

# Uporaba podatkovnih skladišč v univerzitetnem okolju: primer Univerze v Ljubljani

Viljan Mahnič, Marko Pozenel

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Tržaška 25, Ljubljana

Viljan.Mahnic@fri.uni-lj.si, Marko.Pozenel@fri.uni-lj.si

## Povzetek

V članku opisujemo projekt izgradnje podatkovnega skladišča, ki poteka na Univerzi v Ljubljani z namenom, da bi vzpostavili enoten vir podatkov za razne analize študijskega procesa. V prvem delu članka je opisan namen projekta in arhitektura predlaganega podatkovnega skladišča. Nato sledi podrobna predstavitev dimenzijskih podatkovnih modelov za posamezna področna podatkovna skladišča in opis orodij, ki jih uporabljamo za realizacijo. Na koncu podajamo še pregled možnih analiz in nekaj tipičnih vzorcev poročil, ki so na voljo uporabnikom.

Ključne besede: podatkovno skladišče, področno podatkovno skladišče, dimenzijski podatkovni model.

## Abstract

**Data warehousing in university environment: the case of the University of Ljubljana**

We describe a data warehouse project that has been started at the University of Ljubljana with the aim of obtaining unified and integrated source of data for various analyses of the educational process. A description of the scope of the project and the overall data warehouse design is first given, followed by a detailed description of the underlying dimensional model and implementation tools. Finally, a survey of possible analyses is given and some typical end user reports are described.

Keywords: data warehouse, data mart, dimensional model.

## 1 Uvod

Od začetka devetdesetih let prejšnjega stoletja, ko je Bill Inmon [6] izoblikoval koncept podatkovnih skladišč, je skladiščenje podatkov postalo ena izmed najpomembnejših vej na področju informacijskih sistemov [2]. Mnoga velika podjetja so pričela graditi obsežna podatkovna skladišča, ki služijo kot vir podatkov za sisteme za podporo odločanju, in mnogim izmed njih se investicije bogato obrestujejo [17].

Tudi v univerzitetnih okoljih smo priča različnim projektom s tega področja [14, 1, 5, 3, 11]. Medtem ko Stevenson [14] opisuje predvsem izkušnje in kritične dejavnike uspeha pri izgradnji podatkovnega skladišča na eni izmed avstralskih univerz, je v [1] predstavljen vsedržavni projekt izgradnje podatkovnega skladišča za potrebe francoskih univerz. V [3] je osrednja pozornost namenjena zgodnejšim fazam v življenjskem ciklu razvoja podatkovnega skladišča, ostali projekti pa obravnavajo podatkovna skladišča za posamezna področja. Tako je v [5] predstavljeno podatkovno skladišče za spremljanje raziskovalne dejavnosti, v [11] pa podatkovno skladišče za analizo učinkovitosti e-izobraževanja.

V prispevku bomo predstavili projekt izgradnje podatkovnega skladišča, ki poteka na Univerzi v Ljubljani z namenom, da bi vzpostavili enoten vir podatkov za razne analize študijskega procesa. Članek je sestavljen iz petih delov. V drugem delu opisujemo namen projekta in arhitekturo celotnega podatkovnega skladišča. V tretjem in četrtem delu sta podrobno predstavljene področni podatkovni skladišči s podatki o vpisanih študentih in diplomantih, ki sta bili realizirani najprej. V petem delu so opisana orodja, ki smo jih uporabili pri realizaciji, v šestem delu pa je podan pregled možnih analiz in nekaj tipičnih vzorcev poročil, ki so na voljo uporabnikom.

## 2 Namen projekta in zasnova podatkovnega skladišča

Končni cilj našega projekta je izgradnja celovitega podatkovnega skladišča za področje študijske informatike, ki bo sestavljeno iz verige štirih področnih podatkovnih skladišč (angl. data marts): področnega podatkovnega skladišča za analizo prijav za vpis,

področnega podatkovnega skladišča za analizo vpisnih podatkov, področnega podatkovnega skladišča za analizo opravljanja študijskih obveznosti in področnega podatkovnega skladišča za analizo podatkov o diplomantih [10]. Ker je Univerza v Ljubljani sestavljena iz 26 članic, ki imajo vsaka svoj informacijski sistem, je za izgradnjo podatkovnega skladišča potrebno najprej izločiti ustrezne podatke iz informacijskih sistemov posameznih članic, jih prečistiti in preoblikovati ter nato naložiti v posamezna področna podatkovna skladišča. Arhitektura predlaganega podatkovnega skladišča je prikazana na sliki 1.

Za realizacijo smo izbrali strategijo "od spodaj navzgor" [16, 9, 15], ki omogoča postopno izgradnjo posameznih področnih podatkovnih skladišč, ki jih nato povežemo v celovito rešitev s pomočjo med seboj skladnih skupnih dimenzijskih tabel (angl. conformed dimensions). V primerjavi s strategijo "od zgoraj navzdol", ki jo zagovarja Inmon [7], nam ta pristop omogoča, da hitreje, z manj stroški in manj tveganja pridemo do otipljivih rezultatov. Seveda pa je treba že od začetka posvetiti posebno pozornost načrtovanju skupnih dimenzijskih tabel [10].

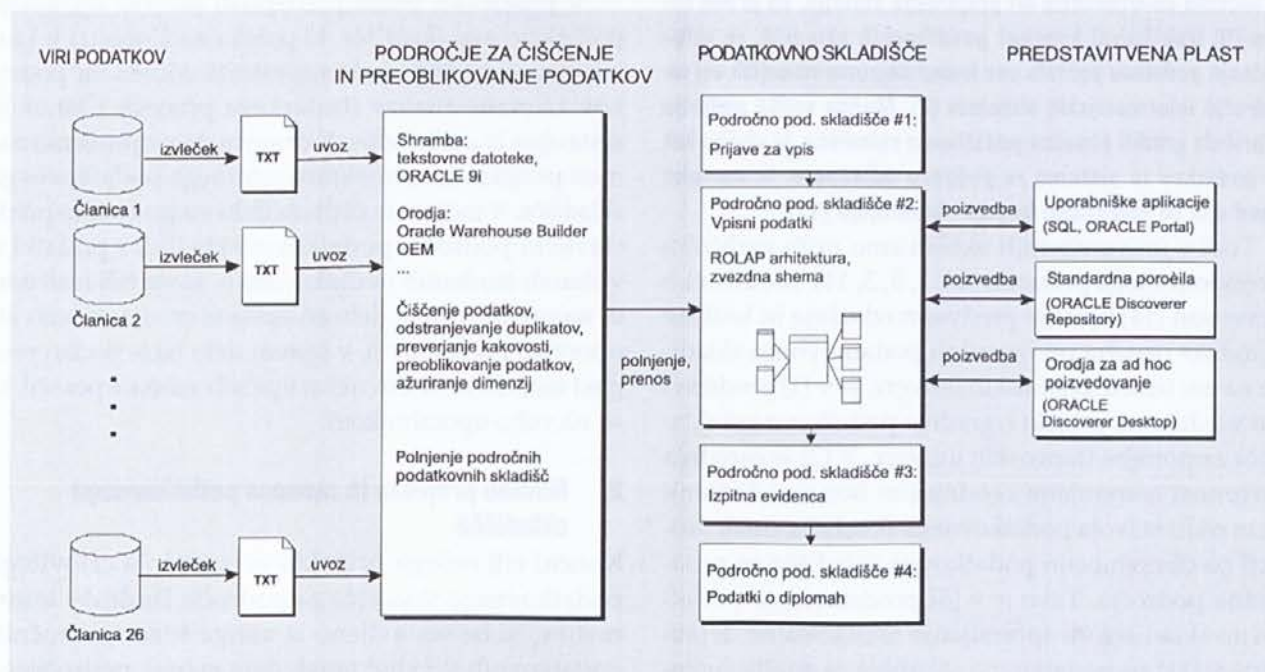
Izgradnjo celotnega podatkovnega skladišča smo tako zasnovali kot iterativen proces, v okviru katerega je moč priti do končne rešitve s pomočjo manjših 4-8-mesečnih projektov, ki dajejo otipljive rezultate

