

Vloga kontroliranih slovarjev v podatkovnih zbirkah

ALENKA ŠAUPERL, DR. IZR. PROF.

Oddelek za bibliotekarstvo, informacijsko znanost in knjigarstvo, Filozofska fakulteta, Univerza v Ljubljani,
Aškerčeva 2, SI - 1000 Ljubljana, Slovenia
e-mail: alenka.sauperl@ff.uni-lj.si

The Role of Controlled Vocabularies in Information Databases

ABSTRACT

Libraries follow international standards and guidelines for subject description of publications and documents, i.e., for subject heading lists and classification systems. IFLA guidelines for subject heading languages are introduced. They guide the establishment and development of subject heading lists, selection of subject headings and building of relationships among them, syntax, and give other recommendations. Functional requirements for subject authority data establish the structure of authority files. These files contain subject heading lists and merge them with user interfaces of library catalogs. Authority control makes it possible for the user to search a certain topic by one term even when this term may be described by different terms (synonyms).

Il ruolo dei vocabolari controllati nei database di informazioni

SINTESI

Le biblioteche seguono standard e linee guida internazionali per la descrizione dei soggetti delle pubblicazioni e dei documenti, ad esempio liste di soggetti e sistemi di classificazione. Vengono presentate le linee guida dell'IFLA le lingue delle liste di soggetti. Esse guidano l'instaurazione e lo sviluppo delle liste di soggetti, la selezione di soggetti e la costruzione di un sistema di interrelazione fra di loro, la sintassi, e danno altre raccomandazioni. I requisiti funzionali per i dati d'autorità stabiliscono la struttura dei file d'autorità. Questi file contengono liste di soggetti e le incorporano con i cataloghi librari degli interfaccia per gli utenti. Il controllo di autorità rende possibile all'utente la ricerca di un determinato argomento attraverso un termine anche quando tale termine può venir descritto da termini differenti (sinonimi).

Vloga kontroliranih slovarjev v podatkovnih zbirkah

IZVLEČEK

Knjižnice sledijo enotnim mednarodnim standardom in smernicam tudi za opis vsebine publikacij in dokumentov, to je za gesla in klasifikacijske sisteme. Predstavljene so IFLA smernice za predmetno označevanje, ki usmerjajo zasnovo in razvoj geslovnikov, izbor besedišča, vzpostavljanje medsebojnih odnosov med pojmi, sintakso in dajejo druga priporočila. Funkcionalne zahteve za predmetne normativne podatke pa podajajo osnovno strukturo za normativne datoteke. Te vsebujejo geslovnike in jih vgrajujejo v uporabniške vmesnike knjižničnih katalogov tako, da lahko uporabniki z enim samim izrazom pridejo do vsega gradiva na neko temo, četudi je tema poimenovana z več sinonimi.

1. UVOD

Kdorkoli se ukvarja z zbiranjem, urejanjem, shranjevanjem in posredovanjem informacij, se srečuje tudi s postopki razvrščanja in grupiranja, npr. klasificiranja in gesljenja publikacij, dokumentov, predmetov, podatkov, skratka informacijskih objektov. Večinoma te postopke laično izvajamo pri vnosu z obrazci in formami in vsak programer, ki snuje podatkovno bazo, se jih loteva na svoj način. Dejansko pa so ti postopki v bibliotekarstvu že stoletja znani in preverjeni, zato imajo tudi jasne teoretične osnove. Kljub temu, da je ta teorija uporabna tudi v drugem okolju, se je niso posluževali niti računalničarji, niti naročniki njihovih storitev. Tako so se metodoloških težav pri gradnji klasifikacijskih sistemov in tezavrov tudi v slovenskih arhivih lotevali posamično in na načine, ki niso bili vedno optimalni za sodobne informacijske sisteme. V zadnjih letih pa skupni nacionalni in mednarod-

ni projekti združujejo različne ustanove, ki se ukvarjajo s kulturno dediščino. Na ta način prihaja do sodelovanja tako na praktični kot teoretični ravni in tudi do prenosa teoretičnih in praktičnih spoznanj med sorodnimi vedami in strokami.

Sisteme za vsebinsko označevanje v bibliotekarstvu ločujemo glede na namen, cilje in izvedbo na geslovnike, tezavre in klasifikacije (Urbanija, 1996, Šauperl 2000, 2007). Vsak izmed njih ima svojo notranjo logiko vzpostavljanja in uporabe. V knjižnicah jih rabimo za opisovanje vsebine publikacij zaradi iskanja v knjižničnih katalogih (geslovniki, tezavri) ter razvrščanje gradiva na policah, do katerih pristopajo bralci sami (klasifikacijski sistemi). Pri vsebinah arhivske vrednosti so pomembni predvsem pri selekciji dokumentarnega in arhivskega gradiva (klasifikacije) ter urejanju evidenc zaradi posredovanja v uporabo (geslovniki) (Novak, 2007; Semlič Rajh in Šauperl, 2011). V računalniških sistemih tako klasifikacije kot geslovnike in tezavre pogosto poimenujemo kar »šifranti«. Ko se ukvarjamo z njimi pa redko pomislimo, da pravzaprav tudi zanje veljajo določene zakonitosti.

