

Kvalitativna analiza semantičke interoperabilnosti zapisa o primjercima: uloga Nacionalne i sveučilišne knjižnice

Jakob Sobota, Nacionalna i sveučilišna knjižnica u Zagrebu

jsobota@nsk.hr

Mario Krajna, Nacionalna i sveučilišna knjižnica u Zagrebu

mkrajna@nsk.hr

Vesna Golubović, Nacionalna i sveučilišna knjižnica u Zagrebu

vgolubovic@nsk.hr

RAZLOZI STVARANJA SUSTAVA I ALATA

U Nacionalnoj i sveučilišnoj knjižnici se prilikom izrade programa za inventarnu knjigu, financijskih izvještaja, biltena prinova i sličnih proizvoda pokazalo da su podaci u bazi podataka primjeraka neprikladni jer su ili nepostojeći ili loše kvalitete.

Izvještaji napravljeni na osnovu takvih podataka su djelomice netočni, neistiniti ili u najboljem slučaju tek približni te onemogućuju kvalitetnu analizu i planiranje.

Pokretanje programa za pronalaženje grešaka jednom godišnje daje mnogo pogrešaka te proces ispravljanja nakon toga dugo traje.

Da bi se to riješilo napravljen je sustav za upravljanje podacima o primjercima. Osmišljen je i cijeli proces pronalaženja grešaka i nedosljednosti u unosu podataka o primjercima te ispravljanja podataka.

OPIS SUSTAVA

- Sustav se sastoji od nekoliko složenih sql-upita za traženje grešaka koji proizvode fajlove u .html obliku.
- Pokretanje programa svake noći u isto vrijeme osigurano je preko unixovog servisa 'cron'.
- Gotovi fajlovi se automatski stavljaju na stalnu intranetsku lokaciju zamjenjujući fajlove od prethodnog dana.
- Pristup intranetskoj lokaciji imaju svi djelatnici NSK i njihova bi obaveza trebala biti da na dnevnoj bazi pregledavaju te fajlove i isprave svoje greške.

Tablice

Zbog djelomice različitog sadržaja napravljeno je nekoliko tablica kojima je glavni ključ inventarni broj:

- 1. tablica **Pogreske_invbroj_2017.html** sadrži podatke o inventariziranom primjerku u 2017.
- 2. tablica **Pogreske_r_invbroj_2017.html** sadrži podatke o reinventariziranom primjerku (NSK je 2014. godine krenula u reinventarizacije sve građe, a struktura inv. broja je ista kao kod inventarizacije s tim da mu prethodi znak 'R'),
- 3. tablica **Pogreske_ne_invbroj_2017.html** sadrži sve ono što nema inventarni broj, a trebao bi imati.

Dodatne tablice

Da bi programski paket bio potpun dodani su upiti:

- upiti koji traže višestruke podatke među onima koji bi po svojoj naravi trebali biti jedinstveni (inventarne brojevi i barcode)
- upiti koji traže nepotpune nizove brojeva (rupe) tamo gdje nizovi moraju biti potpuni (inv. broj, reinv. broj).
- Za potrebe administriranja dodana je i tablica s najvećim iskorištenim inv. brojem prema godinama i vrsti inventarizacije.

Podaci koji su predmet analize

Elementi inventarne knjige

- inventarni broj
- datum inventarizacije
- bibliografski opis
- ISBN/ISSN
- vrsta građe
- vrsta nabave
- dobavljač
- cijena
- status primjerka
- napomena
- lokacija
- signatura

Podaci važni za druge module

- broj narudžbe (order_id)
- barcode
- godina izdanja

Podaci potrebni za administriranje

- id bib. sloga
- id primjerka
- id holdinga
- prvi katalogizator
- zadnji katalogizator

Analiza grešaka

1. Inventarni broj

- Inventarni broj se generira od strane sustava Aleph i dozvoljeni format je 'YYYY/0000000000' pa tu nije moguće pogriješiti. Ipak, postoje situacije kad je potrebno ručno upisati inventarni broj zato što broj nije dodijeljen ili je dodijeljen broj iz krivog niza. Također je nužno upotrijebiti i brojeve koji su izbrisani zajedno s primjercima koji su ih imali (na pr. pri uvezivanju) da bi niz bio potpun.