takoj in (vsaj v začetni fazi) temeljijo na dobro definiranih poslovnih procesih s stabilnim virom podatkov [9]. V skladu s temi priporočili smo najprej realizirali področno podatkovno skladišče za analizo vpisnih podatkov. To področje smo izbrali iz dveh razlogov:

- Obstaja enoten vir podatkov – obrazec "Vpisni list", ki je za vse članice enak, čeprav ga posamezne članice obdelujejo na različne načine in z različnimi programi.
- Na ravni univerze je bila jasno izražena potreba po pregledu nad stanjem vpisa na posameznih članicah.

Na ta način smo dosegli, da je bil projekt na začetku osredotočen na en sam poslovni proces z dobro definiranim virom podatkov, ki lahko pomembno prispevajo k boljšemu poslovanju univerze.

Nato smo realizirali še področno podatkovno skladišče za analizo podatkov o diplomantih, medtem ko je realizacija preostalih področnih podatkovnih skladišč (za analizo prijav za vpis in opravljenih študijskih obveznosti) predvidena v nadaljevanju projekta. Ocenjujemo, da realizacija področnega podatkovnega skladišča s podatki o prijavah za vpis ne bi smela predstavljati velikih težav, saj gre za poslovni proces, ki je strogo centraliziran in enoten za celotno univerzo. Veliko več problemov pa pričakujemo pri realizaciji področnega podatkovnega skladišča za spremljanje



Slika 1: Arhitektura podatkovnega skladišča

opravljanja študijskih obveznosti, saj bodo tu prišle do izraza vse specifičnosti posameznih članic.

Za realizacijo podatkovnega skladišča smo izbrali orodja podjetja Oracle. Podatke iz informacijskih sistemov posameznih članic dobimo v obliki tekstovnih datotek, ki jih očistimo, preoblikujemo in naložimo v podatkovno skladišče s pomočjo orodja Oracle Warehouse Builder [12]. Za dostop do podatkov uporabljamo Oracle Discoverer [13], zahtevnejša poročila pa smo sprogramirali v programskem jeziku PL/SQL.

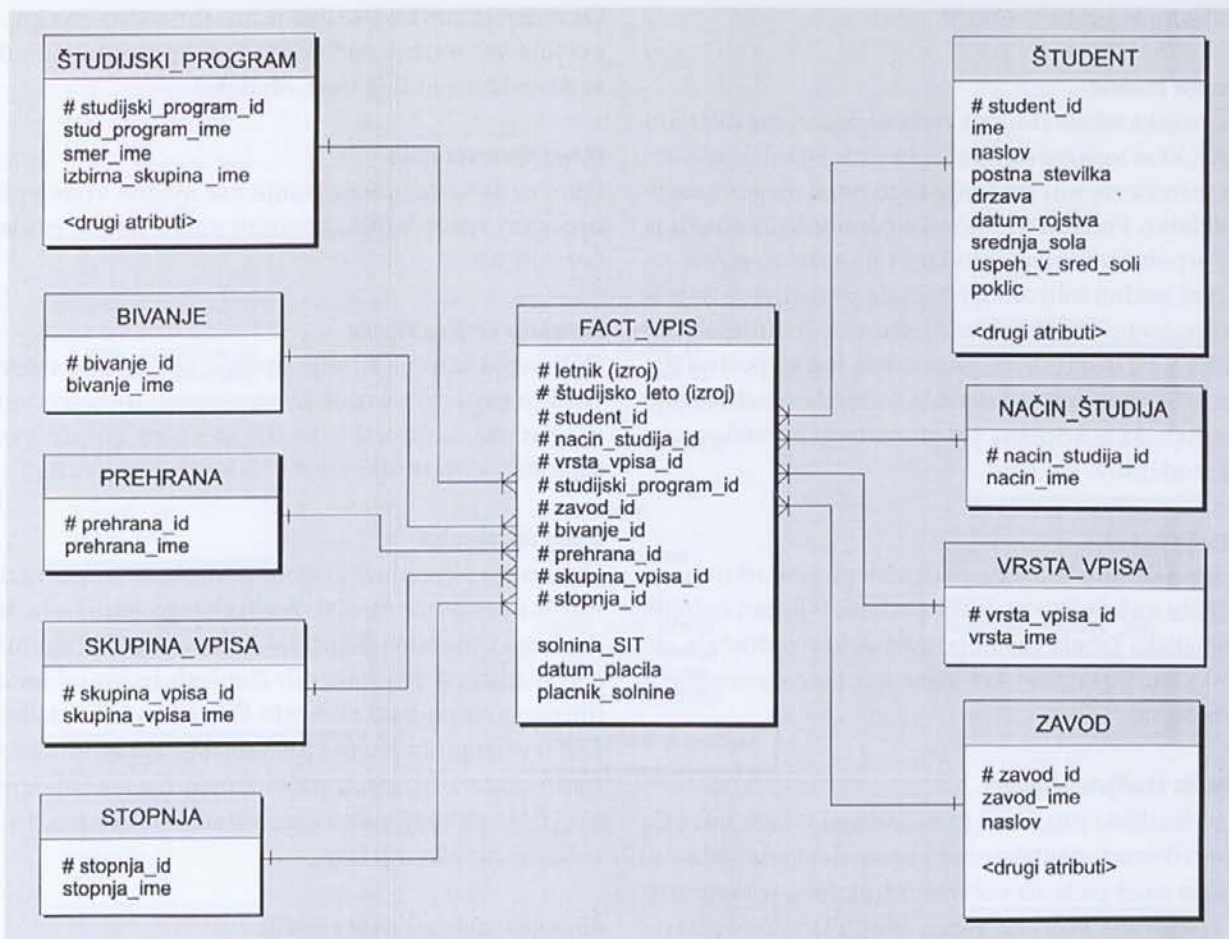
### 3 Področno podatkovno skladišče za analizo vpisnih podatkov

Logični načrt področnega podatkovnega skladišča za analizo vpisnih podatkov prikazuje zvezdna shema na sliki 2. Sestavljajo jo tabela dejstev in 11 dimenzijskih tabel. Za zrno tabele dejstev smo izbrali vpis enega študenta, ki je prikazan kot dogodek na prese-

čišču naslednjih dimenzij: študent, čas (tj. študijsko leto), zavod (tj. fakulteta, akademija ali visoka šola, na katero se je študent vpisal), študijski program, letnik, način študija (redni ali izredni), stopnja študija (visokošolski strokovni, univerzitetni, magistrski, doktorski), skupina (če so študenti v okviru istega letnika in študijskega programa razdeljeni na več skupin), vrsta vpisa, prehrana v času študija in bivanje v času študija. Dimenziji čas in letnik nimata atributov, zato sta predstavljeni kot izrojene dimenziji (angl. degenerate dimensions). Zanju obstaja samo ključ v tabeli dejstev, ne pripada pa jima samostojna dimenzijska tabela.

