naslovu samo Franček.
- Pri **notnem gradivu in glasbenih zvočnih posnetkih** smo uporabili podoben algoritem kot pri neleposlovju, vendar smo pri sekundarnih avtorjih upoštevali skladatelje, prireditelje, aranžerje, urednike, pisce besedil, libretiste, glasbenike, izvajalce, notografe, umetnike in bibliografske predhodnike. Pri glasbenih izvedbah nismo vsake izvedbe obravnavali kot novo delo, zato so se zapisi za izvedbe istega dela, ki so jih izvajali različni izvajalci, uvrstili v isti skupek.
- Pri **serijskih publikacijah** smo pri ugotavljanju dela upoštevali samo ISSN-L, pri prevodih pa ISSN izvirnikov in prevodov.

Izenačili smo izdaje na različnih medijih (ali nosilcih), kot so tiskane in elektronske oblike (Slika 4), pri monografijah pa tudi zvočne izdaje (Slika 3), izdaje v Braillovi pisavi in izdaje s povečanim tiskom.

Naslov ▼	Avtor ↕	Vrsta gradiva ↕	Jezik ↕	Leto ↕
Noč v Reykjaviku	Švigelj-Mérat, Brina	 roman	slovenski	2014
Noč v Reykjaviku [Elektronski vir]	Švigelj-Mérat, Brina	 e-knjiga	slovenski	2014

Slika 4: Tiskana in elektronska izdaja romana *Noč v Reykjaviku* sta izenačeni

5 Analiza rezultatov

5.1 Statistični podatki

Prvi statistični podatki o združevanju bibliografskih zapisov v skupke so nam pokazali, da smo z algoritmom za povezovanje zajeli 2.690.414 zapisov, kar predstavlja nekaj več kot polovico zapisov v bazi podatkov COBIB.SI (Preglednica 1).

Preglednica 1: Število pojavnih oblik v bazi podatkov COBIB.SI glede na vrsto vira (stanje decembra 2016)

Vrsta vira/ število pojavnih oblik	Karto- graf- sko g.	Slikov- no g.	Video g.	Notno g.	Glas- beno g.	Zbrana /izbra- na dela	Lepo- slovje	Nelepo- slovje	Serij- ske publi- kacije	Skupaj
1	12960	31188	24003	22221	50003	30768	198611	1706249	73996	2149999
2–3	3485	6425	9255	2383	8327	3191	68311	203415	46250	351042
4–5	906	1995	2355	489	987	510	23731	38819	251	70043
6–8	767	1530	638	348	444	231	18135	20760	79	42932
9–19	745	2689	227	604	562	142	24511	15825	32	45337
20–49	417	2576	0	135	101	24	13640	2601	1	19495
50–499	497	3273	0	62	9	0	6497	326	0	10664
>500	0	2	0	0	0	0	1	0	0	3

Kot v drugih katalogih (Bennett idr., 2003) ima tudi v bazi podatkov COBIB.SI 20 % identificiranih del več kot eno pojavno obliko, največ oziroma 44 % pri leposlovju (razen pri zbranih/izbranih delih), najmanj oziroma 14 % pa pri neleposlovju. Nekoliko preseneča relativno slabo povezovanje zapisov za notno gradivo in glasbena dela, kjer je približno 16 % del izraženih z več kot eno pojavno obliko, kar je morda posledica podatkov v bibliografskih zapisih ali algoritma. 17,2 % del ima med dve in osem pojavnih oblik, najmanj oziroma 13 % pri notnem gradivu in neleposlovju, največ oziroma 39 % pa pri serijskih publikacijah. Več kot osem pojavnih oblik ima 2,4 % del, najmanj oziroma 0,6 % pri videoposnetkih, največ oziroma 10 % pa pri leposlovju (razen pri zbranih/izbranih delih) in slikovnem gradivu. Med 50 in 500 pojavnih oblik ima 0,4 % del, največ oziroma 6,6 % pri slikovnem gradivu, medtem ko zbrana/izbrana dela, glasbena dela in serijske publikacije v tej skupini sploh nimajo pojavnih oblik. Več kot 500 pojavnih oblik se pojavi le izjemoma, saj jih imajo le tri dela, dve pri slikovnem gradivu in eno pri leposlovju. Deleži pojavnih oblik pri različnih vrstah gradiva med sabo niso čisto primerljivi, ker smo za vsako vrsto gradiva uporabili različne algoritme (glej poglavje 4). Predvsem odstopajo algoritmi za zbrana/izbrana dela, kartografsko in slikovno gradivo ter serijske publikacije, kjer smo določili bolj poenostavljene modele kot npr. za monografije.

Pri oblikovanju koncepta smo se odločili za združevanje samostojnih del in del, ki so izšla v okviru celote. Iz tega koncepta smo izločili dela, ki so v naslovu opredeljena kot zbrana/izbrana dela, vendar pa so v skupku ostala dela, ki jih sestavlja veliko število posameznih del, niso pa označena s skupnim naslovom *zbrana dela*. Tako smo za knjigo z naslovom *The handy volume Shakespeare*, ki je izšla v 13 zvezkih (s 36 samostojnimi deli), v skupku dobili 734 zapisov. Podobne primere smo zasledili tudi pri slikovnem gradivu.

