

# KLASJE

## STREŽNIK ZA STATISTIČNE

### KLASIFIKACIJE

Sandi Čemažar, Oracle Software d.o.o., WTC, Dunajska 156, Ljubljana  
scemazar@si.oracle.com

Lovro Munda, SRC.SI d.o.o., Tržaška 116, Ljubljana  
lovro.munda@link.si

Jožica Klep, Tatjana Šeremet, Andreja Vesel, Statistični urad Republike Slovenije, Vožarski pot 12, Ljubljana  
joza.klep@gov.si, tatjana.seremet@gov.si, andreja.vesel@gov.si

#### Izvleček

V članku so predstavljene temeljne predpostavke, izrazi in definicije ter funkcije strežnika za statistične klasifikacije. Nova programska rešitev osrednje zbirke statističnih klasifikacij je bila izdelana v okviru Programa posodobitve državne statistike. Rešitev, ki smo jo razvili na SURS, je zasnovana na dveh modelih - New Zealand Statistics Classifications and Related Standards System (CARS) in na modelu, kot smo ga povzeli po ISO/IEC 11179 Information technology - Specification and standardisation of data elements.

#### Abstract

*The paper describes basic assumptions, terms, definitions and functioning of the classification server which was defined as a part of development programme of government statistics. The solution is based on New Zealand Statistics Classifications and Related Standards System (CARS) and on the model, described in ISO/IEC 11179 Information technology - Specification and standardisation of data elements.*



#### Uvod

Ena od nalog Statističnega urada Republike Slovenije je tudi oblikovanje in vzdrževanje različnih klasifikacij, ki se uporabljajo za potrebe statističnih obdelav, pa tudi v poslovnem svetu. Take klasifikacije so, na primer, Standardna klasifikacija dejavnosti, Standardna klasifikacija poklicev, Klasifikacija gradbenih objektov, Nomenklatura industrijskih proizvodov, itd. Klasifikacije se sčasoma dopolnjujejo in spreminjajo, kar zahteva na eni strani skrb za organizirano oblikovanje in uveljavljanje novih verzij ter arhiviranje starih, na drugi strani pa je potrebno zagotoviti primerljivost rezultatov statističnih obdelav v daljšem obdobju, v katerem so bile veljavne različne verzije klasifikacije.

Avtorji verjamemo, da je obravnavanje klasifikacij lahko problematično pri gradnji podatkovnih skladišč, zato smo pripravili prikaz rešitve, ki smo jo zgradili. Na pomembnost obravnavanja šifer v svoji knjigi opozarja tudi Kimball, ki pravi, da moramo za vsako šifro vedeti njen pomen<sup>1</sup>.

#### 1. Izrazi in definicije

V nadaljevanju bomo na kratko predstavili uporabljene izraze in njihove definicije (slika 1).

#### Klasifikacija

Klasifikacija je niz povezanih kategorij, ki se uporabljajo za razvrščanje podatkov.

Klasifikacija je lahko: standardna, nestandardna ali neodvisna.

#### Raven

Raven je v klasifikaciji niz kategorij, s katerimi opazovano populacijo lahko razvrstimo do zahtevanih podrobnosti.

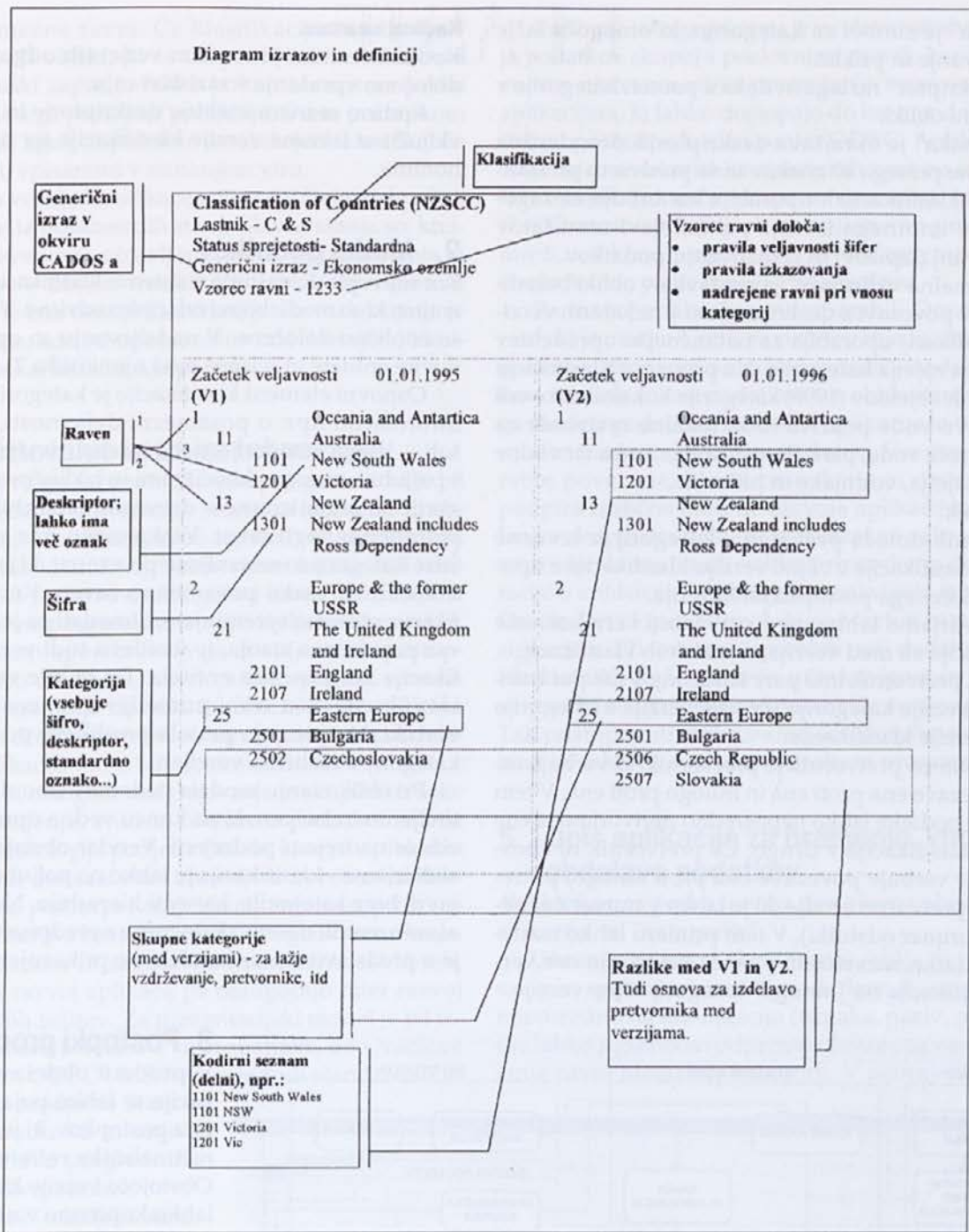
Vsaka klasifikacija ima vsaj eno raven. Ravni imajo lahko različna imena (nazive), na primer: področje, oddelek, skupina, razred, itd.