#### Tabela dejstev

Tabela dejstev predstavlja množico razmerij M:N med prej omenjenimi dimenzijami, vsebuje pa samo dve merljivi dejstvi: znesek šolnine in datum, ko je bila šol-



Slika 2: Zvezdna shema področnega podatkovnega skladišča za analizo podatkov o vpisu

nina plačana. To pomeni, da večina aplikacij, ki uporabljajo to področno podatkovno skladišče, izvaja samo štetje zapisov. Kljub majhnemu številu merljivih dejstev (tabele dejstev, ki ponazarjajo dogodke, običajno nimajo merljivih dejstev) pa lahko s pomočjo te tabele dobimo odgovore na številna vprašanja, npr.:

- Koliko študentov se je vpisalo na posamezno članico ali v posamezen študijski program?
- Kakšna je struktura vpisanih študentov glede na končano srednjo šolo, poklic (ki ga imajo po končani srednji šoli), uspeh v srednji šoli, način študija, vrsto vpisa ipd.?
- Se število študentov na posameznih članicah oziroma študijskih programih povečuje ali zmanjšuje?
- Kako napreduje v višje letnike izbrana generacija študentov?
- Iz katerih krajev prihajajo študentje, ki so vpisani na posamezno članico ali študijski program?
- Kje študentje prebivajo in se hranijo v času študija (v študentskem domu, pri starših, v podnajemniškem stanovanju itd.)?

#### **Dimenzija študent**

Dimenzijska tabela študent vsebuje podatke o vseh študentih, ki so vpisani na Univerzo v Ljubljani. Vsaka vrstica ustreza enemu študentu in vsebuje njegove osebne podatke. Podatki o končani srednji šoli (ime srednje šole, uspeh v zadnjem letniku in na maturi, poklic po končani srednji šoli) omogočajo, da primerjamo uspešnost študentov na univerzi z njihovim srednješolskim profilom. Po drugi strani pa podatki, kot so poštna številka, občina stalnega bivališča in regija predstavljajo hierarhijo, ki je koristna pri analizah geografskega izvora študentov.

#### **Dimenzija čas**

Gledano samo s stališča področnega podatkovnega skladišča za analizo vpisnih podatkov je samostojna dimenzijska tabela čas nepotrebna, ker zadostuje, da imamo študijsko leto kot ključ izrojene dimenzije v tabeli dejstev.

#### **Dimenzija študijski program**

Ker so študijski programi navadno sestavljeni tako, da se (praviloma v višjih letnikih) lahko delijo na več smeri, vsaka smer pa še na več izbirnih skupin, je ta dimenzija zasnovana tako, da vsaka vrstica tabele predstavlja eno izbirno skupino. Več izbirnih skupin, ki pripadajo isti smeri, lahko povežemo med seboj na nivoju

smeri, več smeri, ki pripadajo istemu študijskemu programu, pa na nivoju študijskega programa. Opisana hierarhija študijski program – smer – izbirna skupina omogoča enostavno izdelavo poročil na različnih ravneh podrobnosti, ki je v terminologiji podatkovnih skladišč znana pod imenom vrtanje navzdol (angl. drill down) oziroma vrtanje navzgor (angl. drill up).

#### **Dimenzija zavod**

Dimenzija zavod opisuje posamezne članice ljubljanske univerze. Trenutno univerzo sestavlja 26 članic.

#### **Dimenzija letnik**

Dimenzija letnik je izrojena dimenzija, saj je atribut letnik študija, ki je sestavni del primarnega ključa v tabeli dejstev, edini atribut te dimenzije. Uporablja se kot ključ za grupiranje tistih študentov, ki so vpisani v isti letnik.

#### **Dimenzija način študija**

Dimenzija način študija je majhna dimenzija, ki opisuje vse možne načine študija, npr. redni, izredni (v preteklosti tudi iz dela, ob delu).

#### **Dimenzija vrsta vpisa**

Dimenzija vrsta vpisa opisuje vse možne vrste vpisa, npr. prvi vpis v letnik, ponovni vpis v letnik, evidenčni vpis itd.

#### **Dimenzija stopnja študija**

Dimenzija stopnja študija opisuje vse možne stopnje študija (npr. visokošolski strokovni, univerzitetni, magistrski, doktorski študij) in s tem določa vrsto diplome, ki jo študent dobi po končanem študiju.

#### **Dimenzija skupina**

Dimenzija skupina omogoča podrobnejši opis različnih skupin študentov v okviru istega letnika in študijskega programa. Skupine so odvisne od organizacije študija na posameznih članicah in jih lahko definiramo na podlagi različnih kriterijev, npr. glede na kraj izvajanja študija (v Ljubljani ali v različnih dislociranih oddelkih), glede na vrsto študija (redni, izredni), glede na predhodni potek študija (običajno študij, nadaljevalni študij) itd.

#### **Dimenzija prehrana v času študija**

Dimenzija prehrana v času študija opisuje različne možnosti, ki so na voljo študentom za prehranjevanje

v času študija, npr. v študentskem domu, pri starših, v lastnem gospodinjstvu itd.

#### Dimenzija bivanje v času študija

Dimenzija bivanja v času študija opisuje različne vrste nastanitve v času študija, npr. v študentskem domu, pri starših, v podnajemniškem stanovanju itd.