5.2 Ovire pri realizaciji koncepta

Pisanski in Žumer (2009) navajata, da so »frbrizirani prototipi sicer vnesli veliko mero reda v prikaze bibliografskega sveta. Napake so sicer zdaj bolj izpostavljene, predvsem v obliki napačno identificiranih entitet, ki pa niso posledica frbrizacije, temveč napak v obstoječih bibliografskih podatkih«. ⁵ Pri analizah in testiranju smo tudi mi naleteli na številne bibliografske zapise, ki pripadajo istemu delu, a se niso povezali zaradi pomanjkljivih podatkov ali napak v zapisih, različne uporabe katalogizacijskih pravil ali formata COMARC/B. Prav tako smo ugotovili, da smo v skupke zajeli nekatere neustrezne zapise, zato se nismo odločili za kompleksnejši algoritem, ker bi to upočasnilo hitrost procesa združevanja in prikaza podatkov. Združevanje zapisov namreč poteka ob vsakem iskanju in v skupke vsakič prinaša tudi najnovejše zapise.

Največja ovira pri realizaciji koncepta so bili bibliografski zapisi za dela, ki so izšla pod različnimi naslovi in ki praviloma nimajo enotnega naslova. Zato je določanje dela potekalo v nizu optimizacij podatkov o naslovu in odgovornosti, s katerimi smo poskušali nadomestiti enotne naslove.

Druga bistvena ovira je neenotna kvaliteta zapisov; zapisi so namreč nastajali v različnih časovnih obdobjih in z različnimi nameni, tako da so nekateri precej popolni, drugi pa zelo skopi. Tudi pristop h katalogizaciji se razlikuje od ene do druge vrste gradiva, saj se katalogizacijska pravila in tudi opis katalogizacijske prakse največkrat posvečajo besedilom, precej manj pa drugim vrstam virov. Zato smo bili prisiljeni skleniti kompromis med zastavljenim modelom in stanjem kataloga ter algoritme optimizirati po t. i. »povprečnem zapisu«.

Zaradi nepopolne normativne kontrole osebnih imen smo morali narediti še en kompromis in upoštevati le glavni del osebne značnice. Na enak način smo obravnavali tudi imena v vgrajenih poljih pri dodanih, prikritih in priključenih

⁵ Avtorja sta verjetno mislila napake v bibliografskih zapisih.

delih. Precej izgubljenih povezav je posledica neenotnega pristopa pri določanju avtorstva, kar je še posebej očitno pri predelavah. Tako je npr. v zapisu za zvočno knjigo *Samorastniki* Prežihov Voranc enkrat primarni avtor, prireditelj in režiser Jože Vozny pa sekundarni avtor, v drugem zapisu pa obratno. Precej nedoslednosti pri določanju avtorstva smo zasledili tudi pri glasbi in filmu, zato smo bili prisiljeni uporabiti manj natančen algoritem, posledica pa je nekoliko več napačnih povezav pri tovrstnem gradivu.

Pri združevanju del, ki so na različnih nosilcih, predvsem na nosilcih zvoka in slik, so bili precejšnja ovira kodirani podatki, težave pa so nastopile tudi zaradi manjkajočih podatkov za vrsto vsebine.

V skupke so se uvrstila samostojna dela, ki so izšla v okviru širše celote, kadar so navedena v bloku za povezovanje. Kadar so dela navedena le v opombi, pa jih ni bilo mogoče programsko povezati. Celote z veliko samostojnimi deli povzročajo obsežne skupke, ki so zato za uporabnike manj pregledni.

Težave so tudi pri povezovanju slikovnega in kartografskega gradiva, kjer je večina del anonimnih in se primerjajo samo naslovi. Pri slikovnem gradivu se zapisi povežejo na osnovi ujemanja naslovov, ne pa tudi na osnovi motiva. Motiv je naveden opisno, na osnovi tega pa ne moremo ločiti dela od nekega drugega dela. Kljub temu smo se odločili, da zapise povezujemo na osnovi ujemanja naslovov. Zato se predvsem pri razglednicah v skupku prikazujejo pojavne oblike drugih del.

Pri kartografskem gradivu nismo upoštevali merila in drugih značilnosti kartografskega prikaza, ki so vpisane zelo nekonsistentno.

6 Kako vidimo nadaljnji razvoj

Ko smo se odločili za pripravo skupka »Vse izdaje in prevodi«, smo se zavedali, da napak ne bo malo. Na kvaliteto avtomatske ekstrakcije zahtevne strukture FZBZ poleg zastavljenega modela vplivajo še kvaliteta zapisov in razvitost katalogizacijske prakse. Kljub vidno izboljšani kvaliteti novejših bibliografskih zapisov so stari zapisi in zapisi, ki so nastali z retrospektivno konverzijo listkovnih katalogov, trajna ovira kvalitetni frbrizaciji.