#### Verzija

Verzija klasifikacije je niz kategorij klasifikacije, ki se uporablja, oz. je na razpolago v določenem časovnem intervalu. O novi verziji govorimo, ko se spremenijo kategorije.

Lastnik klasifikacije lahko odloči, ali bo oblikoval novo verzijo klasifikacije kot prvo verzijo nove klasifikacije.

Spremembe kategorij, ki vplivajo na število ravni in na vzorec klasifikacije, se morajo obvezno odraziti kot oblikovanje prve verzije nove klasifikacije.



Slika 1: KLASJE, dciagram izrazov in definicij

Izraz verzija se nanaša na časovno obdobje, v katerem je prišlo do zaporednih ažuriranj ali revizij standardnih, nestandardnih ali neodvisnih klasifikacij. Vsaka klasifikacija se praviloma začne s prvo verzijo, ki se postopoma, v nekem časovnem obdobju, razvije v zaporedne verzije.

Vsaka verzija klasifikacije ima določen datum, od katerega velja. Na primer, prva verzija Klasifikacije gradbenih objektov velja od 1. 1. 1998.

V določenem časovnem obdobju bomo imeli vrsto verzij, vendar bo imela vsaka svoj, določen datum, od katerega velja. Veljavna verzija je običajno tudi zadnja verzija izbrane klasifikacije. Podatki so praviloma šifrirani skladno z veljavno verzijo.

#### Kategorija

Kategorija je del populacije statističnih enot, razdeljenih skladno z izbranimi opazovanimi lastnostmi teh enot.

“Šifra” je simbol za kategorijo, ki omogoča lažje shranjevanje in priklic.

“Deskriptor” razlaga in določa pomen kategorije v tekstualni obliki.

“Oznaka” je okrajšava deskriptorja. Standardna oznaka ne presega 40 znakov in se predvsem pri standardnih klasifikacijah uporablja kot orodje za zagotavljanje enotnega izrazja pri izdelavi rezultatov raziskovanj (izpisov) in izkazovanju podatkov.

“Formalna vključitev” je postavka v obliki besedila, ki se v povezavi z deskriptorjem v nekaterih verzijah klasifikacij uporablja za natančnejšo opredelitev ključnega zajetja kategorije. Na primer: v Klasifikaciji gradbenih objektov (1999) kategorija Lokalni cevovodi za dobavo vode pokriva tudi: lokalne cevovode za dovod vroče vode, pare ali stisnjenega zraka ter vodne stolpe, zajetja, vodnjake in hidrante.

### Pretvornik

Pretvornik določa prevajanje kategorij iz izvorne verzije klasifikacije v ciljno verzijo klasifikacije z uporabo določenega postopka ali kriterija.

Pretvarjamo lahko med različnimi verzijami iste klasifikacije ali med verzijami različnih klasifikacij.

Vsak pretvornik ima pare kategorij. Vsak par kategorij povezuje kategorijo izvorne verzije s kategorijo ciljne verzije klasifikacije.