2. Datum inventarizacije

- Prije svega to je obavezni element u inv. knjizi, a osim toga godina inventarizacije treba biti ista kao i godina koja je dio inv. broja.

3. Bibliografski opis

- Bibliografski opis se slaže iz višerazinskog zapisa do tri razine povezane preko LKR\$b gdje postoji \$r774.
Bibliografski opis se sastoji od tagova 1xx, 245,250,260 i 300.

4. Materijal (vrsta građe)

- Dekodira se i uspoređuje sa stupcima signature, lokacije i sl.

5. Vrsta nabave

- Dekodira se i testira na postojanje podatka te na dozvoljene kodove. Ukoliko je za knjigu napravljena narudžba tada se provjeravaju kodovi za vrstu nabave u primjerku i narudžbi.

6. Dobavljač

- Ukoliko postoji narudžba tada se dobavljač dekodira preko z68 (orders) iz tablice z70 (vendors). Ukoliko ne postoji (obavezni primjerak) tada se kao dobavljač pojavljuje izdavač ili ukoliko ga nema (na pr. vlast. naklada) autor postaje dobavljač.

7. Cijena

- Podatak o cijeni se preuzima sa knjige ukoliko se radi o obaveznom primjerku ili se preuzima iz narudžbe ukoliko se radi o kupnji. U svim ostalim slučajevima u taj stupac se upisuje vrijednost. Cijena mora biti u formatu ('999990.00') te se testira na znakove koji smiju biti samo brojevi i decimalna točka. Cijena ne smije biti manja od 4.00 kn niti veća od 10,000.00 kn.

8. Status primjerka

- Dekodirano i testirano na dozvoljene kodove. Ako je knjiga na signiranju ili izlučena iz fonda tada podatak o signaturi nije obavezan.

9. Status primjerka

- Dekodirano i testirano na dozvoljene kodove. Ako je knjiga na signiranju ili izlučena iz fonda tada podatak o signaturi nije obavezan.

10. Opis primjerka

- Stupac opis primjerka (item_description) je obavezan za serijske publikacije (vrsta građe: 'ISSUE', 'ISSBD') te se testira postojanje podatka.

11. Napomena

- Napomena se ne testira pri inventarizaciji, ali pri reinventarizaciji u nju se upisuje stari inventarni broj. Testira se na format ('X99990/YYYY') gdje je 'X' prefiks koji označava vrstu nabave i vrstu građe.

12. Lokacija

- Lokacija se dekodira, a testira se usporedbom sa signaturom i materijalom

13. Signatura

- Signatura se testira na strukturu, a potom na međuovisnost signature i lokacije ili materijala. Prefiksi signatura kao i dozvoljeni znakovi su standardni za pojedine formate i lokacije te se kao greška prikazuju svi koji odstupaju od sustava. U slučaju potrebe postojanja dodatnih elemenata (\$2udc) (otvoreni pristup poredan po udc) testira se njihovo postojanje i sadržaj.

14. Holding (posjedovanje)

- U NSK je holding obavezan podatak te se testira njegovo postojanje i sadržaj.

15. Barcode (štapasti kod)

- Barcode se testira na strukturu, dužinu i vrstu znakova te početne znakove (dužina 9 znakova, samo brojevi, počinje s '99'). Testira se i nepostojanje barcode-a za određenu vrstu građe i za određene lokacije.

16. Status bibliografske obrade

- Dekodirano i testirano na dozvoljene kodove. Kao greška se javlja i ukoliko sa bibliografskog zapisa nije skinut status 'ACQ_CREATED' više od tri mjeseca od nastanka primjerka.

17. ISBN/ISSN

- Napravljeno je nekoliko testova na format ISBN-a ili EAN-a. Za ISBN je testirana dužina podatka (10 znakova) i znakovi (prisutnost samo brojeva i 'X' na zadnjem mjestu). EAN je također testiran na dužinu (13 znakova) i sadrže li '978' kao prva tri znaka. Treba spomenuti da je napravljen i program za testiranje kontrolnih znakova, ali nije ovdje implementiran zato što pronalazi previše grešaka! Osim toga ISBN se primjenjuje od 1972. godine te su sve knjige koje ga posjeduju, a izdane su prije te godine, sumnjive.

