

PROJEKT IZBOLJŠANJE KAKOVOSTI ZAPISOV V COBIB.SI

Anka Rogina

Institut informacijskih znanosti, Maribor

Kontaktni naslov:
anka.rogina@izum.si

Izveček

Kakovost bibliografskih zapisov v vzajemni bibliografski bazi podatkov COBIB.SI in v vseh lokalnih bibliografskih bazah je temeljnega pomena za kakovostno delovanje celotnega sistema COBISS.SI. Da bi zagotovili čim večjo konsistentnost bibliografskih zapisov in programsko odpravili kar največ napak v zapisih, smo izvedli projekt Izboljšanje kakovosti zapisov v COBIB.SI. V prispevku so opisane analize, ki smo jih izvedli v okviru projekta, dopolnitve programske opreme COBISS2/Katalogizacija, ki smo jih predlagali v okviru projekta, ter konverzija podatkov v bibliografskih zapisih, ki je bila izvedena kot zadnja faza v projektu.

Ključne besede

kakovost bibliografskih zapisov, vzajemna bibliografska baza podatkov COBIB.SI, sistem COBISS.SI, analize bibliografskih zapisov, konverzija podatkov

Abstract

The quality of bibliographic records in the COBIB.SI shared bibliographic database and all local bibliographic databases is key to the high performance COBISS.SI system. In order to ensure the highest possible degree of consistency in the quality of bibliographic records and a software-based error correction in records we have conducted the COBIB.SI Bibliographic Record Quality Improvement project. The paper describes the analyses performed within the project and the proposed upgrading of COBISS2/Cataloguing software as well as data conversion in bibliographic records undertaken as the last project phase.

Keywords

quality of bibliographic records, COBIB.SI shared bibliographic database, system COBISS.SI, analyses of bibliographic records, data conversion

UVOD

Projekt *Izboljšanje kakovosti zapisov v COBIB.SI* je potekal od februarja 2005 do decembra 2006. Cilj projekta je bil v čim večji možni meri programsko odpraviti napake in pomanjkljivosti v bibliografskih zapisih, s tem zagotoviti boljšo kakovost lokalnih bibliografskih baz knjižnic in vzajemne bibliografske baze podatkov COBIB.SI ter uporabnikom omogočiti, da res najdejo tisto, kar iščejo.

Naše osnovno vodilo je bilo izboljšati čim več bibliografskih zapisov, vendar pri tem ne pokvariti obstoječih zapisov. Vse predloge sprememb v zapisih je bilo zato treba zelo natančno premisliti z vseh vidikov in upoštevati medsebojne vplive različnih predlogov. Marsikateri predlog smo prav zaradi teh razlogov tudi opustili ali izpeljali v okrnjeni obliki.

Končno dejanje je bila izvedba konverzije podatkov v bibliografskih zapisih v vseh lokalnih bazah podatkov in vzajemni bibliografski bazi podatkov v sistemu COBISS.SI.

ANALIZE

Delo na projektu se je začelo z zbiranjem predlogov sprememb in izboljšav v zapisih. Nekaj predlogov smo izoblikovali na osnovi sporočil uporabnikov o težavah pri delu, večino predlogov pa so prispevali kolegi, ki so se s posameznimi problemi srečevali že dalj časa. Probleme smo najprej popisali, nato pa smo začeli natančnejše analize z različnih vidikov: osredotočili smo se na probleme pri zapisih za posamezne vrste publikacij, na probleme pri zapisih določenega statusa pa tudi na probleme v posameznih delih zapisov (bibliografski del, podatki o zalogi).

Zapisi za serijske publikacije

V letu 2005 smo se najprej lotili zapisov za serijske publikacije. Tipična lastnost serijskih publikacij je, da se v obdobju izhajanja spreminjajo, zato je treba v bibliografskih zapisih spremembe beležiti. Žal se prepogosto dogaja, da zapisi ostanejo v prvotni verziji, brez ustreznih dopolnil. Želeli smo ugotoviti, ali bi bilo mogoče nedopolnjene zapise na kak način programsko dopolniti. Izvedli smo več analiz, ki so opisane v nadaljevanju.

Zapisi za serijske publikacije z vnesenim ISSN 0000-0000

Serijske publikacije so identificirane z ISSN. Če te identifikacije ne poznamo, je v zapis mogoče vnesti tudi 0000-0000, vendar je tak vnos manj primeren, saj močno zmanjša možnost natančnega iskanja, onemogočena pa je tudi analitična obdelava take serijske publikacije. Z analizo smo želeli ugotoviti, ali je mogoče serijske publikacije natančneje identificirati in podatke v zapisih programsko popraviti s pomočjo podatkov v bazi ISSN, tj. namesto 0000-0000 vpisati v zapise ustrezne številke ISSN iz baze ISSN.

Dne 19. 7. 2005 je bilo v vzajemni bazi podatkov COBIB.SI 11.094 zapisov, ki so v enem od podpolj, v katera vnašamo ISSN (tj. v podpoljih 011c, e, f, y, z) imeli vneseno vrednost 0000-0000. Za vse te zapise smo v bazi ISSN preverili, ali obstaja zapis ali več zapisov, ki se z zapisom iz baze podatkov COBIB.SI ujema v prvih 50 znakih naslova. Ugotovili smo, da je bilo v bazi ISSN 3.988 takih zapisov. Od teh zapisov je bilo 1.747 (tj. 15,75 % od vseh 11.094 zapisov) takih, za katere je v bazi ISSN obstajal le en zapis z ustreznim prvim delom naslova. Za ostalih 2.241 zapisov smo v bazi ISSN našli dva ali več zapisov z ustreznim prvim delom naslova. Z naključnim preverjanjem med 1.747 zapisi smo ugotovili, da tudi pri teh podatki niso dovolj zanesljivi, da bi lahko izvedli programsko vpisovanje (popravljanje) ISSN v zapise. Razlike, ki so se v zapisih pojavljale kljub enake-mu naslovu, so bile npr. različen jezik, različna obdobja izhajanja, različni podatki o založnikih, različni mediji (tiskani – elektronski) in podobno.

Na podlagi rezultatov je bilo očitno, da tudi bolj zapleten algoritem za identifikacijo ustreznih zapisov ne bi zado-stoval za programsko popravljanje zapisov oz. da bi z zapletenim algoritmom lahko popravili premalo zapisov. Sklenili smo, da programsko ne bomo popravljali teh zapisov.

Kontrola izhajanja

Precej zapisov za serijske publikacije v bazi podatkov COBIB.SI po prvem vnosu ni bilo vzdrževanih, kar pomeni, da ni nihče poskrbel, da bi ob spremembi izhajanja publikacije popravil tudi zapis. Predvsem se to odraža v

vzajemni bazi podatkov COBIB.SI, medtem ko v lokalnih bazah podatkov posamezne knjižnice vzdržujejo zapise za zanje pomembne publikacije. Podatki o izhajanju serijskih publikacij v bazi podatkov COBIB.SI in v bazi ISSN se zato ne ujemajo. Za serijske publikacije z verificiranim ISSN (številka ISSN vpisana v podpolju 011e) nas je zanimalo, v kolikšni meri se ujemajo oznake za leto izida publikacije (tj. serijska publikacija, ki še izhaja, ki je prenehala izhajati ali ima neznani status) v bazah podatkov COBIB.SI in ISSN.