Enostaven pretvornik je pretvornik, ki vsebuje samo povezave ena proti ena in mnogo proti ena. V tem primeru podatke lahko neposredno pretvorimo iz ene verzije klasifikacije v drugo. Če pretvornik ni enostaven (če vsebuje povezave ena proti mnogo) potrebujemo pretvorno orodje, ki je lahko v numerični obliki (na primer odstotki). V tem primeru lahko razdelimo podatke, razvrščene v “eno” kategorijo ene verzije klasifikacije na “mного” kategorij ciljne verzije.

### Kodirni seznam

Kodirni seznam je seznam verjetnih odgovorov na določeno vprašanje v raziskovanju.

Kodirni seznam vsebuje deskriptorje in formalne vključitve izbrane verzije klasifikacije ter njihove sinonime.

## 2. Model podatkov

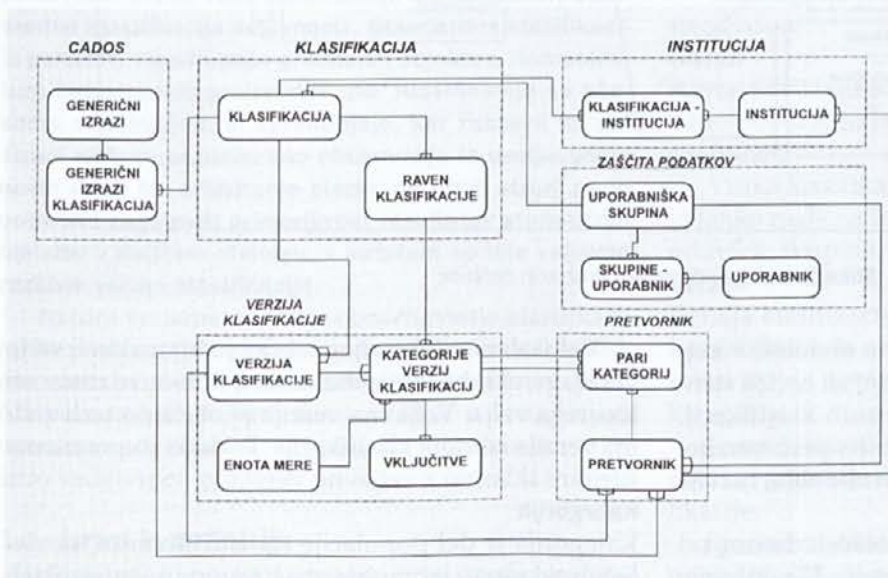
Kot smo opisali, nastopa v sistemu klasifikacij množica entitet, ki so medsebojno relacijsko odvisne. Vse relacije so enolično določene. V nadaljevanju so opisane osnovne entitete in relacije med njimi (slika 2).

Osnovni element klasifikacije je kategorija, ki nosi informacijo npr. o posamezni dejavnosti, poklicu, šoli... Vsako kategorijo je možno bolj detajlno opisati s poljubno mnogo vključitvami in izključitvami. Kategorije so organizirane v drevesno hierarhijo, ki ima poljubno mnogo ravni. Vsaka raven ima svoje ime. Šifre kategorij so sestavljene po vzorcu, ki je enolično določen za vsako posamezno raven. Ker se klasifikacije sčasoma spreminjajo, ohranjati pa je potrebno vsa posamezna stanja, je uvedena tudi verzija klasifikacije kot posebna entiteta. Da bi bile verzije iste klasifikacije med seboj primerljive, so uvedeni pretvorniki, ki opisujejo pravila preslikave posameznih kategorij v različnih verzijah.

Pri oblikovanju modela podatkov smo sledili ugotovljenim relacijam, ki na koncu vedno opredeljujejo odnos nadrejeni-podrejeni. Vendar obstaja v praksi tudi izjema - klasifikacija je lahko na poljubni vmesni ravni brez kategorije, kar ruši hierarhijo. Na taki ravni smo uvedli umetno kategorijo s predpisano šifro, ki je v predstavitvenem modelu ne prikazujemo.

## 3. Postopki procesa

V procesu obdelave klasifikacije se lahko pojavlja množica postopkov, ki jih izdelana računalniška rešitev podpira. Obstoječo verzijo klasifikacije lahko kopiramo v novo verzijo ali celo v novo klasifikacijo. Kategorijam klasifikacije lahko spreminjamo posamezne lastnosti. Kategorije lahko dodajamo ali brišemo, seveda v takem zaporedju, ki ne ruši hierarhije. Do popolnoma nove klasifikacije je možno priti na več načinov. Lahko jo kreiramo od začetka, določimo ravni in vzorec šifer kategorij, nato pa vpisujemo kategorije



Slika 2: Shematični prikaz ključnih entitet

na posamezne ravni. Če klasifikacija že obstaja v kakršnikoli obliki, ki se da pretvoriti v tekstovni računalniški zapis, jo lahko polnimo v sistem preko univerzalnega postopka polnjenja. S tem postopkom lahko tudi zgolj dopolnimo že obstoječo klasifikacijo s podatki, vpisanimi v zunanjem viru.

Vsaka verzija klasifikacije ima svoj življenjski cikel, sestavljen iz posameznih stanj. Glavna stanja so: kreiranje, preverjanje, objavlanje, umik v popraviljanje in opustitev. V danem trenutku je lahko objavljena in veljavna le ena verzija klasifikacije. Za vsako stanje je predpisano, kdo od uporabnikov ima dostop do klasifikacije in kaj lahko z njo počne.