Bibliografsko tranzicijo COBISS/OPAC-a kratkoročno vidimo v izboljšanju in dopolnitvah uveljavljenih algoritmov za združevanje zapisov. Spremljanje statistike skupkov in dodatna analiza bibliografskih zapisov bosta usmerjena na:

- nadaljnjo selekcijo zapisov, ki niso primerni za povezovanje;
- dodatno ugotavljanje elementov v bibliografskih zapisih, ki vplivajo na prepoznavnost dela;
- ugotavljanje konkretnih pomanjkljivosti v bibliografskih zapisih, ki vplivajo na napake pri identifikaciji dela.

Prva cilja sta povezana z izboljšanjem uveljavljenih algoritmov, tretji cilj pa nas vodi k naslednji fazi, oplemenitvi bibliografskih zapisov z določenimi dodatnimi podatki, ki ne bi bili namenjeni samo identifikaciji dela, ampak tudi identifikaciji izraznih oblik. V prvi fazi bi želeli preučiti možnost programskega dodajanja povezav, ki obstajajo v bloku 4 formata UNIMARC. V obsegu podatkov, ki so v bibliografskih zapisih, bi povezali dela, ki so izšla na različnih medijih, različne izdaje dela na istem mediju, originalna dela in reprodukcije, pa tudi originalna in revidirana besedila. Obogatitev bibliografskih zapisov z definiranimi odnosi do drugih zapisov je pogoj za pripravo strukturiranih prikazov, ki bi jih v prvi fazi uporabljali le v skupkih »Vse izdaje in prevodi«.

Nadaljnji razvoj COBISS+ vidimo ne le v frbrizaciji starih bibliografskih zapisov, ampak tudi v sistematičnem uvajanju nove katalogizacijske prakse, utemeljene na novih načelih katalogizacije z bibliografskimi in normativnimi zapisi, ki upoštevajo konceptualne modele iz družine Iflinih FZBZ dokumentov. Zavedamo se, da vse to zahteva spreminjanje miselnosti pri velikem številu katalogizatorjev, ki delujejo v sistemu COBISS, hkrati pa pomeni tudi obvezo nacionalne knjižnice in nacionalnega bibliografskega servisa, da jim to novo miselnost približajo.

7 Zaključek

Združevanje bibliografskih zapisov v skupke je naš prvi poskus, da uporabnikom pri enem samem iskanju ponudimo tudi informacijo o vseh izdajah in prevodih iskanega dela. Upamo, da bomo z obogateno informacijo pomagali uporabnikom, da lažje in bolj vsestransko dostopajo do knjižničnih fondov. Zavedamo se, da je pred nami še veliko dela, tako glede izboljšanja združevanja kot tudi samih prikazov podatkov v katalogu. Kljub vsemu smo zadovoljni, da smo naredili prvi korak, saj smo s tem postavili smernice za nadaljnji razvoj frbrizacije katalogov COBISS.

Navedeni viri

Aalberg, T. in Žumer, M. (2013). The value of MARC data, or, challenges of frbrisation. *Journal of documentation*, 69(6), 851–872.

Bennet, R., Lavoie, B. F. in O'Neill, E. T. (2003). The concept of a work in worldcat: an application of FRBR. *Library collections, acquisitions, and technical services*, 27(1), 45–59.

Final report of the Working Group on Aggregates. (2011). Haag: IFLA. Pridobljeno 23. 3. 2016 s spletne strani: <http://www.ifla.org/files/assets/cataloguing/frbrwg/AggregatesFinalReport.pdf>

FRBR Display Tool, Version 2.0. (2009). Washington: Library of Congress. Pridobljeno 13. februarja 2015 s spletne strani: <http://www.loc.gov/marc/marc-functional-analysis/tool.html>

Funkcionalne zahteve za bibliografske zapise: končno poročilo. (2000). Ljubljana: Filozofska fakulteta, Oddelek za bibliotekarstvo.

Hickery, T. B. in Toves, J. (2009). *FRBR work-set algorithm. Version 2.0*. Dublin, Ohio: OCLC. Pridobljeno 15. 12. 2015 s spletne strani: <http://www.oclc.org/content/dam/research/activities/frbralgorithm/2009-08.pdf>

Pisanski, J. in Žumer, M. (2009). Funkcionalne zahteve za bibliografske zapise (FZBZ): analiza uporabnosti konceptualnega modela bibliografskega sveta. *Knjižnica*, 53(1–2), 61–76.

Andreja Krajnc Vobovnik

Institut informacijskih znanosti, Prešernova 17, 2000 Maribor
e-pošta: andreja.krajnc@izum.si

Gordana Mazić

Institut informacijskih znanosti, Prešernova 17, 2000 Maribor
e-pošta: gordana.mazic@izum.si