Procese standardizacije na področju bibliotekarstva vodi IFLA, Mednarodno združenje bibliotekarskih društev in ustanov. Čeprav je standardizacija na področju bibliografskega opisa stara že 40 let, se standardizacije vsebinskega opisa dolgo nismo lotevali drugače kot z uporabo tradicionalnih bibliografskih klasifikacij ki so prevedene v mnoge jezike. Tako so npr. mednarodno najbolj razširjeni trije klasifikacijski sistemi: Deweyeva decimalna klasifikacija, Univerzalna decimalna klasifikacija in Klasifikacija Kongresne knjižnice. Klasifikacijski sistemi niso tako močno odvisni od jezika kot geslovniki. Zato na področju geslovnikov in tezavrov ni podobno prevladujočega. V 1990-ih letih je prišlo do prvega mednarodnega pregleda geslovnikov in skupina, ki se je pri IFLA s tem ukvarjala, je prišla do ugotovitev o skupnih teoretičnih osnovah. Tako so izšle *Ifline smernice za predmetno označevanje* (2002). Dolgoletna prizadevanja so dosegla vrh na svetovni konferenci avgusta 2009 v Milanu, kjer so predstavili *Funkcionalne zahteve za predmetne normativne in napotilne podatke FRASAD* (Zeng in Žumer, 2009) ter nove smernice za večjezične tezavre (*Guidelines*, 2009). Oba dokumenta bomo predstavili v nadaljevanju. Prikazali pa bomo tudi pomen normativne nad predmetnimi oznakami.

2. UPORABLJENA TERMINOLOGIJA

Ker se arhivska stroka verjetno še ni srečevala z običajno bibliotekarsko terminologijo s področja vsebinskega označevanja, so v nadaljevanju pojasnjeni nekateri osnovni pojmi. Definicije so večinoma iz *Bibliotekarskega terminološkega slovarja* (Kanič, 2009) in *Angleško-slovenskega slovarja bibliotekarske terminologije* (2002). Za italijanske izraze se zahvaljujem kolegici Amalii Petronio iz Osrednje knjižnice Srečka Vilharja v Kopru. Omenjene strokovne pojme je za arhivsko stroko v svojem delu *Preslikave vsebin v arhivskih strokovnih postopkih* opredelil tudi M. Novak (2007).

Deskriptor (*ang. descriptor, ita. descrittore*) je standardizirana predmetna oznaka za določen pojem v tezavru. To je značilno in standardizirano geslo z definirano vsebino, ki označuje njeno enopomensko predstavitev vsebine enega ali več dokumentov. Uporabljamo ga za urejanje in shranjevanje informacij ter za poizvedovanje po njih tako v računalniško podprtih kakor tudi v klasičnih podatkovnih zbirkah. Primer deskriptorja v evropskem tezavru Eurovoc je *kulturna dediščina* (<http://eurovoc.europa.eu/drupal/?q=sl/request&view=pt&termuri=http://eurovoc.europa.eu/379096&language=sl>).

Geselnik ali **geslovník** (*ang. subject heading list, ita. il soggettario*) je seznam besednih enot, določenih za gesla, ki so navadno urejena abecedno. Gesla v geslovníku so med seboj pomensko povezana z napotili tako, da se prikažejo izključene besede, nadpomenke ali podpomenke. Primer novega geslovníka je italijanski *Il soggettario*.

Gesljenje, tudi **predmetno označevanje** (*ang. subject description, ita. indicizzazione per soggetto*), je postopek oblikovanja in dodeljevanja gesel posameznim vsebinam, ki jih obravnavamo.

Geslo, tudi **predmetna oznaka** (*ang. subject heading, ita. la voce*), je beseda ali besedna zveza, ki opredeljuje vsebino dokumenta ali publikacije. Uporablja se kot dogovorjena razpoznavna beseda ali besedna zveza. To pomeni, da se nadzoruje in izključuje rabo sinonimov in homonimov ter podobnih jezikovnih značilnosti. Gesla predstavljajo vsebino dokumenta in jih uporabljamo za iskanje v informacijskih sistemih, katalogih in v informativnih pomagalih.

Primer gesla je »književnost« na sliki 1.

Kazalka (*ang. see reference, ita. richiamo*) je tisto napotilo, ki kaže povezavo s sprejetih na nesprejete pojme. Na ta način povezujejo identične pojme in izraža ekvivalenčne odnose. V geslovnikih in tezavrih se namreč ne dovoljuje uporabe sinonimov. Samo eden od sinonimov je lahko geslo, drugi so kazalke. Uvaja jih beseda »glej«. Njen dopolnilo v paru je »rabi za«. Če smo za geslo izbrali npr. izraz književnost, sta leposlovje in beletristika lahko kazalki. Slika 1 prikazuje zapis tega odnosa v geslovniku.

Beletristika <i>rabi</i> Književnost ... Književnost <i>rabi za</i> beletristika, leposlovje <i>glej tudi</i> Slovenska književnost ... Leposlovje <i>rabi</i> Književnost	Beletristika USE Književnost ... Književnost USE FOR beletristika, leposlovje NT Slovenska književnost ... Leposlovje USE Književnost
---	--

Slika 1: Prikaz ekvivalenčnega odnosa v geslovniku (*književnost* je izbrani veljavni izraz, leposlovje in beletristika sta kazalki, Slovenska književnost je vodilka na podrejeni pojem, levo slovenske oznake odnosov, desno angleške standardizirane oznake odnosov med pojmi)

Klasifikacija (*ang. classification, ita. classificazione*) je razvrstitev, razporeditev česa glede na enake ali podobne lastnosti ali tudi ocenitev, ovrednotenje. Klasifikacije danes v najširšem pomenu besede pomenijo načine razvrščanja predmetov ali njihovih zvez v razrede glede na skupno značilnost, ki je neločljivo združena z vsemi predmeti neke vrste in jih razlikujemo od predmetov druge vrste. Razvrščamo jih tako, da ima vsak razred v sistemu nasproti drugemu razredu točno določen položaj. Značilnost, po kateri klasificiramo, imenujemo načelo delitve.