#### 4. Načrtovanje in izbira tehnologije

Pri načrtovanju računalniško podprtega sistema je bil prvi cilj izdelati univerzalen sistem za hranjenje in obdelavo vseh klasifikacij, ki bo zajemal vse specifičnosti posameznih klasifikacij. Sistem mora zagotavljati konsistentnost podatkov, omogočati mora le dovoljene postopke za posamezno stanje življenjskega cikla, posameznim uporabnikom pa dovoljevati le postopke, ki so jim dovoljeni, kar vključuje vpogled, spreminjanje in kreiranje klasifikacij. Sistem naj bo odprt in dostopen drugim sistemom, npr. statističnim obdelavam. Aplikacija mora biti dokumentirana in razvita z orodjem, ki omogoča nadgrajevanje in skupinsko delo.

Za hranjenje, vnos in obdelavo podatkov je bila izbrana tehnologija podjetja Oracle, ki se je v praksi že mnogokrat dokazala kot zanesljiva in robustna tehnologija na področju kompleksnih informacijskih sistemov. Oracleova baza podatkov omogoča varno hranjenje podatkov in zagotavlja konsistentnost le-teh, orodja za razvoj aplikacij pa omogočajo hiter razvoj aplikativnih rešitev. Za prezentacijski model je bil izbran Microsoft Access z dostopanjem do Oracleove baze preko ODBC<sup>2</sup>. MS Access je tudi standardno in razširjeno orodje na Statističnem uradu RS in omogoča izdelavo preglednih vpoglednih modulov.

Objektno-relacijska baza podatkov Oracle<sup>8i</sup><sup>3</sup> omogoča izdelavo rešitev opisanih zahtev. Tako smo pri oblikovanju aplikacije največ aplikacijske logike in pravil namestili neposredno v bazo podatkov v obliki referenčnih omejitev (constraints), prožilcev (triggers) in procedur, funkcij ter paketov (stored procedures).

Prednost takega pristopa je v celovitosti shranjevanja podatkov skupaj s poslovnimi pravili, kar omogoča enoten nivo varnosti in prezentacije vsem orodjem in aplikacijam, ki lahko dostopajo do baze podatkov, bodisi neposredno bodisi preko ODBC. Aplikacija pri odjemalcu se tako lahko osredotoči zgolj na uporabniški vmesnik. Tak način prispeva nenazadnje tudi k boljši odzivnosti aplikacije, k zmanjšanemu prometu po mreži in h koncentraciji moči strojne opreme na enem strežniku, ne pa na množici odjemalcev.

Za razvoj aplikacije smo uporabili Oracle Designer<sup>4</sup>, ki je orodje CASE za načrtovanje in oblikovanje aplikacij. Omogoča analizo procesov, modela podatkov, funkcij, tokov podatkov ter oblikovanje fizičnega modela podatkov, modulov aplikacije in tudi gradnjo fizičnega modela ter modulov. Vse faze so med seboj povezane, hkrati pa nobena ni obvezna. Ker podpira različne metode razvoja aplikacij in kontrolirano skupinsko delo, se je izkazal za zelo učinkovito orodje, ki nudi tudi veliko pripravljenih dokumentacije o aplikaciji. Aplikacija je sestavljena iz modulov Oracle Developer Forms<sup>5</sup>, ki so v celoti razviti z Designerjem. Pri tem smo uporabili tudi dodatek Oracle Headstart<sup>6</sup>, ki še poveča učinkovitost in nudi že pripravljeno okolje za začetno oblikovanje aplikacije. Tak pristop nudi maksimalno osredotočenost na vsebino in zahteva najmanj rutinskega programiranja.

#### 5. Opis aplikacije za hranjenje, vnos in obdelavo podatkov

Aplikacija ima preko menija dva glavna vhoda - "klasifikacija" in "kategorije".

Obrazec za vnos klasifikacije (slika 3) omogoča vnos in spremembo vseh podatkov, ki se nanašajo neposredno na klasifikacijo (oznaka, naziv, opis...). Z nje lahko z gumbom odpremo obrazec za vnos in urejanje ravni hierarhije kategorij. V zgornjem delu so

The screenshot shows a software interface for managing classification versions. It features a tree view on the left showing the hierarchy: SKP (top level), Vojaske (sub-level), Zakonodajalci (sub-level), and Strokovnjaki (sub-level). Below the tree is a table for 'Podatki o verziji klasifikacije' (Classification Version Data) with columns for 'Verzija' (Version), 'Datum veljavnosti' (Valid from), 'Datum CAIS' (CAIS date), 'Sprejetost' (Approved), and 'Velevnost' (Status). The table shows a version '2000' with a valid date of '01.01.2000', CAIS date of '22.12.1999', and status 'POPRAVLJANJE' (Correction). A separate window titled 'Podatki o verziji klasifikacije' is also visible, showing the same data for version '2000'.