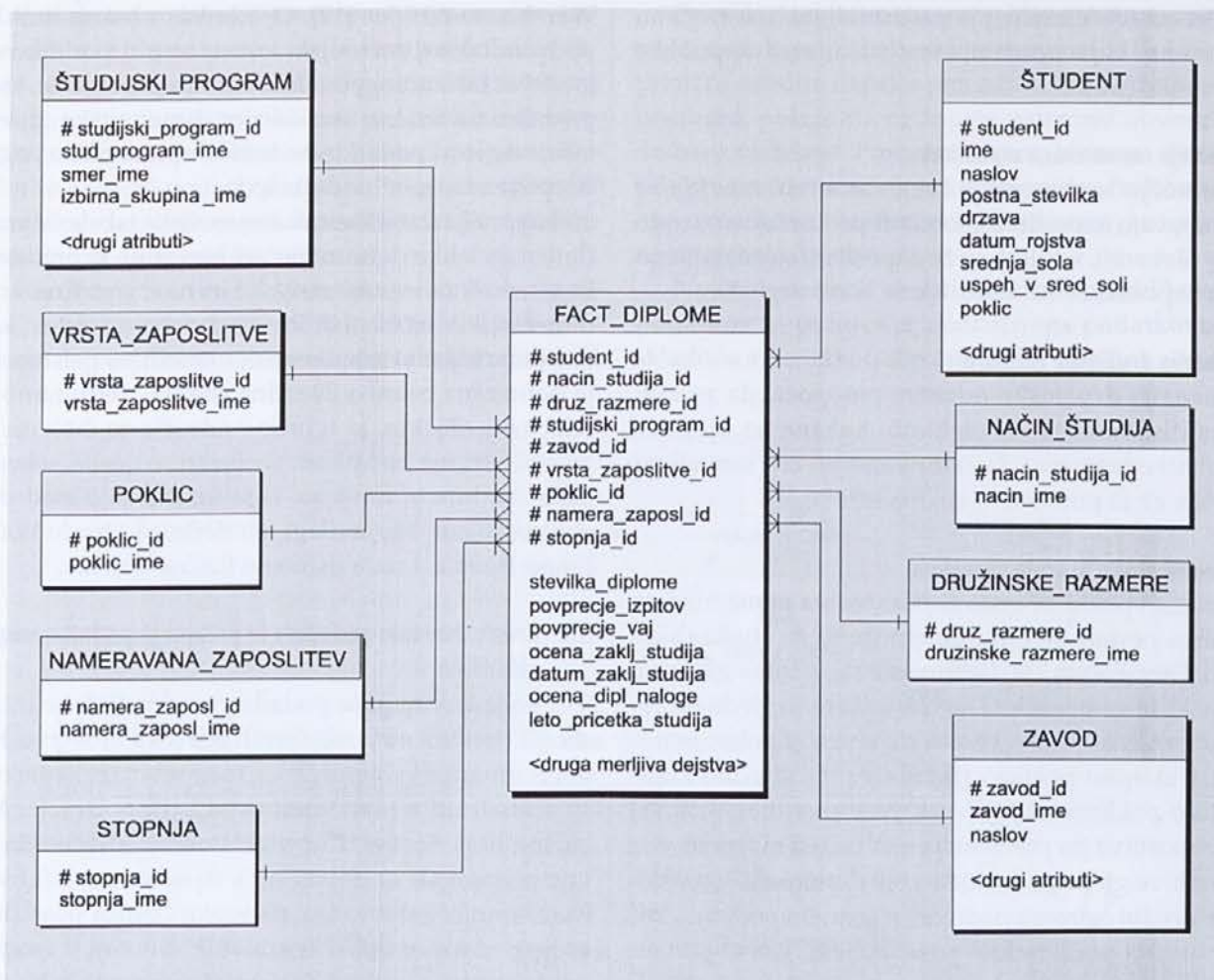
### 4 Področno podatkovno skladišče za analizo podatkov o diplomantih

Logični načrt področnega podatkovnega skladišča za analizo podatkov o diplomantih je prikazan na sliki 3. Sestavljajo ga tabela dejstev in 9 dimenzijskih tabel. Zrno tabele dejstev predstavlja diploma enega študenta, ki je prikazana kot dogodek na presečišču devetih dimenzij. Tabela dejstev vsebuje naslednja merljiva dejstva: številko diplome, datum zagovora, oceno

diplomske naloge, oceno zagovora, povprečno oceno izpitov in vaj, leto pričetka študija itd. Dimenzijske tabele pa predstavljajo dimenzije študent, zavod, študijski program, način študija, stopnja študija, poklic, vrsta zaposlitve, nameravana zaposlitev in družinske razmere.

S pomočjo tega področnega podatkovnega skladišča lahko dobimo odgovore na različna vprašanja, povezana s podatki o diplomantih, npr.:

- Koliko študentov je diplomiralo na posameznih članicah univerze?
- Kolikšno je število diplomantov na posameznih študijskih programih, smereh in izbirnih skupinah?
- Kolikšen je povprečen čas študija pri posameznih generacijah študentov (na posameznih članicah oziroma študijskih programih)?



Slika 3: Zvezdna shema področnega podatkovnega skladišča za analizo podatkov o diplomantih

- Kako uspešni so izredni študenti v primerjavi z rednimi?
- Kakšna je povprečna ocena študentov na različnih članicah, študijskih programih, smereh in izbirnih skupinah?
- Kje se nameravajo diplomanti zaposliti?
- Kakšne so družinske razmere diplomantov?

Dimenzijske tabele študent, zavod, študijski program, način študija in stopnja študija so skupne za obe področni podatkovni skladišči, zato so fizično realizirane samo enkrat in delujejo kot "lepilo", s katerim sta obe področni podatkovni skladišči povezani med seboj. Z njihovo pomočjo lahko izvajamo t. i. vrtnanje počez (angl. drill across) in tako generiramo poročila, ki vključujejo podatke iz obeh zvezdnih shem.

Poleg že omenjenih skupnih dimenzij nastopajo v področnem podatkovnem skladišču s podatki o diplomantih še štiri nove dimenzije: nameravana zaposlitev, družinske razmere, poklic in vrsta zaposlitve. Vsaka od teh dimenzij je predstavljena z ustreznim šifrantom, ki ga uporablja Statistični urad Republike Skovenije (SURS).

#### **Dimenzija nameravana zaposlitev**

S pomočjo te dimenzije lahko analiziramo, kje se nameravajo zaposliti diplomanti po končanem študiju: v Sloveniji, v tujini, so že zaposleni, se ne nameravajo zaposliti, še ne vedo, kje se bodo zaposlili.

#### **Dimenzija družinske razmere**

Dimenzija družinske razmere omogoča, da za vsakega diplomanta zabeležimo, kakšne so njegove družinske razmere v času diplome: ali živi sam ali pri starših, ali je poročen in ali ima otroke.

#### **Dimenziji poklic in vrsta zaposlitve**

Dimenziji poklic in vrsta zaposlitve sta pomembni za analizo podatkov o diplomantih, ki so študirali ob delu. Omogočata, da diplomante razvrstimo glede na delo, ki so ga opravljali med študijem in glede na vrsto zaposlitve. Vsaka vrstica v dimenziji poklic ustreza eni skupini poklicev, skladno s standardno klasifikacijo poklicev SURS. Vsaka vrstica v dimenziji vrsta zaposlitve pa predstavlja eno izmed mogočih vrst zaposlitve glede na to, kje so bili diplomanti zaposleni: v družbi oziroma podjetju, v lastnem podjetju, kot samostojni podjetniki – posamezniki, v svobodnem poklicu itd.

## **5 Opis orodij za realizacijo**

V splošnem velja, da za realizacijo podatkovnega skladišča potrebujemo sistem za upravljanje podatkovne baze, orodja za izločanje, preoblikovanje in nalaganje podatkov (angl. ETL tools) in orodja za predstavitev podatkov končnim uporabnikom (angl. data access tools) [9].

### **5.1 Sistem za upravljanje podatkovne baze in fizični načrt**

Pri izboru sistema za upravljanje podatkovne baze se navadno odločimo za tisti sistem, ki je predpisan kot interni standard organizacije, za katero razvijamo podatkovno skladišče. Vsako odstopanje namreč zahteva tehtno utemeljitev. Zato smo izbrali Oracle DBMS, ki je predlagan kot interni standard za razvoj vseh novih aplikacij na ljubljanski univerzi.

Fizični načrt podatkovnega skladišča in polnjenje skladišča s podatki smo realizirali z orodjem Oracle Warehouse Builder [12]. Oracle Warehouse Builder podpira tako dimenzijski kot relacijski podatkovni model in tako omogoča, da kreiramo vse objekte, ki so potrebni za fizično realizacijo: dimenzijske tabele, tabele dejstev, podatkovne kocke, uporabniške poglede, povezave med podatki ipd.