Klasifikacijska oznaka (*ang. class number, ita. numero di classificazione*) je črka, številka, simbol ali njihova kombinacija, ki izraža razvrstitev v klasifikacijsko skupino. V slovenskih knjižnicah je to navadno oznaka Univerzalne decimalne klasifikacije, ki knjigo razvrsti na ustrezno tematsko polico. V arhivih gre za sistem, po katerem razvrščajo in odlagajo dokumente pri ustvarjalcih. V muzejih je to lahko razvrstitev po zvrsteh kulturne dediščine.

Kvalifikator (*ang. qualifier, ita. qualificazione*) je beseda ali besedna zveza, ki v predmetni oznaki pojasnjuje izbrani izraz z več pomeni za enolično razlikovanje ene entitete od druge.

Napotilo (*ang. reference, ita. rinviamiento*) pojasnjuje, na kateri strani ali na katerem mestu je potrebno iskati vsebino, ki jo označuje deskriptor ali geslo. Napotila so kazalke in vodilke.

Normativna kontrola (*ang. authority control, ita. controllo normativo*) se nanaša na normo, pravilo ali zakon in predstavlja konsistentno uporabo unikatnih oblik pojmov, kot so imena in priimki, predmetne oznake, geografski pojmi, časovne in druge oznake za označevanje in razvrščanje. Z njimi vzpostavljamo vsebinska kontekstualna razmerja in omogočimo najdenje informacije ali dokumenta ne glede na izraz, ki ga uporabimo pri iskanju.

Normativna zbirka (*ang. authority file, ita. raccolta di codici normativi*) predstavlja celoto podatkov, ki so organizirani v normativnih zapisih, ki služijo kot reference tako pri zajemanju kakor tudi pri proizvodovanju v informacijskih sistemih.

Normativni zapis (*ang. authority record, ita. codice normativo*) predstavlja zaokroženo podatkovno celoto, ki jo sestavlja nosilni pojem in temu dodeljene vsebine, ki ga opredeljujejo. Vsak nosilni pojem mora biti v sistemu unikatni. Če se pojavijo enaki nosilni pojmi in

označujejo različne vsebine, je potrebno nosilne pojme opremiti z dodatnimi identifikatorji (kvalifikatorji), tako da zagotovimo unikatnost normativnega zapisa.

Tezaver (*ang. thesaurus, ita. thesauro*) predstavlja referenčni slovar, ki uporabnikom pomaga izraziti zahtevo za informacijo s termini deskriptorskega jezika. V njem so zbrani deskriptorji. Odnosi med pojmi oz. deskriptorji in nedeskriptorji so dosledno in natančno označeni. Od tezavra v jezikoslovju se tezaver v informacijski znanosti loči po tem, da sinonime sicer zbere, a za uporabo dovoli le enega.

Vodilka (*ang. see-also reference, ita. rimando*) je tisto napotilo, ki kaže na sorodnost med sprejetimi pojmi. Tako povezuje pojme, ki so med seboj tematsko tesno povezani, vendar niso identični in se med seboj tudi ne izključujejo. Uvaja jih »glej tudi«. Na sliki 1 sta tako povezana izraza književnost in slovenska književnost.

3. SMERNICE ZA IZDELAVO GESLOVNIKOV IN TEZAVROV

Razlog za smernice leži v funkcijah knjižničnega kataloga (Izjava, 2009). Te pravijo, da mora biti knjižnični katalog učinkovito orodje, ki uporabniku omogoča, da vire najde, jih identificira, izbere in pridobi ter da se giblje znotraj kataloga in izven njega. Te funkcije lahko uporabimo za informacijske sisteme nasploh. V vseh zbirkah informacij želimo doseči isto: informacije najti (osnovni razlog za shranjevanje), jih identificirati (določiti, da so prave), jih izbrati (opraviti izbiro med več različnimi), pridobiti (imeti možnost pregleda, uporabe) ter se gibati znotraj in zunaj kataloga (kar pomeni, da moramo imeti vzpostavljene povezave). Da lahko te funkcije izpolnimo, moramo o shranjenih publikacijah voditi primerno evidenco, dokumentacijo, metapodatke. Če za knjigo nimamo avtorja in naslova, je po teh podatkih ne bomo mogli najti, identificirati ali izbrati. Prav tako se ne bomo mogli na podlagi avtorjevega imena gibati znotraj kataloga (med njegovimi različnimi deli), niti zunaj njega (med katalogom in spletnimi knjigarnami ali stranmi z drugimi informacijami o avtorju). Vstopne točke so tiste, ki določajo, kako bomo lahko publikacijo iskali. Medtem ko je avtorjevo ime v vseh katalogih bistvena vstopna točka, barva platnic to ni in zato tega podatka ne moremo uporabiti za iskanje, identifikacijo ali izbiro, pa tudi ne za pregledovanje kataloga.

Med bistvene vstopne točke - to so tiste, ki so za knjižnične kataloge nujne - sodijo imena ustvarjalcev, naslovi del, leto izida, vsebinske oznake ter identifikacijske številke in podobni označevalci (Izjava, 2009). To so bibliografski podatki, za katere je standardizacija nujna, zato podležejo normativni kontroli. Smisel standardizacije pa je v tem, da je mogoče najti vse bibliografske vire, ki podležejo določenemu iskalnemu kriteriju. To pomeni, da lahko najdemo vsa dela istega avtorja, ne glede na to, ali se je podpisal s pravim imenom ali enim od psevdonomov (npr. France Bevk in Pavle Sedmak), ali da lahko najdemo vsa dela na neko temo, četudi smo uporabili le enega od številnih sinonimov za njeno poimenovanje (npr. leposlovje, beletristika in književnost).