Z analizo smo prišli do naslednjih podatkov: 19. 7. 2005 je bilo v bazi podatkov COBIB.SI 49.566 zapisov za serijske publikacije z verificiranim ISSN. Oznake za leto izida v bazi podatkov COBIB.SI in bazi ISSN se niso ujemale pri 6.645 (13,4 %) zapisih. Za 5.567 serijskih publikacij bi po podatku v bazi ISSN sklepali, da so serijske publikacije prenehale izhajati, medtem ko je v bazi podatkov COBIB.SI za te iste serijske publikacije podatek govoril, da serijska publikacija izhaja redno ali da njen status ni znan.

Po navzkrižnem preverjanju zapisov za nekaj naključno izbranih serijskih publikacij v različnih bazah podatkov (v OCLC-jevi bazi podatkov WorldCat, v bazi podatkov Nacionalne in sveučiliščne knjižnice iz Zagreba, v bazi univerzitetne knjižnice iz Varšave (<http://www.buw.uw.edu.pl/>) in v drugih lokalnih bazah podatkov se je izkazalo, da tudi podatki v bazi ISSN niso zanesljivi in ažurni. Nezaanesljivi so bili predvsem podatki za serijske publikacije v neangleških jezikih. Naj kot primer navedem našo “domačo” serijsko publikacijo: za serijsko publikacijo *COBISS obvestila*, ISSN 1318-8585, je v bazi ISSN bil še vedno podatek, da redno izhaja, medtem ko zapisa za *Organizacijo znanja*, ISSN 1580-979X v bazi ISSN še zmeraj ni bilo. Vemo, da je publikacija *COBISS obvestila* prenehala izhajati leta 2001 ter da jo je nasledila publikacija *Organizacija znanja*, kar velja tako za tiskano kot elektronsko izdajo.

Zaradi nezanesljivosti podatkov smo se odločili, da bibliografskih zapisov v bazi podatkov COBIB.SI na osnovi oznak za leto izida ne bomo avtomatsko popravljali s podatki iz baze ISSN.

Kontrola neverificiranih ISSN

Neverificirani ISSN v zapis za tujo serijsko publikacijo vpišemo, kadar je ISSN na publikaciji natisnjen, vendar ga v času kreiranja bibliografskega zapisa v bazo podatkov COBIB.SI še ne najdemo v bazi ISSN. Pogosto se zgodi, da zapisov za serijske publikacije po prvem vnosu katalogizatorji več ne dopolnjujejo. Zato nas je zanimalo, za koliko serijskih publikacij s številkami ISSN, ki so v bazi podatkov COBIB.SI navedene kot neverificirane, najdemo zapise v bazi ISSN.

Ugotovili smo, da je 19. 7. 2005 v bazi podatkov COBIB.SI obstajalo 1.008 zapisov za serijske publikacije z neverificiranim ISSN. Za 513 izmed teh je obstajal zapis v bazi ISSN. ISSN, ki ga najdemo v bazi ISSN, jemljemo kot verificiran, zato je bilo treba zapise v bazi podatkov COBIB.SI popraviti. Ker je 513 zapisov relativno malo, smo se namesto za konverzijo podatkov odločili za ročno dopolnjevanje in popravljanje. V vzajemni bazi podatkov COBIB.SI je bilo do 8. septembra 2005 dopoljenih in popravljenih 483 zapisov. Podatki za 30 zapisov niso bili zanesljivi, zato jih nismo spreminjali.

Kontrola vsebnosti polj za medsebojno povezovanje

Serijske publikacije se v času svojega "življenja" pogosto preoblikujejo. Pogosto spremenijo naslov, lahko se razcepijo na dve ali več novih publikacij, lahko se s kako publikacijo združijo in začnejo izhajati pod novim naslovom, lahko so prevedene v druge jezike ali natisnjene v različnih pisavah ali pa prirejene za različne medije. Vse te spremembe v bibliografskih zapisih beležimo v polja za medsebojno povezovanje v bloku 4XX.

Kot že pri prejšnjih analizah smo izhajali iz predpostavke, da so zapisi za serijske publikacije v bazi podatkov COBIB.SI slabo vzdrževani. Da bi jih posodobili, nas je zanimalo, ali bi jih bilo mogoče programsko dopolnjevati s podatki o medsebojnih odnosih med publikacijami, ki jih najdemo v zapisih v bazi ISSN.

Dne 19. 7. 2005 je bilo v bazi podatkov COBIB.SI 11.736 zapisov za serijske publikacije, ki so bili prevzeti iz baze ISSN in po prevzemu niso bili nikoli več dopolnjeni. Med njimi je bilo 1.446 takih, ki v bazi podatkov COBIB.SI niso imeli nobenega polja 4XX za medsebojno povezovanje, v bazi ISSN pa so imeli vsaj en podatek o povezavah. Skupno število zapisov, v katerih se število podatkov o povezavah v bazi podatkov COBIB.SI in v bazi ISSN razlikuje, je bilo 2.752 (23,4 %).

Preverili smo nekaj naključno izbranih zapisov. Ugotovili smo, da so razlike v poljih za medsebojno povezovanje tesno povezane z različnimi načini izhajanja serijskih publikacij (npr. podatki o naslovih nadaljevanj v zapisih za serijske publikacije, ki so prenehale izhajati). V primeru programskega popravljanja zapisov bi morali dopolniti in popraviti vse podatke v zapisih. Ker pa smo že prej ugotovili, da tudi podatki v bazi ISSN niso popolnoma zanesljivi, nismo mogli zagotoviti, da pri programskem popravljanju zapisov ne bi vpisovali napačnih podatkov. Zato smo se odločili, da podatkov o medsebojnem povezovanju serijskih publikacij ne bomo programsko popravljali.

Zapisi za Slovensko nacionalno bibliografijo

Ena od idej za izboljšanje kakovosti zapisov v vzajemni bazi podatkov COBIB.SI je bila, da bi podatke iz zapisov, ki so del Slovenske nacionalne bibliografije, iz lokalne baze podatkov NUK vpisali v zapise v vzajemno bazo podatkov COBIB.SI. Predpostavljali smo, da so ti zapisi v celoti dopolnjeni in popravljeni. Izvedli smo kratko primerjavo, za katero smo izbrali dva vzorca zapisov. Primerjavo smo izvedli tako, da smo pregledovali zapis po zapis. Prvi vzorec so predstavljali zapisi za monografske publikacije, drugega pa zapisi za neknjižno gradivo.

Zapisi za monografske publikacije

V NUK-ovi lokalni bazi podatkov smo dne 21. 3. 2006 za vzorec izbrali 200 zapisov za monografske publikacije, ki so ustrezale kriterijem za uvrstitev v nacionalno bibliografijo. Zapise smo primerjali v polnem (FULL) formatu v lokalni in vzajemni bazi.