Nakon sređivanja podataka mogući su detaljni i potpuno točni izvještaji poput: NSK – dio financijskog izvještaja nabave za 2016. godinu

materijal	Hst	broj	prosjeak	vrijednost
sve		24513	136,77	3352712,04
ATLAS		19	315,53	5995,00
BOOK		20154	143,97	2901631,10
BOOK	HST	1944	34,88	67802,68
CASST		2	30,00	60,00
CDISC		1124	51,14	57480,86
CDROM		519	80,49	41773,40
DRAW		3	2100,00	6300,00
DVD		3	58,00	174,00
DVDRM		2	75,00	150,00
KIT	HST	9	53,33	480,00
MAP		332	394,92	131115,00
MNOTE		278	86,61	24077,00
OTHER	HST	26	10,58	275,00
PCARD		4	3550,00	14200,00
PORTF		5	3680,00	18400,00
POST		79	635,44	50200,00
PRINT		8	4050,00	32400,00
SDISC		2	99,00	198,00

Ograničenja koja utječu na funkcionalnost:

- Pošto ne postoji standard za zapise o primjercima, količina elemenata i kvaliteta podataka bitno su određeni izvještajima koji na osnovu njih nastaju (na pr. inv. knjiga). No knjižnica je živ organizam koji se mijenja i nadograđuje te je izuzetno bitno da svaka promjena u knjižničkoj praksi bude praćena promjenom u programu jer će inače javljati greške tamo gdje realno ne postoje.

- Organizacija ispravljanja podataka zamišljena je tako da svaki djelatnik sam pristupa tablicama i ispravlja ono što je program pronašao kao njegovu grešku. Ukoliko to nije postavljeno kao obaveza tada je funkcionalnost smanjena i taj dio je prebačen na ključnog knjižničara koji šalje podatke na ispravak ili ih ispravlja sam.

- Program pokrenut na cijeloj bazi može pronaći previše grešaka što će nas obeshrabriti u ispravljanju. U takvom slučaju je nužno odrediti koji su nam podaci važni i koliko je toga moguće ispraviti uz redovno poslovanje. Zbog velike količine podataka NSK je odlučila urediti bazu na nivou dovoljnom za ispis trenutne inventarne knjige, a ispravljanje ostaloga ostaviti za kasnije. Ipak, dio toga se odrađuje preko reinventarizacije za koju je program za otkrivanje grešaka također napravljen i dio je ovog paketa.

Proces izvještavanja i knjižnice sustava IKS

- Testiranje mogućnosti upotrebe takvog procesa kod drugih knjižnica napravljeno je na podacima Hrvatskog katoličkog sveučilišta (HKS). Proces okuplja programe za traženje grešaka za primjerke te ispisa inventarne knjige i financijskog izvještaja. Proces se sastoji od nekoliko stupnjeva: traženje grešaka, slanje na ispravak, povratna informacija o ispravljenom te, na kraju, izrada inv. knjige i financijskog izvještaja. Procesom je poboljšana kvaliteta podataka i izvještaji su gotovi za daljnju upotrebu.

Izjava o autorstvu

- Treba svakako spomenuti da je alat za pronalaženje grešaka proizveden potpuno u NSK te da je proces izvještavanja osmišljen i proveden bez uzora u alatima programskog paketa Aleph. Isto tako, alat je osmišljen za potrebe NSK i sadrži testiranja koja u tom opsegu vjerojatno neće trebati niti jednoj drugoj knjižnici.

Zaključak:

- Proces otkrivanja grešaka u bazi podataka primjeraka u NSK sveobuhvatan je i izrazito detaljan. Unaprijedio je kvalitetu podataka u bazi primjeraka i drastično skratio vrijeme ispravljanja grešaka. Izvještaji na osnovu tako sređenih podataka su točni, detaljni i pouzdani. Postoje mogućnosti upotrebe takvog ili sličnog procesa u drugim knjižnicama sustava IKS, ali to zahtijeva mnogo prilagodbi zbog specifičnosti podataka i oblika izvještaja.

- Izvještaji za administraciju baze primjeraka NSK na adresi:

- <http://iks.nsk.hr/izvjestaji/>