Funkcionalne zahteve za predmetne normativne podatke (Functional requirements for subject authority data, FRSAD) predstavlja abstraktni temelj vsakega sistema za predmetno označevanje in razlog za standardizacijo besedišča jasno prikaže: Neko delo lahko obravnava eno ali več tem (*thema* po Zeng, Žumer, Salaba 2010) (glej sliko 1). Vsako od njih pa poimenujemo z eno ali več besedami - poimenovanji (*nomen* po Zeng, Žumer, Salaba 2010) (glej sliko 1). Tako je v naravnem jeziku. Toda v sistemih za predmetno označevanje, se poimenovanje normira tako, da je enoznačno (uporabimo npr. samo poimenovanje „književnost“ ne pa tudi „leposlovje“ in „beletristika“). Predpostavimo lahko, da bo gradivo lažje obvladljivo in najdljivo pod enim poimenovanjem, kot bi bilo, če bi bilo označeno z različnimi poimenovanji in posledično raztreseno na različnih koncih evidence, dokumentacije in/ali zbirke. Tudi pri iskanju bi namesto večkratnih iskanj pod različnimi poimenovanji iskali le enkrat, in sicer s tistim poimenovanjem, ki nam pride na misel. Toda to združevanje je možno le, če je v sistem za iskanje informacij vključena normativna datoteka, ki vsebuje podatke o vseh sinonimih. Ti sinonimi se namreč lahko pri iskanju vključijo v delo, ne da bi se uporabnik za to posebej trudil.



Slika 1: Tri osnovne entitete modela FRASAD, *work* (delo), *thema* (tema) in *nomen* (poimenovanje) z relacijami (Zeng, Žumer in Salaba, 2010, p. 15)

Eden od začetnih dokumentov prizadevanja za ureditev vsebinskega označevanja so Iffine smernice za predmetno označevanje (Iffine, 2002), kjer so v enajstih načelih opredeljene temeljne zakonitosti za geslovnike. Čeprav niso popolna, jih je vredno upoštevati (Šauperl, 2005). Predstavili jih bomo v prenovljeni terminologiji FRASAD. Načelo o enoličnosti pravi, da vsak pojem za predmetno označevanje v sistemu zastopa le ena predmetna oznaka, torej eno poimenovanje za eno temo. To tudi pomeni, da moramo izmed sinonimov za normativnega izbrati le enega. Pri homonimih gre za prav nasprotno situacijo. Ker je isto poimenovanje uporabljeno za različne teme, jim moramo ustvariti razlikovanje. To se običajno naredi s kvalifikatorjem, tj. pojasnilom v oklepaju, npr. vile (pravljica bitja), vile (vrtnarsko orodje). Pri imenih imamo kot razlikovanje navadno leto rojstva in smrti ali poklic (npr. Novak, Janez, 1970- in Novak, Janez, duhovnik) (glej sl. 2)¹.

Št.	Značnica
1.	Novak, Janez, 1941- (02882)
2.	Novak, Janez, 1955- (03213)
3.	Lavrač, Vladimir (03571)
4.	Novak, Janez, 1939- (08986)
5.	Novak, Janez, 1970- (15884)
6.	Novak, Janez, dipl. ekonomist, 1952-
7.	Novak, Janez Krstnik
8.	Novak, Janez, duhovnik
9.	Novak, Janez, pravnik, 1970-
10.	Novak, Janez, geodezija, 1983

Slika 2: Izsek iz normativne zbirke osebnih imen CONOR (pridobljeno 12. 2. 2011 s spletne strani <http://www.cobiss.si/scripts/cobiss?ukaz=getid&clani=si>)

Tretje načelo od skupno enajstih v Iffinih smernicah zahteva, da med poimenovanji v sistemu za predmetno označevanje vzpostavimo semantične odnose, torej ekvivalenčne, hierarhične in asociativne. Načelo o sintaksi pa pravi, da imajo besedne zveze prednost pred semantičnimi odnosi. To načelo ni jasno. Manjka jasna odločitev o tem, kdaj imajo prednost besedne zveze in kdaj enostavni pojmi, npr. kdaj bomo ohranili besedno zvezo pokrajinski arhiv in kdaj jo bomo razstavili v dve besedi. Mednarodni standard za enojezične tezavre (ISO 5964:1985) je pri tem jasnejši, saj pravi, da besedno zvezo razdelimo le, če se pri tem ne izgubi smisel. Odločitev je torej odvisna od tega, ali je besedna zveza samostojni poimenovanje ali ne, in kakšno škodo utegnemo povzročiti pri iskanju oz. priklicu iz

1. Na sliki 2 se Vladimir Lavrač pojavi na seznamu Janezov Novakov zato, ker je ime Janez Novak uporabljal kot psevdonim. To vidimo pri pregledu normativnega zapisa.

baze, če jo razstavimo. To lahko razložimo že na naslednjem načelu, ki pravi, da moramo vsebinske oznake oblikovati dosledno v istem smislu. Če se torej odločimo za predmetno oznako »zgodovina arhivov«, bomo morali imeti tudi »zgodovino knjižnic«, »zgodovino muzejev«, »zgodovino gledališč« itd. To nam seveda zelo poveča obseg geslovnika in s tem ceno njegovega vzdrževanja. Inverzijo v besedni zvezi, kot je npr. »avtoceste, gradnja«, smo bibliotekarji opustili, ko so računalniški sistemi začeli enakovredno indeksirati in iskati po katerikoli besedi v besedni zvezi.