Kratek povzetek ugotovitev po 200 pregledanih zapisih:

- Število zapisov, ki so bili v lokalni bazi podatkov NUK in v vzajemni bazi podatkov COBIB.SI popolnoma enaki, je bilo 77 (38,5 %).
- Zapisov, ki so bili po vsebini enaki, razlikovali so se samo po tem, da so bili v vzajemni bazi podatkov COBIB.SI zaklenjeni (LOCK1), je bilo 6 (3 %).
- V preostalih zapisih (117, tj. 58,5 %) je bilo razlik več. Razlike so se pojavljale tako v bloku 2XX (naslovi, založnik), 3XX (opombe), največ pa jih je bilo v bloku 6XX pri predmetnih oznakah in klasifikaciji ter pri avtorstvih v bloku 7XX. Razlike v avtorstvih izhajajo tudi iz tega, ker so bili nekateri zapisi povezani z normativnimi zapisi in so se zaradi sprememb v normativnih zapisih ter zaradi različne sinhronizacije normativnih zapisov (v vzajemni bazi podatkov COBIB.SI in v lokalnih bazah podatkov) pojavile razlike v kazalkah. Razlike v polju 200 (naslovi) in pri avtorstvih (značnice) so bile take, da samo iz zapisov nismo mogli sklepati, ali je "pravilnejši" zapis v lokalni ali v vzajemni bazi podatkov. Prav tako ni bilo mogoče na splošno zaključiti, da so bili podatki v NUK-ovi lokalni bazi podatkov popolnejši in boljši.
- Med pregledanimi zapisi je bilo 6 takih, ki so jih v vzajemni bazi podatkov COBIB.SI zaklenili OSIC-i, ker so verificirali tipologijo del.

Na osnovi pregledanega vzorca zapisov ni bilo mogoče najti smiselnega algoritma za dopolnjevanje zapisov v vzajemni bazi podatkov COBIB.SI. Ob upoštevanju načela, da obstoječih zapisov ne smemo pokvariti, smo se odločili, da teh zapisov programsko ne bomo spreminjali.

Zapisi za neknjižno gradivo

V NUK-ovi lokalni bazi podatkov smo dne 2. 3. 2006 za vzorec izbrali zapise, ki so jih v NUK-u v letu 2005 kreirali za neknjižno gradivo in ki so hkrati ustrezali kriterijem za uvrstitev v nacionalno bibliografijo.

Rezultat je bilo 1.065 zapisov. Primerjali smo jih v polnem formatu v lokalni in vzajemni bazi podatkov.

Zapise smo pregledali in beležili razlike po poljih.

Kratek povzetek ugotovitev po 1.065 pregledanih zapisih:

- Število zapisov, ki so bili v lokalni bazi podatkov NUK in v vzajemni bazi podatkov COBIB.SI popolnoma enaki, je bilo 1.004 (94,3 % od vseh zapisov).
- V preostalih 61 zapisih smo ugotovili naslednje razlike (61 zapisov je sedaj 100 %):
 - različna verzija zapisa: 37 zapisov (60,66 %),
 - v vzajemni bazi podatkov je zaklenjenih (LOCK1): 59 zapisov (96,72 %),
 - razlike samo v kodiranih podatkih v bloku 0XX in polju 100: 11 zapisov (18,03 %),
 - razlike v polju 200: 27 zapisov (44,26 %),
 - različni podatki o kodi namembnosti (podpolje 100e): 7 zapisov (11,47 %),
 - različni podatki o jeziku (polje 101): 10 zapisov (16,39 %),
 - različni drugi kodirani podatki (polja 105 do 135): 13 zapisov (21,31 %),
 - razlike v podatkih o založništvu (polje 210): 15 zapisov (24,59 %),
 - razlike v podatkih o fizičnem opisu (polje 215): 14 zapisov (22,95 %),
 - razlike v opombah (blok 3XX): 15 zapisov (24,59 %),
 - razlike v predmetnih oznakah in klasifikaciji (blok 6XX): 18 zapisov (29,51 %),
 - razlike v avtorstvu (blok 7XX): 23 zapisov (37,70 %).

Glede na to, da je bilo od 1.065 zapisov v lokalni in vzajemni bazi podatkov le 61 zapisov (5,7 %), programsko popravljanje teh zapisov ni bilo smiselno.

Predhodni nepopolni zapisi (CIP-zapisi)

V vzajemni bazi podatkov COBIB.SI obstaja večje število bibliografskih zapisov, označenih kot predhodni nepopolni zapis oz. bolj znanih kot CIP-zapis. Kot že samo ime pove, nastane zapis "predhodno", tj. pred izidom publikacije in je zato seveda nepopoln. Zapis naj bi ta status imel, dokler knjižnica ne dobi fizičnega izvoda publikacije. Takrat je treba zapis dopolniti in mu spremeniti status, dodajo pa

se tudi podatki o zalogi. V skladu z načelom vzajemnosti, ki velja pri katalogizaciji v sistemu COBISS, naj bi prva knjižnica, ki dobi publikacijo, dopolnjen zapis shranila tudi v vzajemno bazo podatkov COBIB.SI.

V realnem življenju prva knjižnica, velikokrat pa tudi naslednje, zapis dopolni v lokalni bazi podatkov, na vzajemnem nivoju pa ga ne spremeni. Tako imamo v vzajemni bazi podatkov COBIB.SI še vedno nepopolne podatke zapisa s statusom CIP, vsaka naslednja knjižnica, ki zapis prevzame v svojo lokalno bazo podatkov, pa mora podatke sama dopolnjevati. Tako dobimo veliko različic istega zapisa v lokalnih bazah podatkov. Problemi, ki se zaradi tega pojavijo, so: nepotrebno multipliciranje časa za obdelavo (ni učinka vzajemnosti), večja možnost kreiranja duplikatov zaradi nepopolnosti podatkov v vzajemni bazi podatkov in ne nazadnje oviranje uporabnikov pri iskanju in pridobivanju želenega gradiva zaradi nepopolnih podatkov v vzajemni bazi in različnih podatkov v lokalnih bazah podatkov.

Tudi tokrat smo nameravali podatke iz zapisov iz lokalne baze podatkov NUK programsko vpisati v zapise v vzajemni bazi podatkov COBIB.SI. Predvidevali smo namreč, da ima NUK kot prejemnik obveznih izvodov publikacij v svoji lokalni bazi podatke že popravljene. Po preverjanju se je izkazalo naslednje:

- Dne 25. 7. 2006 je bilo v vzajemni bazi podatkov COBIB.SI kot CIP-zapis označenih 9.638 zapisov. Od teh je 3.271 zapisov (33,9 %) imelo podatke o zalogi.
- V lokalni bazi podatkov NUK je bilo dne 25. 7. 2006 kot CIP-zapis označenih 8.645 zapisov.
- Od 9.638 zapisov, ki so bili v COBIB.SI označeni kot CIP, jih je v NUK-ovi lokalni bazi podatkov obstajalo 6.748 (70 %). Od teh 6.748 zapisov jih je v lokalni bazi podatkov NUK samo 220 (3 %) imelo status različen od CIP.