Najprej je treba kreirati dimenzijske tabele. V vsaki dimenziji lahko definiramo več hierarhij, ki predstavljajo temelj za vrtnanje navzdol in navzgor. Kreiranju dimenzijskih tabel sledi kreiranje tabele dejstev, kjer določimo ključne vse dimenzij in merljiva dejstva. Ko je definirana celotna zvezdna shema, definiramo še relacijske objekte, ki tvorijo področje za čiščenje in preoblikovanje podatkov. Ti objekti so tabele, sekvence, procedure in funkcije, ki se uporabljajo med preoblikovanjem. Na podlagi teh definicij Oracle Warehouse Builder kreira ustrezne fizične objekte.

### **5.2 Preoblikovanje podatkov in polnjenje podatkovnega skladišča**

Viri podatkov za naše podatkovno skladišče so transakcijski sistemi na posameznih članicah univerze. Ker se informacijski sistemi članic med seboj razlikujejo in so realizirani z različnimi orodji (npr. Oracle, Microsoft SQL Server, Clipper), smo se odločili, da ne bomo posegali neposredno v njihove podatkovne baze, ampak zahtevamo, da vsaka članica posreduje svoje podatke v obliki tekstovnih datotek v vnaprej predpisanem formatu. Te datoteke zbiramo in z njimi

polnimo podatkovno skladišče v vnaprej predpisanih časovnih intervalih. Za področno podatkovno skladišče z vpisnimi podatki so ti intervali bolj gosti v času vpisa (tj. septembra in v začetku oktobra), medtem ko se med študijskim letom vpisni podatki praktično ne spreminjajo. Področno podatkovno skladišče s podatki o diplomantih pa osvežujemo vsak mesec.

Odločitvi za uporabo tekstovnih datotek sta botrovala še dva razloga: tekstovne datoteke so preproste in se jih da enostavno kreirati, Oracle pa ima na voljo učinkovita orodja za njihovo nalaganje v relacijske tabele.

Med postopkom preoblikovanja podatkov se tekstovne datoteke s podatki posameznih članic najprej naložijo v relacijske tabele s pomočjo orodja Oracle SQL\*Loader. Pri tem SQL Loader preveri format podatkov in prisotnost vseh obveznih atributov. Nepopolne vrstice in vrstice z neustreznim formatom se izločijo in zabeležijo v posebnem dnevniku obdelave, ki služi kot osnova za odpravo napak.

V naslednjem koraku preverimo veljavnost podatkov (ali so njihove vrednosti znotraj predpisane domene in ustrezajo dogovorjenim poslovnim pravilom) in referenčno integriteto. Nazadnje opravimo še kontrolo podvajanja in izločimo duplikate. Tudi v tem koraku se napačni podatki skupaj z ustreznimi sporočili o napakah zapišejo na posebno datoteko, ki se potem uporabi pri njihovem čiščenju.

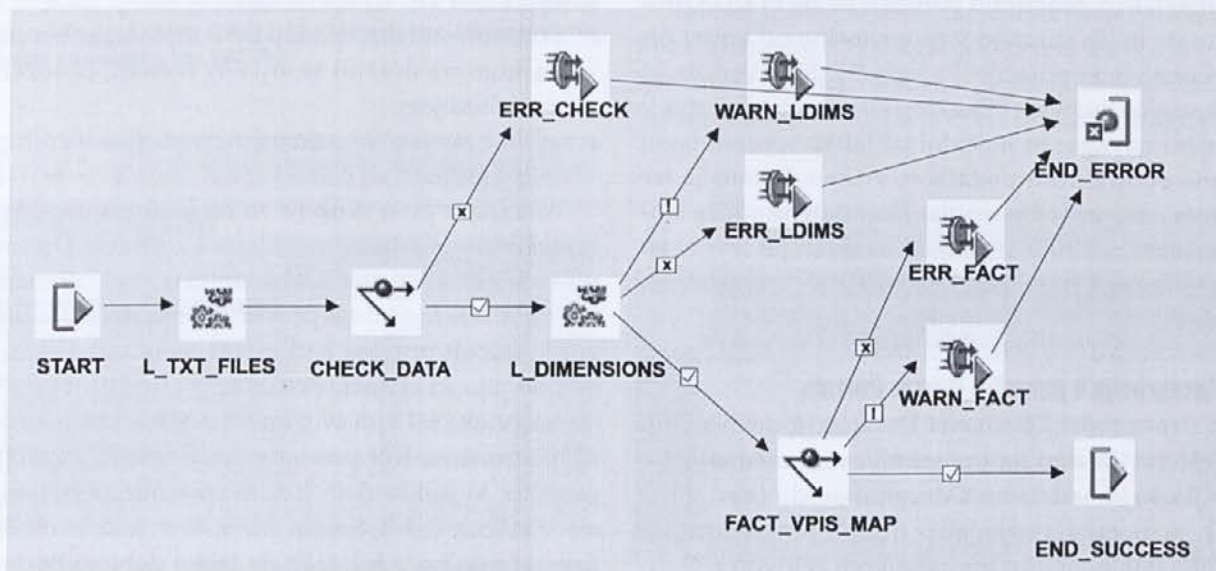
Podatki, ki so uspešno prestali vse kontrole, se na koncu naložijo v podatkovno skladišče. Med nalaganjem se najprej osvežijo dimenzijske tabele, nato pa še

tabela dejstev. V primeru sprememb v dimenzijskih tabelah nove vrednosti "povezijo" stare, kar ustreza tipu 1 počasi spreminjajočih se dimenzij [8]. Do takih sprememb največkrat pride v dimenziji študent, medtem ko so ostale dimenzije relativno majhne in se le redko spreminjajo.

Celoten postopek preoblikovanja in nalaganja podatkov v področno podatkovno skladišče s podatki o vpisu je prikazan na sliki 4. Proces `l_txt_files` sproži nalaganje tekstovnih datotek v relacijske tabele. Preslikava `check_data` preveri veljavnost podatkov, referenčno integriteto in morebitna podvajanja. V primeru napake se sproži procedura `err_check`, ki v poseben dnevnik obdelave zapiše vrsto napake in podatke, pri katerih je bila napaka ugotovljena. Proces `l_dimensions` sproži osveževanje dimenzij. Sestavljen je iz več podprocesov (angl. child processes) in preslikav (angl. mappings), ki osvežujejo posamezne dimenzije. Preslikava `fact_vpis_map` na koncu osveži tabelo dejstev.

Proceduri `err_ldims` in `warn_ldims` služita za obravnavo napak pri osveževanju dimenzijskih tabel. V primeru usodne napake procedura `err_ldims` prekine postopek polnjenja in zapiše ustrezno obvestilo v dnevnik obdelave. Procedura `warn_ldims` pa samo zapiše opozorilo in nadaljuje s polnjenjem. Podobno nalogo opravljata proceduri `err_fact` in `warn_fact` pri polnjenju tabele dejstev.

Na enak način je realizirano tudi preoblikovanje podatkov in polnjenje področnega podatkovnega skladišča s podatki o diplomantih.