Slika 3: Vnos podatkov o verziji klasifikacije

prikazani osnovni podatki izbrane klasifikacije, spodaj pa je seznam ravni z vsemi pripadajočimi podatki. Poleg imena je pomembna številka ravni, ki določa globino v hierarhiji. Ko so na posamezne ravni dodeljene tudi kategorije, zaporedja ravni ni več možno spreminjati, za kar skrbi vgrajena funkcionalnost baze podatkov Oracle8i. Pomemben podatek ravni je tudi vzorec šifre kategorije, ki se uporablja za kontrolo vpisane šifre nove kategorije ter za določanje ravni pri polnjenju kategorij iz zunanjega vira.

Drugi vhod dostopa do kategorij in vodi do vseh postopkov, povezanih z obdelavo kategorije (slika 4). Obrazec je sestavljen iz treh sinhroniziranih blokov. V gornjem bloku je seznam klasifikacij (do katerih ima uporabnik dostop). V srednjem bloku se nahaja seznam verzij klasifikacije iz gornjega bloka. Od tu je možen dostop do detajlnih podatkov verzije, med katerimi je tudi informacija o stanju življenjskega cikla, v katerem se nahaja ta verzija. V spodnjem bloku je seznam kategorij na prvi ravni verzije klasifikacije iz srednjega bloka z gumbom za dostop do podrobnih podatkov o kategoriji. Taka razporeditev nudi visoko preglednost nad ključnimi entitetami in zmanjšuje možnost napačne gradnje hierarhije. Hkrati omogoča vpisovanje in pregled velike količine raznovrstnih podatkov, ki opisujejo vsako kategorijo. Ob izbrani kategoriji na prvi ravni lahko dostopamo do obrazca za vnos in pregled kategorij na nižjih ravneh. Ta ima zgoraj osnovne podatke o nadrejeni kategoriji (samo en zapis) skupaj s podatki o ravni, na kateri se nahaja. V spodnjem delu je seznam neposredno podrejenih kategorij. Na obrazcu sta tudi dva gumba za preskok na nižjo oz. višjo raven. Tak način omogoča zanesljivost pri razvrščanju kategorij, saj je vedno dobro vid-

na nadrejena kategorija, ki ji uporabnik mora vpisati podrejene kategorije. Preko gumbov pa se lahko spušča ali dviga po hierarhiji v zeleni smeri.

Iz obrazca za obdelavo kategorij je možen tudi dostop do obrazca za vnos vključitev in izključitev ter do modula za polnjenje kategorij, ki so na razpolago v zunanjem viru.

## 6. Problemi in rešitve

V nadaljevanju so opisane še nekatere zahteve, problemi in načini reševanja.

*Varnost podatkov* - za vsako fazo življenjskega cikla klasifikacije je predpisana matrika uporabnikov in načina dostopa do te klasifikacije (vpogled, spreminjanje, kreiranje).

Varnost podatkov smo dosegli z oblikovanjem podatkovnega modela vlog (roles) in uporabnikov z dodeljenimi vlogami ter z Oracleovo vgrajeno funkcionalnostjo za omejitve dostopa glede na dodeljeno vlogo.

*Enostavna in zanesljiva rešitev*, da uporabnik razvrsti kategorijo na pravilno raven in pravilni nadrejeni kategoriji.

Enostavnost in hkrati zanesljivost je dosežena z opisanim načinom prikaza nadrejene in podrejenih kategorij v formi za vnos kategorij na nižjih ravneh. Uporabnik vidi le eno nadrejeno kategorijo (skupaj s podatki o ravni, na kateri se nahaja), kateri vpisuje podrejene kategorije.

*Polnjenje podatkov kategorij*, ki že obstajajo v zunanjem viru; omogočeno mora biti polnjenje še neobstoječih kategorij, pa tudi polnjenje dotlej praznih stolpcev (atributov) že obstoječih kategorij.

*Novo verzijo klasifikacije naj bo možno kreirati s kopiranjem in naknadnim spreminjanjem obstoječe.*

Novo verzijo obstoječe klasifikacije največkrat seveda ne kreiramo "iz nič". Zato je izdelan postopek kopiranja obstoječe verzije v novo. Uporabnik najprej odpre novo verzijo, nato pa v pripravljenem obrazcu izbere izvorno verzijo klasifikacije. Ko sproži kopiranje, se vse kategorije izvorne verzije klasifikacije kopirajo v novo. Tudi ta postopek je izveden z baznim paketom PL/SQL in metodo dinamičnega SQL-a. Dinamični SQL zagotavlja neobčutljivost na spremembe definicije tabele in tudi hitreše izvajanje programa zaradi optimizacije, vgrajene v Oracleovo bazo podatkov. Ker se celotni

Slika 4: Vnos podatkov za kategorijo

postopek kopiranja izvaja na bazi, tudi ni odvečnega prometa po omrežju.

## 7. Uporabniški vmesnik

V MS Access<sup>7</sup>-u je narejen uporabniški vmesnik za dostop do baze klasifikacij, ki omogoča različnim notranjim in zunanjim skupinam uporabnikov različne načine preiskovanja in izpisovanja.