Iz zadnjega podatka sledi, da konverzija podatkov iz zapisov, ki so v lokalni bazi podatkov NUK že dopolnjeni, v zapise v vzajemni bazi podatkov COBIB.SI ni smiselna zaradi premajhnega števila dopolnjenih zapisov.

Problema CIP-zapisov se zavedamo in smo ga želeli zmanjšati s pomočjo programskih kontrol pri katalogizaciji. Programska kontrola, ki opozori na status zapisa, ko vnesemo inventarno številko, je zaradi različnih pooblastil, ki jih imajo katalogizatorji, le opozorilo in jo katalogizatorji lahko tudi ignorirajo. Ko smo razmišljali, kako odpraviti problem nepopolnih zapisov, smo spoznali, da programsko popravljanje zapisov ne pride v poštev ter da problema ni mogoče odpraviti brez sodelovanja katalogizatorjev. Sklenili smo poskusiti z mesečnim obveščanjem in prošnjo, naj katalogizatorji dopolnijo zapise za tiste publikacije, za katere v knjižnici že imajo zalogo, zapis pa ima še vedno status CIP-zapisa.

Analiza vsebine v nekaterih poljih v bibliografskih zapisih

Z razvojem programskega segmenta COBISS2/Katalogizacija pa tudi z razvojem katalogizacijskih pravil je v različnih obdobjih prihajalo do različnih sprememb, tako glede pomena ali načina uporabe posameznih kod v poljih s kodiranimi podatki, kot tudi do sprememb pri pravilih vnosa v format COMARC/B. Spremembe so se navadno uveljavile od določenega datuma dalje, medtem ko so že kreirani zapisi ostajali večinoma nespremenjeni.

Z analizo vsebine nekaterih polj v bibliografskih zapisih smo želeli poiskati čim več polj, ki so doživela spremembe, in preveriti, v kolikšni meri je mogoče podatke poenotiti. S konverzijo podatkov v vseh bibliografskih bazah, tako vzajemni kot lokalnih, smo želeli doseči predvsem večjo konsistentnost podatkov v zapisih.

Pripravili smo več predlogov za konvertiranje podatkov z namenom, da jih pred realizacijo predstavimo knjižnicam, katalogizatorje pa prosimo za mnenje in pripombe.

Uporaba kode "d" v podpolju 105a

S kodo v podpolju 105a opisujemo ilustracije v publikaciji. V COMARC/B je bila koda "d" definirana kot "grafični prikazi" in naj bi se uporabljala za tabele, sheme, diagrame itd. Po UNIMARC ima koda "d" pomen "navigacijske karte" (angl. "charts – Special purpose maps designed for use of navigators"). Neskladje najverjetneje izhaja iz različnih tolmačenj pomena in uporabe kode. Da bi se izognili prihodnjim nesporazumom, smo predlagali, da se v bodoče koda "d" dosledno uporablja v pomenu "navigacijske karte", s čimer smo vnos uskladili z vnosom po UNIMARC. Za opis dejstva, da publikacija vsebuje tabele, sheme, diagrame itd. bomo v prihodnje uporabljali pomensko širšo kodo "a" – *ilustracije*.

Taka rešitev je zahtevala definiranje konverzije kode "d" v kodo "a" v podpolju 105a. Seveda so se takoj pojavili pomisleki o tem, v kakšnem pomenu so katalogizatorji kodo "d" vnašali doslej. Preverili smo vnose v bazi podatkov COBIB.SI s kombinacijo kode za grafične prikaze in ključnimi besedami oz. vrstilci UDK s področja navigacije. Ugotovili smo, da je bilo 9. 8. 2006 v bazi podatkov COBIB.SI 151.772 zapisov s kodo "d". Ko smo iskanje zožili na področje navigacije, smo našli 406 zapisov. Te zapise smo pregledali v polnem formatu in ugotovili, da katalogizatorji kode "d" doslej niso uporabljali za navigacijske karte. Ugotovitev je pomenila, da je predvidena konverzija kod mogoča.

Brisanje kode "y" iz podpolja 105f

V podpolju 105f vnašamo kodo za literarno vrsto. Decembra 1998 je bila ukinjena koda "y" – *neliterarna besedila*.

Kode od takrat ni mogoče več vnesti, v bazi podatkov COBIB.SI pa je avgusta 2006 obstajalo še 49.518 "starih" zapisov, v katerih je ta koda še bila vnesena. Koda je bila umaknjena iz šifranta za vnos, s tem pa je bila umaknjena tudi razrešitev kode. Posledica je bila, da se pri "starih" zapisih v polnem formatu ni več izpisal pomen kode.

Da bi poenotili podatke v zapisih, smo predlagali, da v vseh zapisih iz podpolja 105f brišemo kodo "y".

Vnos v podpolje 115b

V podpolje 115b vnašamo čas trajanja filmov in videoposnetkov v minutah. Po UNIMARC je obvezen vnos treh števk, zato po potrebi vnesemo tudi vodilno ničlo.

Pri pregledovanju zapisov smo ugotovili, da so podatki v 231 zapisih napačno vneseni. Namesto treh števk jih je bilo vnesenih manj, ponekod pa so bile vnesene tudi črke. Da se napačni vnosi ne bi nadaljevali, smo takoj uvedli programsko kontrolo vnosa v podpolje 115b. Predvideli smo tudi programsko popravljanje obstoječih napačnih vnosov tako, da bomo po potrebi vnesli vodilno ničlo.

Poenotenje polj za opombe

Polja 304 – *Opomba k naslovu in navedbi odgovornosti*, 305 – *Opomba k izdaji in bibliografski preteklosti* in 306 – *Opomba k založništvu, distribuciji itd.* smo uporabljali samo v zapisih za neknjižno gradivo. V zapisih za druge vrste gradiva smo opombe z istim pomenom vnašali v polje 300 – *Splošna opomba*.

Najprej smo preverili, v kolikšnem obsegu so katalogizatorji doslej uporabljali našeta polja. Ugotovili smo, da je le okoli 5.400 zapisov od približno 3 milijonov zapisov v bazi podatkov COBIB.SI vsebovalo ta polja. To predstavlja le 0,18-odstotni delež zapisov. Da bi poenotili in poenostavili vnos opomb za vse vrste gradiva, smo predlagali, da podatke iz polj 304, 305 in 306 prenesemo v polje 300, nato pa polja 304, 305 in 306 ukinemo.

Konverzija kod v podpolju 500m v razrešitev kod

V podpolje 500m smo vnašali kodo jezika publikacije, kadar je bil jezik potreben kot del značnice. Podpolje m v polju 500 ni ponovljivo, kar pomeni, da je bilo mogoče vnesti samo eno kodo jezika. Tak vnos v primerih, ko je publikacija npr. vzporedno natisnjena v več jezikih, ni bil ustrezen. Prav tako vnos kode za jezik namesto izpisane oblike jezika ni bil v skladu s pravili vnosa po UNIMARC.