Imamo zelo elegantno rešitev zadrege s sestavljenimi pojmi kakršen je zgornji primer, ki se je premalo poslužujemo, prinaša pa ga fasetna zasnova. Fasete grupirajo pojme po neki značilnosti, pri čemer razvrščanje poteka s kombinacijo teh grup (Guidelines, 2009). Najbolje lahko fasetni princip prikažemo z vsebino in geografskimi pojmi. Imamo lahko npr. sistem, v katerem dosledno prikazujemo književnost po deželah oz. jezikih - npr. danska književnost, estonska književnost, finska književnost itd.; potem moramo zaradi prejšnjega načela imeti tudi dansko glasbo, estonsko glasbo, finsko glasbo itd. ter dansko slikarstvo, estonsko slikarstvo, finsko slikarstvo itd. Toda če se odločimo za fasete, pripravimo seznam dežel - Danska, Estonija, Finska itd. ter seznam umetnosti, kot so književnost, glasba, slikarstvo. Dežele pa lahko razen z umetnostmi poljubno kombiniramo še z drugimi temami.

Dve načeli govorita tudi o tem, od kod je treba jemati besedišče za predmetne oznake, torej o izvoru poimenovanj. Izhajali naj bi iz knjižničnega gradiva, a bi morali biti tudi prilagojeni besedišču, ki so ga vajeni uporabniki. To pomeni, da imamo lahko več poimenovanj, a je za normativno izbrana ena oblika, druge so kot kazalke vključene v normativno zbirko. Da bi bilo to za vse uporabnike tudi jasno, je smiselno zagotoviti tak izpis na zaslon, kot ga predvideva programski sistem ScopeArchiv v meniju iskanja po deskriptorjih (slika 3).

The screenshot shows a list of search results for 'Gospodstvo' (Manor) in the ScopeArchiv system. The results are as follows:

- Gospodstvo Bajnhof (Pravne osebe\)
- Gospodstvo Bequnje (1338-1765) (Pravne osebe\)
- Gospodstvo Bekštani (Pravne osebe\)
uporabljeno za:
Gospodstvo Finkenstein (Pravne osebe\)
- Gospodstvo Bekštani (Finkenstein) (Pravne osebe\)
- Gospodstvo Bela Peč (1431-1887) (Pravne osebe\)
- Gospodstvo Belnek (Pravne osebe\)
glej tudi:
Graščina Belnek (Pravne osebe\)
- Gospodstvo Bistra (Pravne osebe\)
- Gospodstvo Bistriški grad (1256-1945) (Pravne osebe\)
- Gospodstvo Bled (1004-1871) (Pravne osebe\)

Annotations in the image:

- A box pointing to 'Gospodstvo Bekštani (Pravne osebe\)' contains the text: 'Neveljavna oblika imena kaže na veljavno (kazalka)'. An arrow points from this box to the entry 'Gospodstvo Bekštani (Pravne osebe\)'.
- A box pointing to 'Gospodstvo Belnek (Pravne osebe\)' contains the text: 'Eno veljavno ime kaže na drugo, sorodno in tudi veljavno (vodilka)'. An arrow points from this box to the entry 'Gospodstvo Belnek (Pravne osebe\)'.

Slika 3: Prikaz veljavnih in neveljavnih gesel ter napotil v sistemu ScopeArchiv Arhiva Republike Slovenije (pridobljeno 5. 7. 2011 s spletne strani <http://arsq.gov.si/Query/deskriptorensuche.aspx>)

Posebno načelo o specifičnosti pravi, da je vedno treba izbrati najbolj podrobno predmetno oznako, ki jo sistem za predmetno označevanje premore. Če torej naša publikacija obravnava domačo muco, naš sistem pa vsebuje tako geslo domače mačke kot tudi geslo mačke, mora katalogizator izbrati geslo domače mačke. Ni pa pravilno ustvarjati verižne oznake, ki prikazuje celotno hierarhijo, npr. domače mačke, mačke, zveri, sesalci. Prikazovanje te hierarhične verige je zagotovljeno z odnosi med pojmi, ki so ustvarjeni v geslovniku in normativni datoteki, ne pa s predmetnimi oznakami v posameznem opisu publikacije (kataložnem zapisu). Slovenski knjižničarji se tega pravila ne zavedajo dobro in

delajo veliko napak. Pravilno pa je zasnovan Splošni geslovnik COBISS.SI, ki žal še ni implementiran kot normativna kontrola (slika 4).

Uporabniški		GSARE
Občno ime	Bibliotekarstvo	
Občno ime (ENG)	Library science	
Pomenski obseg	Pod to značilno uršamo splošna dela o znanju in veščinah, potrebnih za organizacijo in poslovanje knjižnic. Dela o storitvah, ki jih knjižnice nudijo svojim uporabnikom, uršamo pod Library services.	
Uporabljeno za	Bibliotekarska znanost	
Uporabljeno za	Knjižničarstvo	
Širši izraz	Dokumentacija	
Širši izraz	Informacijska znanost	
Sorodni izraz	Bibliografija	
Sorodni izraz	Knjižnične storitve	
Ožji izraz	Katalogizacija	
Ožji izraz	Knjižnične ankete	
Ožji izraz	Obdelava gradiva	
Ustrezna oznaka LCSH	Library science	
Izvirna oznaka SEARS	Library science	
SGC-ID	1053028	

SPREMI NIŠKALNO ZAHTEVO

Legenda: Uporabljeno za = neveljavni izraz; Širši izraz, Ožji izraz – hierarhični odnos, Sorodni izraz-- asociativni odnos
 Slika 4: Prikaz semantičnih odnosov v Splošnem geslovniku COBISS.SI (pridobljeno 31. 01. 2011 s spletne strani <http://cobiss4.izum.si/scripts/cobiss?ukaz=DISS&id=0840279254490400&rec=2&sid=1&fmt=1>)

Ker je torej način predmetnega označevanja po eni strani odvisen od gradiva,

Ker je torej način predmetnega označevanja po eni strani odvisen od gradiva, po drugi od uporabnika, po tretji pa od informacijskega sistema, v katerem delamo (oz. računalniškega vmesnika), in podpornih sistemov in datotek, morajo vsak sistem za predmetno označevanje spremljati tudi navodila za uporabo. To so navodila za knjižničarja. Znano je namreč, da uporabniki knjižnic (bralci) neradi uporabljajo priročnike. Za končnega uporabnika mora pomoč in vodenje zagotoviti vmesnik.