Poslovna pravila za izdelavo klasifikacij so zelo specifična in stroga, zato se nam je zdelo smiselno, da uporabniški del aplikacije, ki naj bi omogočal pregledovanje in izpis, ločimo od vnosa in vzdrževanja. Ena izmed prednosti te delitve je, da smo se zato pri izdelavi uporabniškega vmesnika lahko osredotočili na preglednost ter s tem ponudili uporabniku prijazno aplikacijo.

V uporabniškem delu aplikacije KLASJE lahko klasifikacijo poiščemo na dva načina.

V seznamu generičnih izrazov označimo izbrani izraz in prikaže se nam seznam vseh klasifikacij, ki so povezane s tem generičnim izrazom, in seznam tezavrov, v katerih se ta izraz pojavi (slika 5).

V ta namen smo klasifikacije povezali z generičnimi izrazi Eurostatovega tezavra CADOS<sup>8</sup>, ki smo ga na statističnem uradu prevedli in priredili v okviru projekta Informacijsko dokumentacijski center - IDC<sup>9,10</sup>.

V drugem koraku iz prikazanega seznama klasifikacij izberemo tisto, katere lastnosti in vsebino želimo pregledati.

Lastnosti izbrane klasifikacije so prikazane v treh sklopih, med katerimi preklaplamo s klikanjem na

jeziček kartice. Razporejene so v sklope: verzije, klasifikacija in ravni.

Kartica "Verzija" prikazuje podatke za vsako posamezno verzijo (oznaka verzije, začetni in končni datum veljavnosti, ali je klasifikacija standardna ali nestandardna in veljavnost verzije).

Kartica "Klasifikacija" prikazuje poleg kratice in naziva klasifikacije še naziv skupine uporabnikov, zadolžene za klasifikacijo in naziv institucije, ki je klasifikacijo sprejela.

Kartica "Ravni" prikazuje že omenjena stroga pravila gradnje ravni, ki določajo globino hierarhije pri vnosu kategorij za vsako posamezno klasifikacijo. Prikazani so število ravni, naziv ravni, vzorec ravni, šifre ravni in dolžina deskriptorja.

Z izbiro določene verzije klasifikacije se odpre nov obrazec, na katerem lahko pregledujemo vse kategorije izbrane verzije.

Drugi način iskanja klasifikacij je, da z gumbom "Iskanje klasifikacij" odpremo obrazec, na katerem poiščemo klasifikacijo iz seznama vseh klasifikacij v bazi ali pa vnesemo kratico ali del kratice v pripravljeno polje za vnos.

Pregled kategorij izbrane verzije klasifikacije je možen z odpiranjem drevesne strukture ravni ali pa z izbiro posamezne ravni (slika 6).

Pri vnosu novih in vzdrževanju obstoječih klasifikacij je uvedena stroga kontrola vnosa kategorij po točno predpisani hierarhiji. Ker ima vsaka posamezna kategorija urejena "družinska razmerja" - kdo je komu oče in kdo sin (parent - child relationship), smo

v Access-u uporabili objekt, imenovan TreeView<sup>11</sup>, ki omogoča drevesni prikaz podatkov. Način odpiranja kategorij znotraj posameznih ravni je tak, kot je npr. v Raziskovalcu urejeno odpiranje map. Iskanje kategorij je možno tudi z vnosom teksta ali delom teksta. Z izborom kategorije se odpre seznam vseh njej neposredno podrejenih kategorij, ki so na naslednji ravni. Lastnosti vsake posamezne kategorije postanejo vidne, ko izberemo kategorijo iz seznama.

Odpre se obrazec, na katerem je definicija kategorije, kratica nadrejene kategorije, seznam vključitev, ki podrobneje opredeljujejo

**Generični izrazi:**

- delovna aktivnost
- demografija
- ekonomsko ozemlje
- gospodarska panoga
- gradbeništvo
- industrijski proizvod
- izobraževanje
- prebivalstvo
- turizem
- zunanja trgovina

**Povezani tezavri:**

- SPLOŠNE STATISTIKE
- PREBIVALSTVO IN SOCIALNE RAZME

**Klasifikacije:**

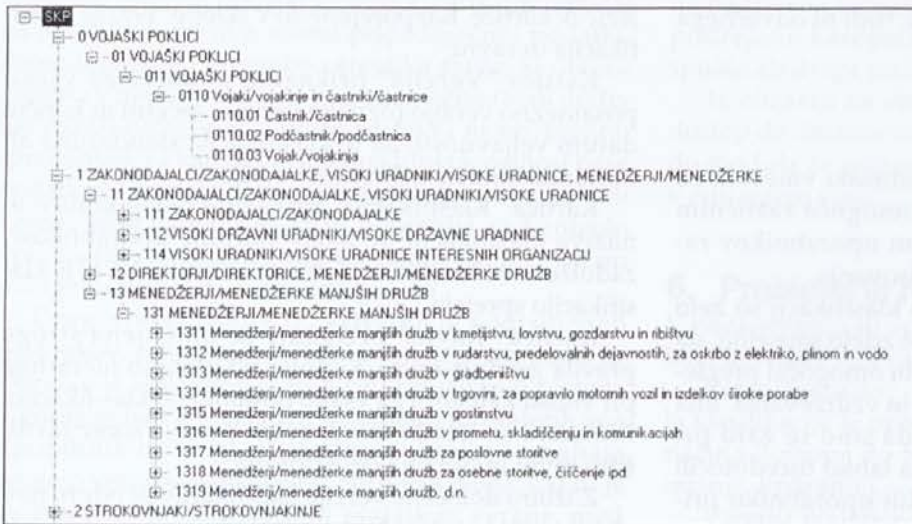
- demografija
  - SKP - Standardna klasifikacija poklicev
  - Rojstvo - Vrtni red rojstva