Da bi našete nepravilnosti odpravili, smo predlagali, da kode, ki so bile vnesene v podpolje 500m, nadomestimo z njihovimi razrešitvami po veljavnem šifrantu.

Ukinitev podpolja 610b

Podpolje 610b je bilo v format COMARC/B dodano za potrebe izpisa bibliografije Jugoslavije in je bilo namenjeno formalnim oznakam. Polje ni bilo v skladu s formatom UNIMARC, bibliografije Jugoslavije pa tudi ni več. V podpolje 610b so bili podatki vpisani le v 680 zapisih. Predvideli smo, da podatke iz podpolj 610b prenesemo v podpolja 610a, podpolje 610b pa ukinemo.

Konverzija kode "fik" v najbližji ustrezen vrstilec v podpolju 675c

Vnos vrstilca UDK po šifrantu v podpolje 675c – *Vrstilec za iskanje* je obvezen. Ker imajo nekatere knjižnice vsebinsko obdelavo ločeno od formalne, je v podpolje 675c mogoče vnesti tudi fiktivno kodo "fik", ki naj bi jo katalogizatorji pri vsebinski obdelavi nadomestili z ustreznim vrstilcem za iskanje, kar pa se v praksi ne izvaja dosledno. Podpolje 675c služi predvsem za iskanje, pri čemer pa fiktivna koda ne koristi kaj dosti, nasprotno, uporabnike zavaja in jim onemogoča priklic podatkov o želenem gradivu.

Ugotovili smo, da je bil 12. 7. 2006 v COBIB.SI 20.801 zapis, ki je imel v podpolju 675c vneseno kodo "fik", hkrati pa je vrstilec v podpolju 675a – *Vrstilec* vsebovalo 6.490 zapisov, kar je približno 31 %. To je praktično pomenilo, da je bilo približno eno tretjino zapisov nemogoče najti na osnovi vrstilca za iskanje, čeprav so imeli vpisan širši vrstilec UDK.

Da bi izboljšali najdljivost teh zapisov, smo definirali konverzijsko tabelo za tiste zapise, v katerih je poleg kode "fik" v podpolju 675c obstajalo vsaj eno podpolje 675a z vnesenim vrstilcem UDK. Predlagali smo, da v zapisih nato kodo "fik" nadomestimo z ustreznim vrstilcem za iskanje.

Splošna konverzija vrstilcev v podpolju 675c

V podpolje 675c vnašamo vrstilec, ki je enoten v sistemu COBISS.SI in je namenjen iskanju v vzajemni bazi podatkov COBIB.SI in v lokalnih bazah knjižnic. Vrstilci, ki jih je mogoče uporabiti, so zbrani v šifrantu. Ker so se v zadnjih nekaj letih vrstilci UDK kar precej spremenili, je bilo novejšemu stanju treba prilagoditi tudi šifrant vrstilcev za iskanje. Najbolj pereči so bili vrstilci za religije, ki so se preimenovali, zaradi česar je bilo treba izvesti tudi konverzijo teh vrstilcev.

Potrebno konverzijsko tabelo in nov šifrant so nam posredovali kolegi iz NUK. Po nekaj usklajevanjih in manjših popravkih konverzijske tabele in šifranta smo bili pripravljene na predvideno splošno konverzijo vrstilcev za iskanje.

Konverzija podpolj 675bsu v polja za nacionalno rabo

Podpolja 675bsu so bila dodana v format COMARC/B in niso v skladu s formatom UNIMARC. Podatki v podpoljih 675b – *Skupina*, 675s – *Statistika* in 675u – *UDK za lokalne kataloge* so lokalnega značaja in se večinoma vnašajo iz lokalnih šifrantov posameznih knjižnic. Ker so podatki lokalni, se razlikujejo v večini lokalnih baz podatkov, v vzajemni bazi podatkov COBIB.SI pa imajo neko vrednost, pač tisto, ki jo je vnesla knjižnica, ki je zapis zadnja shranila v vzajemno bazo podatkov.

Glede na tako nekonsistentne podatke se je izoblikovala zahteva, da naj podatke lokalnega značaja iz podpolj 675bsu prenesemo v neka nova polja v bloku 9XX, tj. v bloku podatkov za nacionalno rabo.

Kreiranje novih polj v bloku 9XX in nato prenos podatkov načeloma niso predstavljali problema. Ker pa podpolja 675bsu vsebujejo vrstilce UDK, ki so namenjeni izpisovanju bibliografij, raznih statističnih preglednic in drugih izpisov za lokalne potrebe, so omenjena podpolja močno povezana s programskim segmentom COBISS2/Izpis. Izkazalo se je, da bi sprememba polja, v katerega se vpisujejo podatki, vplivala na polovico vseh izpisov. Popravljanje programske opreme za vse te izpise bi zahtevalo veliko časa, zato smo predlagali, da se ta konverzija izvede kasneje.

Konverzija kod za vrsto avtorstva

Kode za vrsto avtorstva vnašamo v podpolja 7XX4 v skladu s šifrantom, ki je enoten v sistemu COBISS. Večina kod v šifrantu je v skladu s kodami, ki jih predpisuje UNIMARC, težimo pa k temu, da bi šifranta popolnoma poenotili. V zadnjih desetih letih je bilo precej kod spremenjenih, dodanih ali ukinjenih. Vse te spremembe smo sproti vnašali v šifrant, vendar pa so zapisi, ki so bili kreirani pred posameznimi spremembami, obdržali stare kode. Da bi poenotili podatke v vseh zapisih, ne glede na datum kreiranja, smo predvideli konverzijo starih kod v nove, kjer je bilo seveda to mogoče.

Zapisi z neosebni kreatorji

Pri pregledovanju zapisov smo naleteli na nekaj zapisov, katerih kreator ni bil določena oseba, temveč so imeli kot kreatorja navedeno neosebno uporabniško ime "DEFAULT". V skladu s projektom uvajanja enoznačnih uporabniških imen ter pooblastil je bilo treba neosebna uporabniška imena spremeniti v osebna. Neosebna uporabniška imena smo v dogovoru s posameznimi knjižnicami spremenili v osebna in tako bibliografske zapise, kreirane pod neosebni uporabniškim imenom pripisali enemu izmed katalogizatorjev v knjižnici.

Podatki o stanju zaloge

Podrobni podatki o stanju zaloge se beležijo za vsako knjižnico v njeni lokalni bazi podatkov, na vzajemni nivo pa se prenašajo le zbirni podatki. Programsko poseganje v podatke o zalogi za celoten sistem je odgovorno in nevhvaležno početje. Tokrat je bilo nujno, saj smo le tako lahko zagotovili, da po uvedbi nove denarne enote ne bo prihajalo do napačnih interpretacij podatkov.