Slika 4: Postopek preoblikovanja in nalaganja podatkov

### 5.3 Dostop do podatkov

Podatkovno skladišče samo po sebi ni koristno, če ne moremo na primeren način prikazati podatkov, ki so shranjeni v njem. V našem primeru so poročila, ki jih potrebujejo končni uporabniki, realizirana na tri načine:

- kot vnaprej definirane uporabniške aplikacije, napisane v jeziku PL/SQL;
- kot vnaprej definirana poročila, zgrajena z orodjem Oracle Discoverer in shranjena v repozitoriju;
- kot ad hoc poizvedbe, ki jih po potrebi generiramo z Oracle Discovererjem.

Prvo možnost uporabljamo za zahtevna poizvedovanja, ki jih je težko realizirati z običajnimi orodji za povpraševanje in generiranje poročil. Druga možnost je primerna za večino standardnih poročil, medtem ko se tretje možnosti poslužujemo v primeru, ko se zahteve za neko poročilo pojavijo spontano.

Orodje za dostop do podatkov Oracle Discoverer [13] sestavlja dve komponenti: Discoverer Administrator in Discoverer Desktop. Discoverer Administrator je namenjen računalniškim strokovnjakom in se uporablja za vrsto nalog, kot so npr. vzdrževanje meta podatkov (tj. določanje dimenzij in kock), nadzor nad dostopom in vzdrževanje agregatov. Vsi metapodatki so shranjeni v repozitoriju, imenovanem Discoverer End User Layer, ki predstavlja jedro sistema za dostop do podatkov v podatkovnem skladišču.

Končni uporabniki in analitiki uporabljajo Discoverer Desktop. Le-ta dostopa do metapodatkov repozitoriju, ki jih je pripravil Discoverer Administrator. Analitiki lahko generirajo poročila, ne da bi znali programirati, in jih shranijo v repozitoriju, od koder jih lahko ponovno uporabijo, kadarkoli želijo. Poročila, ki jih generira Discoverer Desktop so parametrizirana, z vrtnjem navzgor in navzdol pa lahko spreminjamo stopnjo podrobnosti podatkov, vrtimo dimenzije ter spreminjamo izgled poročila. Rezultate poročila lahko predstavimo tudi v grafični obliki in jih izvozimo v druga orodja (npr. Excel, glej sliko 9) za nadaljnjo obdelavo.

### 5.4 Integracija s pomočjo Oracle Portala

Slaba stran orodja Discoverer Desktop je, da mora biti nameščeno lokalno na uporabnikovem računalniku, poročila, ki jih izdelamo z drugimi orodji (npr. v PL/SQL), pa moramo poganjati iz drugih programov. Da bi lahko dali končnim uporabnikom celovito rešitev, ki bi omogočala dostop do podatkovnega skladišča

prek ene same vstopne točke, smo se odločili za integracijo v okolju Oracle 9iAS Portal [4].

V okviru portala smo izdelali aplikacijo, ki omogoča dostop do podatkovnega skladišča prek svetovnega spleta in povezuje v enovito rešitev vsa poročila ne glede na to, s katerim orodjem so bila izdelana. Vsa poročila (vnaprej sprogramirana v PL/SQL in avtomatično generirana z orodjem Discoverer) so dostopna prek enotnega uporabniškega vmesnika, tako da uporabniku ni treba vedeti, kakšna je njihova dejanska realizacija.

Oracle 9iAS Discoverer uporablja iste metapodatke kot Discoverer Desktop, zato lahko do poročil, ki so shranjena v repozitoriju, dostopamo prek svetovnega spleta, ne da bi jih spreminjali. Analitiki in programerji lahko pripravljajo poročila in jih shranjujejo v repozitoriju s pomočjo ustrezne okenske aplikacije, medtem ko jih končni uporabniki lahko uporabljajo prek svetovnega spleta. Zahtevna poročila pa lahko realiziramo kot dinamične strani v PL/SQL.

## 6 Pregled analiz in tipičnih poročil za končne uporabnike

V tem delu podajamo pregled najpomembnejših analiz, ki jih omogoča naše podatkovno skladišče. Zaradi večje preglednosti smo analize grupirali v 4 skupine:

- analize, ki omogočajo pregled nad številom vpisanih študentov in številom diplomantov na posameznih članicah,
- podrobne analize vpisnih podatkov s poudarkom na analizi prehodnosti iz letnika v letnik in trendih vpisa,
- podrobne analize podatkov o diplomantih s poudarkom na dolžini študija in ocenah, doseženih med študijem,
- ad hoc analize, ki nastajajo zaradi specifičnih potreb posameznih članic.

Analize iz prve skupine so razmeroma enostavne in se jih da učinkovito realizirati z Oracle Discovererjem, vendar so za vodstvo univerze zelo pomembne, saj je šele izgradnja podatkovnega skladišča omogočila celovit pregled nad vsemi članicami. Glede na avtonomijo, ki jo imajo posamezne članice, in dejstvo, da ima vsaka od njih svoj informacijski sistem, to prej ni bilo možno. Kot primer te vrste analiz navajamo poročila, ki prikazujejo število vpisanih na posameznih članicah v določenem študijskem letu in omogočajo vrtnje navzdol, tako da lahko dobimo podatke o vpisu na različnih stopnjah podrobnosti:



- za članico kot celoto,
- po posameznih študijskih programih, ki jih izvaja določena članica,
- po smereh znotraj študijskega programa,
- po izbirnih skupinah znotraj vsake smeri,
- po letnikih in študijskih programih.

Primer vrtnja navzdol prikazuje slike 5, 6 in 7. Medtem ko so na sliki 5 prikazani samo sumarni podatki o številu vpisanih na posameznih članicah, je iz slike 6 razvidno število vpisanih po posameznih študijskih programih, iz slike 7 pa tudi število vpisanih na posameznih smereh študija.

Število vpisanih študentov po članicah			
Page Item: Študijsko leto: 2003 ▼			
	Število vpisanih		
	redni študij	izredni študij	Skupaj
Akademija za glasbo Ljubljana	402	15	417
Akademija za gled., radio, film in TV	87		87
Akademija za likovno umetnost Lj.	395	16	411
Biotehniška fakulteta	3.155	264	3.419
Ekonomski fakulteta Ljubljana	5.693	3.167	8.860
Fakul. za gradbeništvo in geodezijo	1.558	249	1.807
Fakulteta za arhitekturo	847	52	899
Fakulteta za družbene vede Lj.	3.651	784	4.435
Fakulteta za elektrotehniko	2.125	59	2.184
Fakulteta za farmacijo	1.040	32	1.072
Fakulteta za kemijo in kem. tehn.	1.369	132	1.501
Fakulteta za matematiko in fiziko	963	57	1.020
Fakulteta za pomorstvo in promet	1.277	784	2.061
Fakulteta za računalništvo in informatiko	1.404	49	1.453
Fakulteta za socialno delo	607	528	1.135
Fakulteta za strojništvo Ljubljana	1.754	73	1.827
Fakulteta za šport Ljubljana	945	123	1.068
Fakulteta za upravo	1.267	2.113	3.380

Slika 5: Število vpisanih po posameznih članicah. Prvi stolpec prikazuje število rednih, drugi stolpec število izrednih in tretji stolpec skupno število študentov v študijskem letu 2003/04.