Za primer si zamislimo iskanje s ključno besedo umetnost. COBIB s skoraj 4 milijoni zapisov vsebuje nad 65.000 ustreznih zapisov. Nihče ne bo pregledal celotnega seznama, da bi našel tisto, kar bi bilo primerno zanj. Zato bi sistem moral sam ustvariti podskupine po različnih kriterijih. Ponudil bi lahko naslednje skupine:

- po vrsti umetnosti (fotografija, glasba, slikarstvo itd.),
- po deželi (slovenska, hrvaška, italijanska itd.),
- po žanru (strokovna razprava, razstavniki katalog, izvedba itd.),
- po umetniškem slogu (barok, secesija, popart itd.).

Na podlagi podatkov, ki obstojijo v kataložnem zapisu, bi bila možna tudi drugačna grupiranja (glej tudi Šauperl, 2007). Če se torej odločimo za določen način grupiranja, moramo po eni strani zagotoviti ustrezne podatke in po drugi ustrezno oblikovanje uporabniškega vmesnika.

To prikazuje primer iskanja v katalogu neke ameriške univerzitetne knjižnice (slika 5). Najdeno gradivo lahko razvrstimo po različnih kriterijih (klasifikacijska oznaka, gesla, žanr) in s tem udobneje pregledujemo krajše sezname sicer neobvladljive množice skoraj 270.000 najdenih publikacij. Tu vidimo dober način pregledovanja in izbire, izvrstno omogočanje dveh izmed funkcij knjižničnega kataloga.

Alenka ŠAUPERL: Vloga kontroliranih slovarjev v podatkovnih zbirkah, 411-421

The screenshot shows a library search interface. On the left, there are three filter boxes: 'Your Current Search' (set to 'Anywhere'), 'Narrow Your Search' (with checkboxes for 'Currently available', 'Available online', and 'New titles'), and 'Subject' (with a list of categories like 'History (162733)', 'Biography (13654)', etc.). Below these are three 'Grupiranje po...' buttons. The main area displays search results for 'Manhood in America: a cultural history' and 'Sources in the history of the modern Middle East'. Each result includes author, publication year, format, and library location. The 'Manhood in America' result shows it is available in the 'D.H. Hill Library' stacks. The 'Sources in the history of the modern Middle East' result shows it is available as eBooks on various Kindles. The interface includes a 'Sort By' dropdown set to 'Relevance' and an 'Add to List' button for each item.

Slika 5: Dolg seznam priklicanih zapisov je mogoče pregledovati v manjših skupinah, oblikovanih po različnih kriterijih (pridobljeno 5. 0. 2011 s spletne strani <http://www.lib.ncsu.edu/>)

Iflina načela pa ne govorijo o tem, kakšno vlogo imajo gesla v katalogu. O tem govori že omenjeni FRSAD: najti, identificirati in izbrati ustrezno temo oz. poimenovanje ter raziskovati odnose med pojmi oz. temami. Če naj torej uporabnik v informacijskem sistemu raziskuje, najde, identificira in izbere ustrezno geslo, morajo imeti gesla določene lastnosti. FRSAD jih razvrsti v attribute entitet tema in poimenovanje ter odnose med entitetami delo in tema, tema in delo, delo in delo ter tema in tema. Atributi entitete tema so ilustrirani s primeri iz znanih geslovnikov in tezavrov, npr. s fasetami iz *Art and Architecture Thesaurus: Styles and Periods, Agents, Activities, Materials, Objects, Physical Attributes in Associated Concepts* (About, 2010). V *ScopeArchivu* lahko kot fasete vidimo družine, fizične in pravne osebe, stvarna gesla ter zemljepisna imena. Verjetno bi lahko dodali tudi žanr (odlok, zapišnik, pismo itd.) in še kaj. Tudi poimenovanje ima lahko več atributov, predvideni so: shema (npr. Univerzalna decimalna klasifikacija ali *Art and Architecture Thesaurus*), vir (kje smo poimenovanje našli), jezik, pisava, oblika (popolno, okrajšava), čas uporabe (kako poimenovanje lahko zastari, npr. fiskultura), uporabniška skupina (npr. bibliotekarji, arhivisti), status (npr. predlog, zastarel). Odnosi so lahko hierarhični, asociativni in ekvivalenčni. Model pa predstavlja le osnovo, na kateri je treba zasnovati nacionalni ali mednarodni pravilnik. Daje nam torej osnovo, da lahko zasnujemo skupni geslovník ali tezaver za knjižnično, arhivistično ali muzejsko dejavnost. Narekuje pa tudi, naj IFLA v nekaj letih prilagodi starejše smernice in standarde, ki so že regulirali to področje.

Smernice za večjezične tezavre (*Guidelines, 2009*) so usmerjene le v tezavre. Razlika med geslovníki in tezavri je velika in neznatna hkrati: neznatna, ker lahko geslovníki upoštevajo pravila za gradnjo tezavrov in sprejmejo podoben način prikaza, velika pa zaradi komaj opazne podrobnosti, predmetnih nizov. Predmetni niz je npr. »Pokrajinski arhiv - Maribor - Vodniki«. V programskem sistemu *ScopeArchiv*, ki ga rabijo v arhivih, med stvarnimi gesli vidimo več invertiranih gesel, ki bi lahko bila predmetni nizi, npr. »avtoceste, gradnja« ali »meja, nelegalen prehod«.