Pripis oznake valute "SIT" k cenam brez valute

Podatke o cenah v podatkih o zalogi vnašamo v poljih 996, 997 in 998 v podpolje 3 – *Cena z računa* in v podpolje 0 – *Številka in datum predračuna*, v element C – *Cena predračuna*. Po pravilih za vnos se je oznaka valute vnašala v skladu z ustreznim šifrantom takoj za oznako podpolja oz. elementa podpolja. Po pravilih vnos oznake za domačo valuto ni bil potreben.

Da pri prehodu na valuto evro pri seštevkih cen ne bi prišlo do napak, je bilo treba v podpolja 3 oz. v element C podpolja 0, v katerih se je nahajal neki znesek brez valute, dodati oznako za domačo valuto, tj. "SIT". Pri tem je bilo treba paziti še na predpisan vnos podatkov, saj je v ponovljiva podpolja 3 namreč mogoče vnašati tudi podatke o popustih. Da v prihodnje ne bi več prihajalo do podobnih težav, smo spremenili način vnosa podatkov o cenah, tako da je treba sedaj, tudi v primeru domače valute, obvezno vnesti oznako valute.

Brisanje podatkov o zalogi v neslovenskih knjižnicah

Za vsak bibliografski zapis se v trenutku, ko ga posamezna knjižnica kopira v svojo lokalno bazo podatkov in mu doda podatke o zalogi, med podatke o zalogi na vzajemnem nivoju vpiše podatek o tem, katera knjižnica je zapis prevzela in koliko izvodov publikacije ima. Kot podatek o knjižnici se med te podatke vpiše sigla knjižnice. Podatki omogočajo uporabnikom, da zeleno gradivo iščejo v ustrezni knjižnici.

Način zapisovanja je enak že od začetkov sistema COBISS in je veljal že v času Jugoslavije, zato so med podatki o zalogi na vzajemnem nivoju bile vpisane tudi sigle knjižnic iz bivše Jugoslavije. Z razpadom Jugoslavije se je delovanje sistema omejilo najprej samo na Slovenijo, kasneje pa se je oblikovala mreža COBISS. Net. V času sprememb so se bistveno spremenili podatki o knjižnicah v sistemu, pa tudi o zalogi v teh knjižnicah. Predvidena konverzija podatkov je bila idealna priložnost za istočasno posodobitev podatkov o zalogi na vzajemnem nivoju. Predlagali smo, da se v sistemu COBISS.SI iz podatkov o zalogi na vzajemnem nivoju brišejo podatki o zalogah v neslovenskih knjižnicah.

Analize normativnih zapisov

Natančnejše analize samih normativnih zapisov se tokrat nismo lotili, obravnavali smo jih le v povezavi z bibliografskimi zapisi.

Normativni zapisi, označeni za brisanje

2. 8. 2006 je bilo v normativni bazi podatkov CONOR.SI 2.076 zapisov označenih za brisanje. To je predstavljalo le približno 0,5 % celotne baze.

Večina zapisov označenih za brisanje ne moti dela z normativnimi zapisi. Izjema so zapisi, ki so bili prevzeti iz baze podatkov LC Names in vsebujejo polje 035 – *Kontrolne številke zapisov iz drugih sistemov* ter so bili kasneje označeni za brisanje. V takih primerih namreč iz baze podatkov LC Names ni mogoče še enkrat prevzeti zapisa, katerega identifikacijska številka je že vsebovana v polju 035, kljub temu da je zapis označen za brisanje in bi potrebovali ustrezní normativni zapis.

Vseh zapisov, označenih za brisanje, ne smemo brisati iz baze podatkov CONOR.SI, ker so nekateri označeni zapisi še povezani z bibliografskimi zapisi v lokalnih bazah. Brisanje normativnih zapisov bi v praksi pomenilo, da bi iz bibliografskih zapisov, na katere so normativni zapisi povezani, izbrisali avtorje. Programsko je treba torej najprej preveriti, kateri izmed normativnih zapisov, označenih za brisanje, niso povezani z nobenim bibliografskim zapisom v nobeni lokalni bazi podatkov niti v vzajemni bazi podatkov COBIB.SI, šele nato smemo normativne zapise brisati iz baze podatkov CONOR.SI.

Nepovezani normativni zapisi

Zanimalo nas je, koliko je v normativni bazi podatkov CONOR.SI normativnih zapisov, ki niso povezani z nobenim bibliografskim zapisom v nobeni lokalni bazi podatkov niti v vzajemni bazi podatkov COBIB.SI. Predvideli smo, da bi jih morda bilo mogoče brisati in tako "počistiti" normativno bazo podatkov.

Seznam nepovezanih normativnih zapisov je bil narejen 23. 7. 2006, razdeljen pa je na dva dela. Prvi del zajema normativne zapise, ki so bili kreirani pred uvedbo normativne kontrole (tj. pred 22. 4. 2003, v fazi začetne priprave normativne baze podatkov) in vsebuje 923 zapisov. Drugi del zajema normativne zapise, ki so bili kreirani po uvedbi normativne kontrole (od 22. 4. 2003 dalje, ko je vnos normativnih zapisov omogočen vsem katalogizatorjem) in vsebuje 2.489 zapisov.

Zapise smo pregledali in preverili, ali v vzajemni bibliografski bazi podatkov COBIB.SI res ni zapisov za dela navedenih avtorjev.

Ugotovili smo, da so med nepovezanimi normativnimi zapisi tudi:

- Zapisi z napačno oblikovano značnico za osebna imena (npr. Andrej, Orozel). Ti zapisi pogosto predstavljajo duplikate k zapisom s pravilno oblikovano značnico.
- Zapisi, v katerih je v masko za osebno ime vneseno ime korporacije (npr. Zavod dr. Marijana Borštnarja, Dornava).
- Zapisi za avtorje, ki se pojavljajo v bibliografskih zapisih kot avtorji prikritih del (npr. Zorn, Hilarij).
- Zapisi za avtorje, ki se v bibliografskih zapisih pojavljajo kot avtorji, vendar bibliografski in normativni zapis nista povezana (npr. Zwitter, Mirt).

Iz navedenega sledi, da teh zapisov iz normativne baze podatkov nikakor ne smemo brisati. Seznama lahko služita predvsem kot seznam zapisov, ki jih je ob redakciji normativne baze podatkov CONOR.SI treba obravnavati prednostno.

SPOZNANJA OB ANALIZAH

Ko smo začeli delo pri projektu, smo probleme najprej popisali. Že na začetku se je nabralo okoli 30 različnih predlogov za izboljšave. Ob natančnejšem premisleku in po začetnih analizah so se nekateri izmed njih izkazali za neizvedljive, tako da smo jih kmalu opustili. Pojavljale so se nove ideje, ki so posamezne probleme poskušale reševati iz drugega zornega kota, pa tudi take, ki so problem samo nakazovale. Popisali smo 39 idej za izboljšave. Štiri smo opustili takoj po začetnem premisleku, tri smo zavrnilo po natančnejših analizah, eno napako pa smo odpravili takoj. 17 predlogov smo zajeli v konverzijo podatkov ali predlagali drugačno rešitev, 14 predlogov pa je ostalo še neuresničenih.