**Verzije | Klasifikacija | Ravni**

Št.ravni:	Naziv ravni:	Vzorec ravni:
1	glavna skupina poklicev	1
2	podskupina poklicev	12
3	področna skupina poklicev	123
4	enota podr. skupine poklicev	1234
5	naziv poklica	1234.55

Record: 1 of 5

Slika 5: Obrazec, na katerem poiščemo klasifikacijo; izbrali smo kartico z opisom ravni



Slika 6: Drevesni način prikaza hierarhije kategorij v standardni klasifikaciji poklicev

ključno zajetje kategorije, ter mejnih objektov, ki jih je potrebno uvrstiti drugam (seznam izključitev).

Drugi način pregledovanja klasifikacije je pregled vseh kategorij posamezne ravni. Iz seznama, ki prikazuje število ravni, naziv ravni, vzorec ravni in števila vseh kategorij za posamezno raven, izberemo eno in prikaže se seznam vseh pripadajočih kategorij izbrane ravni.

Uporabniški vmesnik omogoča tudi izdelavo različnih poročil za izbrano verzijo klasifikacije (slika 7), ki jih lahko izpišemo ali prenesemo v program Word ali Excel.

V uporabniški vmesnik smo vključili pregled tezavra CADOS, iz katerega smo za potrebe strežnika izbrali 214 generičnih izrazov. Za vsak izbran izraz dobimo seznam tezavrov, v katerem se ti izrazi pojavijo. Tudi vsebino izbranega tezavra lahko pregledujemo z odpiranjem drevesne strukture.

## 8. Zaključek

Izdelana aplikacija omogoča enostaven vnos in zanesljivo obdelavo podatkov poljubnih klasifikacij. Celotna definicija aplikacije je skupaj z definicijo podatkovnega modela shranjena v repozitoriju orodja Oracle Designer, kar zagotavlja zaščito opravljenega dela, dokumentiranost in transparentnost tudi za razvijalce, ki niso sodelovali na projektu.

Vse to pa skupaj z dejstvom, da so vsi moduli zgrajeni iz definicije v repozitoriju brez kasnejših popravkov, obljublja enostavno in poceni vzdrževanje aplikacije. Varno shranjene podatke pa lahko prikazuje poljubna aplikacija - v načrtu je tudi izdelava vpogleda v veljavne klasifikacije preko interneta.

### Povezava MS Access in Oracle preko vmesnika ODBC

Pri povezavi ODBC na podatkovni strežnik Oracle, smo naleteli na težave. Ugotovili smo, da Access v nekaterih primerih ne privzame

pravilnega primarnega ključa iz podatkovne baze Oracle. V takem primeru so lahko podatki, ki jih vrne strežnik, nepopolni ali pa celo napačni. Kasnejša poizvedovanja so pokazala, da je tako zaradi zgradbe vmesnika ODBC in da ne gre za hrošča v Accessu ali Oraclu. Vmesnik ODBC namreč ne razlikuje med različnimi vrstami indeksov iz tabele Oracleove baze. Access zato za primarni ključ izbere prvega iz abecedno urejenega seznama indeksov. Ena izmed rešitev, ki se ponuja sama od sebe, je ta, da preimenujemo indeks, ki predstavlja primarni ključ v Oraclu tako, da je v abecedno urejenem seznamu na prvem mestu.

Ugotovili smo torej, da:

- sami ne moremo vplivati na izbiro primarnega ključa,

Slika 7: Obrazec, na katerem izberemo vsebino in obliko poročila

- se upošteva pravilni primarni ključ na tabeli le, če v isti tabeli nimamo še kakšnega drugega enoličnega polja (unique constraint), ali če je indeks v abecednem redu na prvem mestu.

Strežnik je v fazi delujočega prototipa<sup>12</sup>. Večjih sprememb ne pričakujemo več, polna funkcionalnost prototipa pa bo dosežena z gradnjo modula za pretvornike in nekaterih dodatnih funkcionalnosti uporabniškega vmesnika. Predvidena je tudi izdelava rešitve za dostop in pregledovanje podatkov preko interneta.

Nova programska rešitev osrednje zbirke statističnih klasifikacij je bila izdelana v okviru Programa posodobitve državne statistike. Rešitev, ki smo jo razvili na SURS, je zasnovana na dveh modelih - New Zealand Statistics Classifications And Related Standards System (CARS)<sup>13</sup> in na modelu, kot smo ga povzeli po ISO/IEC 11179 Information technology - Specification and standardisation of data elements<sup>14</sup>. Naša rešitev ni identična kopija novozelandske<sup>15</sup> (razlik je precej), saj smo morali upoštevati okolje, ki velja na SURS in v državni upravi. To pomeni, da ni bilo moč kar kupiti gotovega izdelka (podprt je s Sybase), pač pa je bilo potrebno rešitev ponovno postaviti in informacijsko podpreti z Oraclom in MS Access-om, ki sta predvideni standardni programski opremljeni za delo pri nas. Ob tem smo veliko pridobili tudi ob študiju pravil in postopkov, s katerimi so Novozelandski podprli svoje delo.

