

NSK je objavila dokument o poteškoćama pri konverziji zapisa iz CDS/ISIS-a:
Očekivane teškoće kod migracije bibliografskih zapisa iz ISISa u Voyager / Hrvoje Brozovič

<http://www.nsk.hr/Info.aspx?id=488>

U dokumentu je dat pregled mogućih teškoća a te su predložena neka rješenja.

Na ovim stranicama opisani su pokušaji rješavanja prepoznatih teškoća a na podacima FF-a.

Teškoća a:

- ISISov format eksporta u ISO 2709 datoteku sadrži prekide redova (CR+LF-dva bajta) nakon svakih 80 znakova, te naposljetku na kraju svakog sloga. Stoga je podatak o duljini ISO zapisa sadržan u laebli sloga strogo gledano netočan, jer je broj bajtova u datoteci veći za $2 * (\text{duljina zapisa} \div 80) + \text{još } 2$ ako postoji ostatak pri dijeljenju duljine zapisa sa 80.

Rješenje:

Pri konverziji se mogu koristiti originalni ISIS podaci (MST). Konvertirana datoteka je ispravni ISO 2709.

Teškoća a:

- ISISove oznake za početak polja, kraj polja i kraj sloga nisu standardne vrijednosti iz ISO 2709 standarda, već obični znakovi koji se mogu pojaviti u samom tekstu, što upravo oni koji se u DOS praksi koriste za načina slova, što pak dovodi do raznorodnih pristupa kodiranju dijakritika.

Nismo imali problema jer su svi podaci u windows-1250 enkodingu.

Teškoća a:

- indikatori koji se mogu pretpostaviti

Rješenje:

Kad nemaju značaja, postavljaju se na praznine.

70x - 2. ind. postavlja se na 1

(^a je prezime, ^b je ime)

Teškoća a:

- u polju 71x, gdje bi indikatori trebali objasniti da li se radi o korporativnom tijelu ili sastanku, a o čemu pak ovisi u koje se polje MARC21 uopće podatak iz UNIMARC 71x prenosi, indikatori su od odlučujuće važnosti

Rješenje:

Radi li se o korporativnom tijelu ili sastanku pretpostavlja se na temelju postojećih zapisa. Ako zapis ima sjedište i godinu održavanja sastanka, onda se radi o sastanku. Zapis ima relativno malo, pa je moguće podatke jednostavno izlistati i provjeriti da li su se dobro konvertirali.

Teškoća a:

- ISIS ne sadrži praktično nikakve podatke u labeli zapisa, niti one koje propisuje ISO 2709, niti one koje propisuje UNIMARC. Dakle, podaci o tipu zapisa, bibliografskoj i hijerarhijskoj razini, te stanju zapisa naprosto nisu kodirani u labeli

Rješenje:

Podatak o stanju zapisa dobiva se iz polja 9xx.

Vrsta zapisa je za sve zapise a (tekstualna građa)

Bibliografska razina dobiva se prema bazama - monografije, periodila i analika su u posebnim bazama.

Podataka o hijerarhijskoj razini nema. Svi zapisi u originalnoj bazi su iste razine.

Teškoća:

- Mnoga polja unutar ISISa nisu u skladu s UNIMARC i MARC21 implementacijom, da polja 00x nemaju potpolja, već samo vrijednost, a sva ostala polja da imaju potpolja. Tako u ISIS aplikacijama često nailazimo na polje 100 u koje su kodirani podaci (zapravo najčešće samo godina) upisani bez potpolja \$a. Također, među lokalnim poljima, ovo je redovna praksa.

Rješenje:

U postupku konverzije nadomještaju se nedostajuća potpolja. Moguće je to napraviti i za slučaj kad neka polja imaju a neka nemaju potpolje.

Teškoća:

- *ISIS upis oznaka za potpolje prepuzta samom katalogizatoru, tj osim samog sadržaja pojedinog potpolja, potrebno je unijeti i oznaku za potpolje (najčešće e ^) te samu oznaku potpolja, te tek kao treći znak unosa sam tekst potpolja. Uz svu moguću pažnju katalogizatora, ovo naravno proizvodi određeni postotak zapisa koji nemaju ispravno označena potpolja. Neka se dolje navedene greške potkradu u 1% slučajeva, i nek prosječan slog ima 5 polja, to dovodi do toga da svaki 20-ti zapis ima grešku o dolje nabrojanih greškama.*

- ◆ Nekada fali sama oznaka za potpolje (\$, ^) pa tada imamo aHR umjesto \$aHR, što se tumači polje uopće potpolja, već da je njegova jedinstvena vrijednost aHR
- ◆ Često est je i slučaj da se upiše oznaka za potpolje, ali ne i sam identifikator potpolja, već odmah ide tekst, pa imamo \$HR, što pak znači da potpolje \$H ima vrijednost R. Ovdje je problem što identifikatori potpolja mogu po UNIMARCU biti samo mala slova ili zamnke, pa je potpolje \$H nedopušteno.
- ◆ Neke ISIS aplikacije omogućuju izbor potpolja iz padajućeg menija, ali i tada se samo oznaka i identifikator potpolja jednostavno umeću u tekst, pa se događa i situacija HR\$a, što bi značilo prvi indikator H, drugi R (iako indikatora nema u ISISu, a i nisu po UNIMARCU dopušteni ovakvi, već samo praznina i znamenke), te potpuno potpolje \$a bez ikakvog teksta, što je opet nedopušteno.

Rješenje:

Većina greškama proizila iz CDS/ISIS-ovog nedostaka vezanih uz unos podataka, riješena je kako je opisano u poglavlju [Greške u CDS/ISIS zapisima](#).

Postotak ovih greškama nije još rađunat. Moguće je prebrojati zapise s greškama (višestruke greške okupljaju se oko jednog zapisa).

Teškoća:

- Polje kodiranih podataka 100, najčešće e u ISISu sadrži samo godinu unosa u datoteku, dakle 4 znaka, i eventualno datum, što daje 8 znakova, dok UNIMARC traži u

ovom polju 36 kodiranih znakova. Nisu svi ovi kodovi obavezni, već samo datum unosa, koji je najčešće i jedini naveden, te jezik katalogiziranja i skup znakova, koji se pak mogu pretpostaviti za cijelu bazu.

Rješenje:

Datum unosa u datoteku - postoji u polju ??

Jezik katalogiziranja - hrvatski

Pismo stvarnog naslova - polje ??