Število vpisanih po članicah in študijskih programih			
Page Item: Študijsko leto: 2003 ▼			
	Št. vpisanih		
	redni študij	izredni študij	Skupaj
Fakulteta za elektrotehniko	2.125	59	2.184
▶ ELEKTROTEHNIKA UN	1.102		1.102
▶ ELEKTROTEHNIKA VS	1.023	59	1.082
Fakulteta za računalništvo in informatiko	1.404	49	1.453
▶ RAČUNAL. IN INFORMATIKA UN	808	24	832
▶ RAČUNAL. IN INFORMATIKA VS	596	25	621
Medicinska fakulteta Ljubljana	1.685		1.685
▶ MEDICINA - VIS	1.321		1.321
▶ STOMATOLOGIJA - VIS	364		364
Skupaj vsi	5.214	108	5.322

Slika 6: Primer vrtnja navzdol: število vpisanih v posamezne študijske programe na izbranih članicah

Število vpisanih po članicah in študijskih programih			
Page Item: Študijsko leto: 2003 ▼			
	Št. vpisanih		
	redni študij	izredni študij	Skupaj
Fakulteta za elektrotehniko	2.125	59	2.184
▶ ELEKTROTEHNIKA UN	1.102		1.102
▶ Avtomatika	194		194
▶ Elektronika	124		124
▶ Močnostna elektrotehnika	106		106
▶ NI smeri	536		536
▶ Telekomunikacije	142		142
▶ ELEKTROTEHNIKA VS	1.023	59	1.082
▶ Avtomatika	150	10	160
▶ Elektronika	133	26	159
▶ Energetska teh. in avt. post.	145	11	156
▶ NI smeri	394		394
▶ Telekomunikacije	137	12	149
▶ Zagotavljanje kakovosti	64		64
Fakulteta za računalništvo in informatiko	1.404	49	1.453
▶ RAČUNAL. IN INFORMATIKA UN	808	24	832

Slika 7: Nadaljevanje vrtnja navzdol: število vpisanih po posameznih smereh študija

V to skupino sodijo tudi posebne analize števila vpisanih na Filozofsko fakulteto, za katero je značilen dvopredmetni študij, ki omogoča, da študenti izberejo poljubno kombinacijo dveh študijskih programov. Te analize omogočajo pregled nad vsemi izbranimi kombinacijami in vrtnje navzdol po obeh študijskih programih.

Analize iz druge skupine skušajo dati odgovore na dve vrsti vprašanj:

- Kako določena generacija študentov, ki so vpisani v neki študijski program, napreduje iz letnika v letnik?
- Kako se spreminjajo trendi vpisa oziroma zanimanje študentov za študij na posameznih fakultetah?

Primer analize, ki daje odgovor na prvo vprašanje, je prikazan na sliki 8. Iz slike je jasno razviden tipičen problem univerzitetnega študija v Sloveniji, to je izredno velik osip študentov na prehodu iz prvega v drugi letnik.

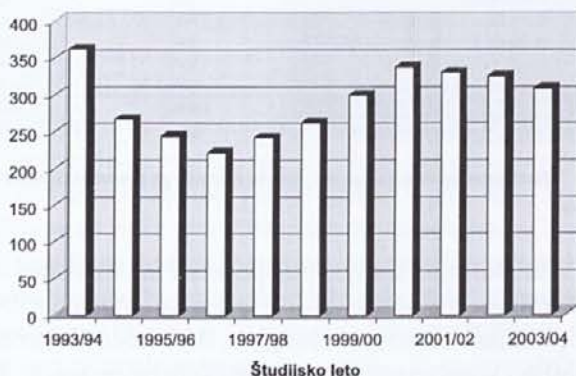
Spremljanje napredovanja generacije							
Page Item: Zvezi: Fakulteta za računalništvo in informatiko ▼ VIP: RAČUNAL. IN INFORMATIKA UN ▼ Leto							
	Število vpisanih						
	1. letnik	2. letnik	3. letnik	4. letnik	5. letnik	Absolvent UNB	Skupaj
1995/96	163						163
1996/97	25	73					98
1997/98	6	34	55				95
1998/99		3	24	51			78
1999/00		1	2	22	51		76
2000/01			1	3	22	41	67
2001/02			1		2	21	24
2002/03						2	2
2003/04				2			2
Skupaj	194	111	83	78	75	64	605

Slika 8: Napredovanje izbrane generacije študentov

Primer analize, ki daje odgovor na drugo vprašanje, pa je na sliki 9. Slika jasno kaže na upad zanimanja za študij tehnike po osamosvojitvi Slovenije in spremembi družbenega sistema. Šele v zadnjih letih se je situacija nekoliko popravila.

V tretjo skupino spadajo analize, ki prikazujejo podatke o dolžini študija in uspešnosti posameznih generacij diplomantov. Primer take analize je prikazan na sliki 10.

**Število vpisanih na študijski program ELEKTROTEHNIKA**



Slika 9: Število vpisanih na univerzitetni študij elektrotehnike v letih 1993–2003

Univerza v Ljubljani

Čas trajanja študija

Page Item: Leto dipliranja: 2003 Vrsta študija: visokoš. (univ.) št.

Zavod	Študijski program	Trajanje študija
Akademija za glasbo i ljubljana	CERKVENA GLASBA - VIS	6,75
	GLASBENA PEDAGOGIKA -VIS	7,50
	GOGALA IN DR INSTR. S STR -VIS	6,65
	INSTRUMENTI S TIFKAMI - VIS	6,46
	KOMPOZICJA IN GLASB.TEOR.-VIS	6,06
	PETJE - VIS	4,25
Akademija za gled.,radio,film in TV	PIHALA, TROBILA IN TOLKALA-VIS	7,15
	DRAMATURGIJA	6,41
	DRAMSKA IGRA IN UMETNJSKA BEG	7,32
	FILMSKA IN TELEVIZJSKA REŽIJA	7,75
	GLEDALJSKA IN RADJSKA REŽIJA	7,75
		Povprečje: 6,12
		Povprečje: 7,31

Slika 10: Trajanje študija v letih za posamezne študijske programe na izbranih članicah

Zadnjo skupino tvorijo analize, ki nastajajo spontano kot rezultat zahtev, ki se pojavljajo med vsakodnevnim poslovanjem. Te analize skušamo realizirati takoj, ko se pojavi tovrstna potreba, z uporabo Oracle Discovererja. Primer take analize prikazuje slika 11. Analiza je nastala kot posledica zahteve Fakultete za računalništvo in informatiko, ki je želela podrobneje analizirati strukturo vpisanih v prvem letniku visokošolskega strokovnega študija glede na poklic, ki ga

Univerza v Ljubljani

Število vpisanih študentov glede na srednješolski poklic

Page Item: Zavod: Fakulteta za računalništvo in informatiko Študijsko leto: 2003/04 VIP: RACUNA

	Št. vpisanih		
	redni študij	izredni študij	Skupaj
5.70.06 EKONOMSKO-KOMERC.TEHNİKAD 1991	35	3	38
5.26.01 RAČUNALNIŠKI TEHNİK	22	3	25
5.25.38 ELEKTROTEHNİK ELEKTRONİK	20	4	24
5.99.11 GIMNAZIJSKI MATURANT	20	7	27
5.11.01 LESARSKI TEHNİK	8		8
5.73.15 POSLOVNI TEHNİK (3+2)	7		7
5.20.01 STROJNI TEHNİK	6	2	8
5.70.08 EKONOMSKI TEHNİK - PR	5		5
5.94.01 ZDRAVSTVENI TEHNİK	5		5
5.99.31 TEHNIŠKA GIMNAZIJA	4	1	5
5.45.01 GRAFIČNI TEHNİK	3		3

Slika 11: Struktura vpisanih v 1. letnik visokošolskega strokovnega študija računalništva glede na poklic po končani srednji šoli

imajo študenti po končani srednji šoli. Analiza je dala zanimive rezultate: največ študentov prihaja iz ekonomske srednje šole, šele nato sledijo gimnazijski maturanti in računalniški tehniki. Ta in podobne analize dajejo vodstvu fakultete možnost, da sprejme ustrezne odločitve: lahko se odloči za prilagoditev vsebine nekaterih tehničnih predmetov (npr. osnove programiranja in osnove arhitekture računalniških sistemov) predznanju študentov, ki so končali program ekonomsko-komercialni tehnik, ali pa predlaga spremembo pogojev za vpis v ta študijski program tako, da bi se zmanjšal vpis študentov, ki nimajo predhodne tehnične izobrazbe.