## Viri:

- 1 Kimball R., The Data Warehouse Toolkit, Practical techniques for Building Dimensional Data Warehouses, John Wiley&Sons, Inc., 1996
- 2 Microsoft Developers Network (MSDN), <http://ddmsdn.microsoft.com>
- 3 Koch G., Loney K., Oracle8i - The Complete Reference; Oracle Press, Osborne
- 4 Peter Koletzke, dr. Paul Dorsey, Designer Handbook; Kenneth Atkins, Paul Dirksen, Zikri Askin Ince, Designer Generation; Oracle Press, Osborne
- 5 Robert J. Muller, Oracle Developer/2000 Handbook, Oracle Press, Osborne
- 6 <http://www.oracle.si/svetovanje/headstart.html>
- 7 Billings S., Rhemann J., et al., Access@ 97 Programming, SAMS Publishing, Indianapolis, 1997
- 8 CADOS, Theasaurus 1-9, Collection, Eurostat, March 89, Office for Official Publications of the European Communities, Luxembourg
- 9 Sirnik I.; Berce J., Informacijska ponudba na področju približevanja Slovenije Evropski uniji, Dnevi slovenske informatike 2000, Portorož, 19.-22. april 2000, zbornik posvetovanja, str. 143-148
- 10 Služba Vlade RS za evropske zadeve in Center Vlade RS za informatiko, Vodič po IDC - Informacijska podpora približevanju Republike Slovenije v Evropsko unijo, Ljubljana, December 1999
- 11 Mesojedec U., Visual basic, sodobno programiranje: priručnik za obvladovanje programiranja z visual basicom, Pasadena, Ljubljana, 1998
- 12 Cajhen J., Jereb Puhar V., Kavšek A., Klep J., Šeremet T., Vesel A., Strežnik za statistične klasifikacije - opis prototipne rešitve, v: Elektronsko poslovanje in statistika, Statistični dnevi '99, Radenci 1999
- 13 Dunnet G., The Components which lead to Statistics New Zealand's Output Database Environment, Output Database Workshop, Avstralija, 1999
- 14 ISO/IEC 11179-(E) Information technology - Specification and standardisation of data elements, Parts 1, 2, 3, 4, 5, 6
- 15 Dokumentacija CARS, ki nam jo je brezplačno odstopila Novozelandska statistika:  
CARS demo; CARS entity.xls; CARS/progman1.doc; logical model; physical model; views and stored procedures; accessing CARS using SAS; accessing CARS using Microsoft products; accessing CARS data using SQL Windows; coding and WINSCM; appendix one: list of main CARS views CARS/ur.mast.doc; CARS requirement specifications; CARS/urdefs.doc; Appendix A - terms and definitions  
CARS/urrules2.doc; Appendix B - CARS business rules; CARS/userdoc.doc

*Sandi Čemažar je diplomiral na Fakulteti za strojništvo in Visoki šoli za organizacijo dela, smer informatika. Dela v podjetju Oracle Software d.o.o. kot svetovalec za gradnjo informacijskih sistemov in specialist za orodja Oracle Designer, Oracle Headstart ter Oracle Developer v klasični (odjemalec-strežnik) ter v internetni tehnologiji.*

*Lovro Munda je ing. matematike. Po končanem študiju je delal na področju računalniškega izobraževanja. Ukvarja se z načrtovanjem informacijskih sistemov, podatkovnih baz in spletnih strežnikov. V projektu je sodeloval kot konzultant za MS Access.*

*Jožica Klep, prof.soc.ped., je svetovalka direktorja na Statističnem uradu RS. Sodeluje v programu posodobitve državne statistike. V programu projektov Informacijsko-dokumentacijski center (IDC), ki ga je leta 1998 definirala Vlada RS kot celovito informacijsko podporo približevanju Slovenije Evropski uniji, dela v projektu IDC6 - Dokumentni viri. Sodelovala je tudi pri projektu IDC2 - Prevodi in terminologija.*

*Tatjana Šeremet je diplomirana organizatorica-menedžerka informatike. Zaposlena je na Statističnem uradu RS v oddelku za Razvoj in administracijo baz podatkov. Ima več let delovnih izkušenj na področju priprave programske opreme za operativno izvajanje statističnih raziskovanj. Opravlja analizo in načrtovanje podatkovnih baz in izdeluje aplikacije v MS Access-u.*

*Andreja Vesel, univ.dipl.ing.rač., je diplomirala na Fakulteti za računalništvo in informatiko v Ljubljani v letu 1991. Dela na Statističnem uradu RS v Ljubljani v oddelku za Razvoj in administracijo baz podatkov. Opravlja predvsem analizo in načrtovanje podatkovnih baz (podatkovna baza Oracle 8i) in razvija programe za vnos in polnjenje podatkov z Oraclovimi razvojnimi orodji.*