Geslovníki navadno ne prikazujejo semantičnih odnosov med pojmi dosledno pri vsakem pojmu, kar je za tezaver obvezno. Ti odnosi so v geslovníkih navadno prikazani le z napotili »rabi« ali »glej« in »rabi za« (vzratna povezava ekvivalenčnega odnosa, torej kazalka) ter »glej tudi« (za označevanje hierarhičnih in asociativnih odnosov, torej vodilka). V tezavrih imamo vedno standardizirane oznake za hierarhični odnos BT (širši izraz), NT (ožji izraz), RT (sorodni izraz) za asociativni odnos in USE/UF (sinonim) za ekvivalenčni odnos. Geslovníki vedno vsebujejo tudi navodila za oblikovanje predmetnih nizov in pogosto standardizirajo celotni predmetni niz. Sintaktičnih pravil v tezavrih nikoli ni, ker je vseeno, v katerem vrstnem redu navedemo uporabljene deskriptorje. Tudi Ifline smernice za večjezične tezavre vsebujejo priporočeni nabor oznak za semantične odnose. Semantični odnosi so

prikazani že v smernicah za predmetno označevanje. Ker so enaki, jih tu ne bomo ponavljali. Ker pa se smernice za večjezične tezavre osredotočajo na komplementarnost oz. kompatibilnost tezavrov v različnih jezikih, opozarjajo predvsem na problem različnih kultur in jezikov, ki ne omogočajo vedno neposrednega prevajanja.

4. NORMATIVNA KONTROLA

Normativna kontrola predmetnih oznak torej po eni strani pomeni vzpostavljanje seznama normiranih (torej sprejetih oz. standardiziranih) poimenovanj, po drugi strani pa tudi vključitev v uporabniški vmesnik za iskanje, združevanje in izbiro bibliografskih virov oz. informacijskih objektov. Normativna datoteka, ki to omogoča, je strukturirana na standardizirani način. IFLA je zato smernice pripravila že precej pred izidom FRSA Da (GARR). Izdala pa je tudi računalniški format za izmenjavo normativnih podatkov, UNIMARC Authorities. Slovenski vzajemni katalog COBIB se je tudi relativno zgodaj vključil v ta svetovni trend. Od leta 2003 deluje CONOR, normativna zbirka osebnih imen, ki je omenjena že zgoraj in prikazana na sliki 2. Normativna datoteka za predmetne oznake pa v Sloveniji še ni pripravljena. Toda tudi slovenska normativna datoteka osebnih imen v COBIB ni vključena na optimalni način. Naj to razložimo na primeru iskanja del Pavleta Sedmaka (glej Sliko 6). Ko iščemo s tem imenom, prikličemo seznam vseh del Franceta Bevka. Nikjer na seznamu pa ne vidimo imena Pavle Sedmak. Če pogledamo posamezno delo, se nam samo v načinu pregleda COMARC (to je format za mednarodno izmenjavo, v katerem so vsi podatki razporejeni po posameznih poljih in podpoljih), prikaže seznam vseh imen, ki jih je France Bevk uporabljal. Našli smo vsa njegova dela, ne vemo pa, v katerem se je podpisal z imenom Pavle Sedmak. Prikaz bi moral biti jasnejši: razen informacije, da je Pavle Sedmak eden od psevdonimov Franceta Bevka, bi morali izvedeti tudi, katera druga imena je uporabljal in katera dela so izšla pod katerim od imen.

Rezultati izbirnega iskanja

Iskali ste: Avtor=**pavle sedmak** IN jezik=**slv** IN Izbor zapisov=**knjige**

Število najdenih zapisov: **376**

Razvrstitev po: **COBISS.SI-ID**

[zapisi: 1-10]

	Št.	Avtor	Naslov	Vrsta gradiva
<input type="checkbox"/>	1.	Bevk, France	Krvavi jezdec : prva knjiga romana iz XIV. stoletja "Znamenja na nebu"	roman
<input type="checkbox"/>	2.	Bevk, France	Črni bratje	kratka proza
<input type="checkbox"/>	3.	Bevk, France	Lukec in njegov škorec : roman za otroke	kratka proza
<input type="checkbox"/>	4.	Bevk, France	V mestu gorijo luči	knjiga
<input type="checkbox"/>	5.	Bevk, France	Kaplan Martin Čedermac	roman
<input type="checkbox"/>	6.	Bevk, France	Peter Klepec	slikanica
<input type="checkbox"/>	7.	Žgur, Fran	Fran Žgur, vipavski slavček	poezija
<input type="checkbox"/>	8.	Bevk, France Cencič, Mira	Črni bratje : povest in resnična zgodba	kratka proza
<input type="checkbox"/>	9.	Bevk, France	Kaplan Martin Čedermac : roman	roman
<input type="checkbox"/>	10.	Bevk, France	Otroška leta	kratka proza

Slika 6: Poizvedba za deli Pavleta Sedmaka obrodi seznam vseh del Franceta Bevka (Pavle Sedmak je eden od Bevkovih psevdonimov)

V majhnih podatkovnih zbirkah imamo navadno občutek, da obvladujemo podatke in besedišče, zato se nam zdi dolgočasno in zahtevno delo z normativno kontrolo odveč. Toda ko pri iskanju s preprostim izrazom pridobimo nekaj deset, sto ali več zadetkov, postane vloga normativne kontrole ključna. Tedaj nam odnosi, ki so v njej vzpostavljeni med pojmi, omogočajo pregledovanje in gibanje

znotraj podatkovne zbirke. Omogočajo nam, da v kopici sena najdemo iglo. Geslovniki, ki so osnova za normativne datoteke, zahtevajo zelo veliko intelektualnega dela, zato se jim marsikje izogibajo. Toda to je nujno delo, saj avtomatski postopki tu niso uspešni. Intelektualne osnove organizacije informacij ostajajo že stoletja enake. Spreminja se le tehnologija.