Pri pripravi predloga rešitve smo za vsak problem posebej najprej preverili, od kod izhaja: ali gre za neupoštevanje pravil pri vnosu ali je vzrok različna praksa v različnih obdobjih ali gre morebiti za programsko pomanjkljivost ali pa za pomanjkanje natančnih pravil in navodil. Nato smo definirali, kakšna je pravilna rešitev problema ter jo na koncu tudi predlagali.

Dopolnitve programske opreme COBISS2/Katalogizacija

Kot končna faza projekta je bila predvidena konverzija podatkov v bibliografskih zapisih v vseh lokalnih bazah podatkov knjižnic v sistemu COBISS.SI ter v vzajemni bazi podatkov COBIB.SI. Ob zavedanju, da gre za preko 300 lokalnih bibliografskih baz ter za vzajemno bazo podatkov z več kot tremi milijoni zapisov, smo vsi sodelujo-

či pri projektu čutili veliko odgovornost. Če pri kakšnem predlogu rešitve nismo bili prepričani, da ne bomo ustvarili še dodatnih napak v zapisih, smo predlog zavrnili. Ob tako strogem ocenjevanju predlogov se je zdelo, da bo mogoče konvertirati le malo podatkov, čeprav je napak in nedoslednosti v bazi podatkov veliko. Da bi preprečili vsaj tvorjenje novih napak in da bi katalogizatorje sproti opozarjali na nelogičnosti pri vnosu, smo predlagali nekaj dopolnitev programske opreme COBISS2/Katalogizacija.

Predlagali smo, da se tudi pri shranjevanju bibliografskega zapisa, ki je bil kreiran z ukazom NEW/INCLUDE izvede preverjanje pravilnosti vnosa kodiranih polj. S tem bi preprečili možnost kopiranja starih kod v nove zapise (npr. stare kode za državo izida v podpolju 102a).

Nadalje smo predlagali, da se spremeni kontrola, ki opozarja na nepravilen vnos ISSN 0000-0000 v podpolja 011a, s, e, f, c (kontrola 53). Ta kontrola je bila izvedena le kot opozorilo. Kontrolo smo razdelili na dva dela. Za podpolja 011e, f, c, torej za vnos v zapise za kontinuirane vire, je kontrola ostala enaka in katalogizatorja le opozori. Za podpolji 011a, s, torej za vnos v zapise za sestavne dele, se je kontrola poostrila in ne dovoli več vpisa 0000-0000. Na ta način članka ni več mogoče vezati na matično publikacijo, ki z vnosom 0000-0000 dejansko ni identificirana.

Dodali smo novo programsko kontrolo, ki bo katalogizatorje opozarjala na vnos kode za tip disertacije oz. raziskovalne naloge, če bodo v polje 702 vnesli kodo avtorstva za mentorje. Vnos kode za tip disertacije je koristen predvsem za tiste knjižnice, ki izpisujejo osebne bibliografije profesorjev, saj koda omogoča izločanje določenega tipa del (npr. diplomskih nalog).

Drugi predlogi

Nekaterih problemov nismo mogli rešiti niti s konverzijo podatkov, niti s spremembami programske opreme. Predlagali smo izdelavo kontrolnih seznamov, ki bodo katalogizatorjem v pomoč pri redakciji zapisov.

Za lažjo kontrolo neverificiranih ISSN smo predlagali uvedbo kontrolne datoteke, ki bi se programsko generirala ob vsakem posodabljanju (angl. *update*) baze podatkov ISSN. Program naj bi vsakič v bazi podatkov COBIB.SI poiskal zapise z vnesenim neverificiranim ISSN v polju 011f in preveril, ali se pojavijo v bazi podatkov ISSN. V kontrolno datoteko bi se zapisali zapisi, ki bi ustrezali pogoju. Te zapise bi bilo treba nato sproti ročno dopolnjevati.

Tudi za reševanje problema CIP-zapisov smo predlagali uvedbo novega mesečnega seznama neažuriranih zapisov.

Seznam bi sestavili tako, da bi najprej preverili, kateri zapisi so v vzajemni bazi podatkov COBIB.SI označeni kot CIP, nato pa na osnovi podatkov o zalogi in na osnovi podatkov o pooblastilih katalogizatorjev seznam prilagodili posameznim knjižnicam. Seznam zapisov bi vsak mesec pošiljali v knjižnice in jih prosili, naj zapise dopolnijo in shranijo v vzajemno bazo podatkov.

Normativne zapise, ki so označeni za brisanje in niso povezani z nobenim bibliografskim zapisom v nobeni lokalni bazi podatkov niti v vzajemni bazi podatkov COBIB.SI bomo iz baze podatkov CONOR.SI izbrisali. Predlagali smo, da take zapise najprej programsko identificiramo in jih zapišemo v ustrezno kontrolno datoteko. Na osnovi skupnega števila identificiranih zapisov se bomo odločili, ali jih bomo izbrisali ročno ali programsko.

Za zmanjšanje razlik med zapisi v lokalnih bazah podatkov in zapisi v vzajemni bazi podatkov smo predlagali, da čim prej uvedemo sinhronizacije normativnih zapisov tudi v vseh lokalnih bazah podatkov, razen v tistih, ki tega izrecno ne bi želele. Sinhronizacija normativnih zapisov pomeni, da se podatki o avtorjih v bibliografskih zapisih sprti usklajujejo s podatki o istih avtorjih v normativnih zapisih. Usklajevanje je pomembno, ker omogoča sprotno upoštevanje sprememb (npr. spremembe priimkov avtorjev, dodatna variantna imena avtorjev) in s tem lažje iskanje. Sinhronizacija normativnih zapisov v vzajemni bazi podatkov COBIB.SI poteka vsako noč, medtem ko redna sinhronizacija normativnih zapisov poteka le v nekaterih lokalnih bazah podatkov. Podatki v lokalnih bazah zato zastarevajo, knjižnice pa ne morejo občutiti učinkov normativne kontrole.

KONVERZIJA

Po izvedenih analizah smo zavrnili predloge za programsko izvajanje popravkov v zapisih za serijske publikacije in v zapisih za Slovensko nacionalno bibliografijo v bazi podatkov COBIB.SI. V konverzijo smo zajeli vse predloge sprememb v bibliografskih zapisih, ki smo jih analizirali, razen predloga konverzije podpolj 675bsu v polja za nacionalno rabo, za katero smo predlagali, da se izvede kasneje. V konverzijo smo zajeli tudi obe predlagani spremembi v podatkih o stanju zaloge.