## 7 Sklep

Zaradi decentralizirane organizacije je izgradnja podatkovnega skladišča za Univerzo v Ljubljani še posebej pomembna, saj predstavlja podatkovno skladišče sredstvo za integracijo podatkov, ki sicer nastajajo in se obdelujejo v okviru informacijskih sistemov posameznih članic. Na ta način je vodstvu univerze omogočen vpogled v poslovanje univerze kot celote, ki pred izgradnjo podatkovnega skladišča ni bil možen.

Zaradi heterogenih virov podatkov (informacijski sistemi posameznih članic se med seboj razlikujejo) pa je izgradnja podatkovnega skladišča zahtevna in tvegana naloga. Izkušnje namreč kažejo, da napor za pripravo podatkov narašča eksponentno s številom različnih izvornih sistemov [9]. Da bi zmanjšali tveganje in čim prej prišli do otipljivih rezultatov, smo izbrali pristop "od spodaj navzgor", ki predvideva postopno izgradnjo posameznih področnih podatkovnih skladišč in njihovo povezovanje prek skupnih

dimenzijskih tabel. Najprej smo realizirali področno podatkovno skladišče z vpisnimi podatki, saj za to področje obstaja enoten in razmeroma stabilen vir podatkov – obrazec Vpisni list. Temu smo nato dodali še področno podatkovno skladišče s podatki o diplomantih. V članku je podrobno predstavljen logični podatkovni model za obe področni podatkovni skladišči in opisan postopek njune realizacije s pomočjo orodij iz družine Oracle.

Končni cilj projekta pa je izgradnja verige področnih podatkovnih skladišč, s pomočjo katerih bo moč spremljati študijski proces kot vrednostno verigo, ki se prične s prijavo za vpis na univerzo, nadaljuje z vpisom in opravljanjem izpitov v posameznih letnikih ter konča z diplomom.

## Literatura

- [1] Desnos, J.-F. A Data Warehouse for French Universities. V: J. Knop, P. Schirmbacher in V. Mahnič (ur.). *The Changing Universities – The Role of Technology*, Gesellschaft für Informatik, Bonn, 2002, str. 146–152.
- [2] Eckerson, W. W. *Evolution of Data Warehousing, The Trend Toward Analytical Applications*, Boston, MA: The Patricia Seybold Group (April 28, 1999), str. 1–8.
- [3] Cardoso, E., Galhardas, E., Trigueiros, M. J., Silva, R. A. *Decision Support System for IST Academic Information*. V: *Beyond the Network – Innovative IT Services*, Proceedings of the 9th International Conference of European University Information Systems, Amsterdam, julij 2003, str. 416–427.
- [4] Greenwald R., Milbrey J., *Oracle 9iAS Portal Bible*, John Wiley & Sons, 2001.
- [5] Flory, A., Soupirot, P., Tchounikine, A. *Design and Implementation of a Data Warehouse for Research Administration Universities*. V: J. Knop, P. Schirmbacher in V. Mahnič (ur.). *The Changing Universities – The Role of Technology*, Gesellschaft für Informatik, Bonn, 2002, str. 259–267.
- [6] Inmon, B. *Building the Data Warehouse*, QED Publishing Group, 1992.
- [7] Inmon, B. *Building the Data Warehouse, Third Edition*, John Wiley & Sons, 2002.
- [8] Kimball, R. et al. *The Data Warehouse Toolkit*, John Wiley & Sons, 1996.
- [9] Kimball, R. et al. *The Data Warehouse Lifecycle Toolkit*, John Wiley & Sons, 1998.
- [10] Mahnič, V. *Data warehouse: a tool for analysing study process as a value chain*. V: *Beyond the Network – Innovative IT Services*, Proceedings of the 9th International Conference of European University Information Systems, Amsterdam, julij 2003, str. 124–130.
- [11] Oliveira, C. C., Domingues, M. *Multidimensional Analysis of Administrative Data in eLearning Systems*. V: V. Mahnič in B. Vilfan (ur.) *IT Innovation in a Changing World*, Proceedings of the 10th International Conference of European University Information Systems, Bled, Slovenija, 29. junij–2. julij 2004, str. 172–178.
- [12] Oracle, *Oracle9i Warehouse Builder User's Guide, Release 2 (9.0.4)*, 2003.
- [13] Oracle, *Oracle9i Discoverer Plus User's Guide, Version 9.0.2*, 2002.
- [14] Stevenson, D. *Data Warehouse and Executive Information Systems – Ignoring the Hype*. V: J.-F. Desnos in Y. Epelboin (ur.). *European Cooperation in Higher Education Information Systems*, Grenoble, Francija, September 1997, str. 202–207.
- [15] Watson, H. J., Wixom, B. H., Buonamici, J. D., Revak, J. R. *Sherwin-Williams' Data Mart Strategy: Creating Intelligence across the Supply Chain*, Communications of the Association for Information Systems, Vol. 5, Article 9, May 2001.
- [16] Watson, H. J. *Recent Developments in Data Warehousing*, Communications of the Association for Information Systems, Vol. 8, 2001, pp. 1–25.
- [17] Watson, H. J., Goodhue, D. L., Wixom, B. H. *The Benefits of Data Warehousing: Why Some Organizations Realize Exceptional Payoffs*, Information & Management, Vol. 39, 2002, pp. 491–502.

Marko Požnel je diplomiral leta 2001 in magistriral leta 2004 na Fakulteti za računalništvo in informatiko Univerze v Ljubljani, kjer je tudi zaposlen kot asistent. Ukvarja se s podatkovnimi skladišči in razvojem spletnih informacijskih sistemov. Poleg pedagoškega in raziskovalnega dela aktivno sodeluje tudi pri prenovi informacijskega sistema za področje študijske informatike na ljubljanski univerzi.

Viljan Mahnič je izredni profesor in predstojnik Laboratorija za tehnologijo programske opreme na Fakulteti za računalništvo in informatiko Univerze v Ljubljani. V letih 1999–2003 je bil tudi prodekan za pedagoško delo. Ukvarja se z razvojem programske opreme za računalniško podprte informacijske sisteme s posebnim poudarkom na informacijskih sistemih za področje visokega šolstva. Od leta 1996 je predstavnik Slovenije v EUNIS (European University Information Systems Association), od leta 2002 pa tudi član sveta direktorjev omenjene organizacije.