VIRI

About the AAT. V *Art and Architecture Thesaurus Online*. Los Angeles, the Getty Trust. Dostopno 5. 7. 2011 na <http://www.getty.edu/research/tools/vocabularies/aat/about.html#scope>.

Angleško-slovenski slovar bibliotekarske terminologije. Ljubljana 2002. Dostopno 5. 7. 2011 na <http://www2.arnes.si/~ljnuk4/dictionary/slovenian/>

Guidelines for multilingual thesauri. The Hague 2009. Dostopno 5. 7. 2011 na <http://archive.ifla.org/VII/s29/pubs/Profrep115.pdf>.

Ifline smernice za predmetno označevanje. V *Spletni splošni slovenski geslovník*. Ljubljana, 2002. Dostopno 5. 7. 2011 na <http://www.nuk.uni-lj.si/ssg/navodila.html#0%20IFLINE%20SMERNICE%20ZA%20PREDMETNO%20OZNA%20C4%8CEVANJE>

Izjava o mednarodnih katalogizacijskih načelih. Haag 2009. Dostopno 5. 7. 2011 na http://www.ifla.org/files/cataloguing/icp/icp_2009-sl.pdf

ISO 5964:1985 Documentation - Guidelines for the establishment and development of multilingual thesauri. Geneva : ISO.

IVAN KANIČ ... [in dr.], *Bibliotekarski terminološki slovar*. Ljubljana 2009.

TATJANA KOVAČ ... [in dr.], *Navodila za predmetno označevanje*. Ljubljana 2002. Dostopno 5. 7. 2011 na <http://www.nuk.uni-lj.si/sssg/navodila.html>.

PATRICE LANDRY, *Guidelines for multilingual thesauri*. V *IFLA World library and information congress. 75th general conference and assembly*, Milan 2009. Dostopno 5. 7. 2011 na <http://www.ifla.org/files/hq/papers/ifla75/215-landry-en.pdf>

MIROSLAV NOVAK, *Preslikave vsebin v arhivskih strokovnih postopkih*. Maribor 2007.

ZDENKA SEMLIČ RAJH in ALENKA ŠAUPERL, *Problematika gesljenja in oblikovanja deskriptorjev*. V S. Tovšak ... [in dr.] (ur.), *Tehnični in vsebinski problemi klasičnega in elektronskega arhiviranja: Zbornik referatov z dopolnilnega izobraževanja* (str. 187-213). Maribor 2011

ALENKA ŠAUPERL, *Klasifikacija knjižničnega gradiva*. Ljubljana 2000.

ALENKA ŠAUPERL, *Ifline smernice za predmetno označevanje: prednosti in slabosti v luči spoznanj o vedênju uporabnikov*, „Knjižnica”, 49(2005), št. 1-2, str. 9-26.

ALENKA ŠAUPERL, *Malo znane lastnosti klasifikacije - Kratek vpogled v teorijo klasifikacije in njeno uporabo v knjižničnih katalogih*. V: S. Tovšak ... [in dr.] (ur.), *Tehnični in vsebinski problemi klasičnega in elektronskega arhiviranja: Zbornik referatov z dopolnilnega izobraževanja* (str. 83-95). Maribor 2007

JOŽE URBANIJA, *Metodologija izdelave tezavra*. Bibliothecaria 1. Ljubljana 1996.

MATJAŽ ZALOKAR, *Priprava splošnega geslovníka COBISS.SI*. Dostopno 5. 7. 2011 na http://splet02.izum.si/cobiss-oz/news.jsp?apl=/2004_4/ar09.jsp.

M.L. ZENG, M. ŽUMER, *Introducing FRISAD and mapping it with SKOS*. In: *World Library and Information Congress, 75th IFLA General Conference and Assembly*, Milan 2009. Dostopno 5. 7. 2011 na <http://www.ifla.org/files/hq/papers/ifla75/200-zeng-en.pdf>

M.L. ZENG, M. ŽUMER. in A. SALABA, (eds.). *Functional Requirements for Subject Authority Data (FRISAD) : A Conceptual Model*. The Hague. Dostopno 5. 7. 2011 na <http://www.ifla.org/files/classification-and-indexing/functional-requirements-for-subject-authority-data/frisad-final-report.pdf>.

SUMMARY

Libraries follow international standards and guidelines for subject description of publications and documents, i.e., for subject heading lists and classification systems. IFLA guidelines for subject heading languages are introduced. Eleven rules guide the establishment and development of subject heading lists, selection of subject headings and building of relationships among them, syntax, and give other recommendations. Functional requirements for subject authority data are a newer document. They establish the structure of authority files. These files contain subject heading lists and merge them with user interfaces of library catalogs. Authority control was established in 1980s, when the first UNIMARC Authority was published. This is a standard for international exchange of authority data. Authority control is crucial for large databases. It makes it possible for the user to search a certain topic by one term even when this term may be described by different terms (synonyms). The other advantage is that the relationships among terms may be used to navigate from one subject to a related one, to a broader or narrower topic. Names of creators can be controlled in the same way. Basic functions of the library catalogs and any other information databases, namely to find, identify, select, acquire and navigate within or outside the database, are realized through such authority control.

Original scientific article

Submitting date: 23.07.2011

Acceptance date: 30.07.2011