Ko smo pripravili dokončni predlog, kaj bomo v konverzijo podatkov zajeli, smo začeli pripravljati dokumentacijo. Za vsak predlog oz. poseg v zapise smo pripravili ustrezno specifikacijo in kjer je bilo potrebno, tudi ustrezne konverzijske tabele. Preveriti je bilo treba še vrsto drugih programov, na katere bi morebiti tudi vplivale spremembe (npr. programi za konverzijo bibliografskih zapisov med različnimi formati, npr. za konverzijo med formatoma MARC21 in COMARC, ki se uporablja za

prevzemanje zapisov iz baze podatkov WorldCat). Prav tako je bilo treba dopolniti in popraviti uporabniško dokumentacijo, predvsem uporabniška priložnika COMARC/B Format in COBISS2/Katalogizacija. Nazadnje smo 24. 10. 2006 o nameri obvestili tudi vse polnopravne članice sistema COBISS.SI in jih prosili, naj podajo svoje pripombe in predloge ter nam sporočijo, ali želijo sinhronizacijo normativnih zapisov v svojih lokalnih bazah podatkov [1].

V odgovor na naše sporočilo smo le iz treh knjižnic prejeli vprašanja in komentarje, ki so se navezovali na konverzijo podatkov, iz 22 knjižnic pa smo prejeli prošnjo za sinhronizacijo normativnih zapisov v lokalnih bazah podatkov. Pripomb na predlagano konverzijo podatkov ni bilo, kar smo razumeli kot sprejetje predloga in potrditev konverzije.

Medtem ko smo pripravljali dokumentacijo, so programerke pripravile testno konverzijo nekaj testnih baz podatkov. Sledilo je obdobje intenzivnega testiranja in odpravljanja napak. Testiranje smo uspešno zaključili in za izvedbo konverzije izbrali 16. in 17. 12. 2006, konec tedna, ko večina knjižnic ne dela.

15. decembra smo vse polnopravne članice sistema COBISS.SI obvestili, da bomo konec tedna izvedli konverzijo. Obvestilu smo priložili dodatna pojasnila o bistvenih spremembah pri vnosu podatkov [2].

Konverzijo smo začeli izvajati v soboto, 16. decembra po 15. uri, ko večina knjižnic preneha delati. V času konverzije je bil namreč zaprt dostop do vseh lokalnih baz podatkov. Sama konverzija in nato še indeksiranje baz podatkov sta potekala v soboto popoldan in ponoči. V nedeljo zjutraj je bilo treba preveriti, ali so se vsi postopki pravilno izvršili in zaključili. Med konverzijo ni prišlo do napak, tako da smo že v nedeljo zjutraj ob 9. uri ponovno odprli dostop do vseh lokalnih baz podatkov in ponovno vzpostavili normalno delovanje sistema.

Konverzijo podatkov smo opravili v 311 lokalnih bazah podatkov in v vzajemni bazi podatkov COBIB.SI. V vzajemni bazi podatkov COBIB.SI je bilo spremenjenih 423.604 zapisov (približno 14 %) od skupno 3.009.847 zapisov v bazi podatkov. Nekaj števil:

- Koda "d" v podpolju 105a je bila s kodo "a" zamenjana 154.241-krat.
- Zamenjanih je bilo 56.913 starih kod za vrsto avtorstva.
- Vrstilec za iskanje "fik" v polju 675c je bil z ustreznim vrstilcem zamenjan 6.235-krat.
- Koda "y" je bila iz podpolja 105f izbrisana 49.465-krat.

- V podpolju 500m je bila koda jezika v svojo razrešitev spremenjena 34.755-krat.

Dobro smo "počistili" tudi podatke o zalogi na vzajemnem nivoju, saj smo izbrisali skupno 543.569 polj s podatki o zalogah v neslovenskih knjižnicah.

Prve dopolnitve programske opreme, ki smo jih predlagali, so bile zajete v programski opremi COBISS2/Katalogizacija že v verziji 8.4-01, vse druge pa bodo aktivne po instalaciji verzije 8.5. Prav tako smo že začeli sinhronizirati normativne zapise v lokalnih bazah podatkov za knjižnice, ki so to želele. V 14 knjižnicah sinhronizacije že redno potekajo. V eni knjižnici so se po pregledu predvidenih sprememb odločili, da sinhronizacije ne želijo, v 7 knjižnicah pa še pregledujejo predvidene spremembe značnic in njihove dokončne odločitve še nismo prejeli. Predlogi za dodatne kontrolne sezname oz. datoteke so bili sprejeti, a so trenutno še v fazi izdelave.

Projekt *Izboljšanje kakovosti zapisov v COBIB.SI* je potekal od februarja 2005 do decembra 2006. V tem času je pri izvajanju projekta sodelovalo 14 ljudi, ki so za delo na projektu porabili približno 860 ur (oz. 5,4 človek/mesecev). Pri tem izračunu nismo upoštevali kolegov, ki so samo prispevali svoje pripombe in predloge za izboljšave, ter časa, ki smo ga porabili za občasna posvetovanja z njimi.

ZA KONEC

Čeprav je projekt potekal skoraj dve leti, nas je vseskozi preganjal čas. To je pri tovrstnem početju menda kar običajno. Posamezne probleme in dileme smo želeli čim bolje preučiti iz različnih vidikov. Pri tem smo brskali po priročnikih in pravilnikih in tuhtali, kako naj zahtevke za posamezne analize programerkam definiramo tako, da nam bodo rezultati prinesli uporabno rešitev. Seveda pa vsega svojega časa nismo mogli posvetiti samo temu projektu, tako da so nam druge naloge včasih pretrgale "rdečo nit". Vse to se odraža v dejstvu, da je 14 evidentiranih problemov ostalo nerešenih.

Ali smo se med projektom kaj naučili? Še bolj kot na začetku se zavedamo, da kakovost celotne baze podatkov temelji na kakovosti vsakega posameznega zapisa. Večine napak ni mogoče odpraviti programsko in najboljše bi bilo, da do njih sploh ne bi prišlo. Zato je treba katalogizatorjem omogočiti neprestano izobraževanje, da svoje znanje osvežijo in poglobijo.

Naučili smo se tudi, da dogovarjanj, pojasnjevanj in usklajevanj ni nikoli preveč. Vsak posameznik s svojimi izkušnjami in pogledi vidi problem v drugačni luči. Prav tako vsak razume dogovore v luči svojega predznanja in

prav lahko se zgodi (in nam se je zgodilo), da je rezultat dogovora drugačen od tega, kar so si posamezniki predstavljali. Kljub vsemu pa je za uspešno delo na projektu ključno dobro sodelovanje. Za naš projekt lahko rečemo, da smo ga uspešno zaključili, za kar gre zahvala vsem sodelujočim.

Projekt se je zaključil, s tem pa niso izginili vzroki, ki so botrovali njegovemu začetku. Ni nam vseeno za kakovost podatkov v vzajemni bazi podatkov, zato se bomo z zagotavljanjem čim večje kakovosti na tak ali drugačen način ukvarjali tudi v prihodnje.

Referenci

- [1] Dostopno na spletnem naslovu: http://home.izum.si/cobiss/obvestila_novosti/dokumenti/Konverzija_bibliografskih-baz_%20podatkov_dopis_20061017.pdf.
- [2] Dostopno na spletnem naslovu: http://home.izum.si/cobiss/obvestila_novosti/dokumenti/Konverzije_baz_podatkov_pojasnila_2006_12_15_dopis_ar_dbf.pdf.