Baze podataka

Predavanja

1. Uvod

ožujak 2021.



"Način na koji prikupljate informacije, upravljate njima i koristite ih, odredit će hoćete li pobijediti ili izgubiti."

Bill Gates

Baze podataka su svuda oko nas ...

Obrada podataka – sveučilište, knjižnica

Sveučilište

- podaci o studentima
- podaci o nastavnicima
- uspjeh na ispitu
- ISVU Informacijski sustav visokih učilišta RH

Knjižnica

- podaci o korisniku knjižnice
- podaci o knjigama
- traženje knjiga i časopisa
- posuđivanje knjiga



Obrada podataka - bankarstvo



- otvaranje računa
- novčane transakcije
- praćenje stanja na računu
- praćenje kupnji ostvarenih putem kreditnih kartica



Obrada podataka - telekomunikacije





- podaci o pozivima
- telefonski računi
- podaci o mreži
- podaci o kvarovima

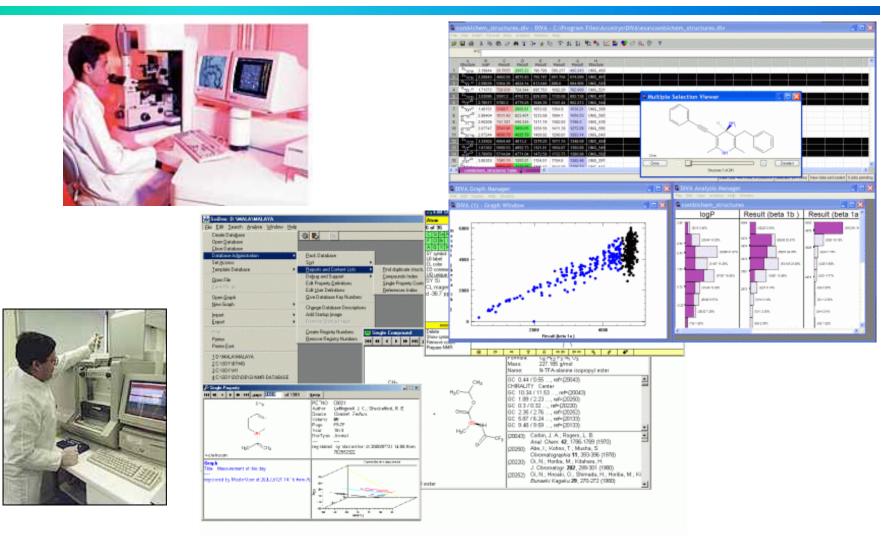








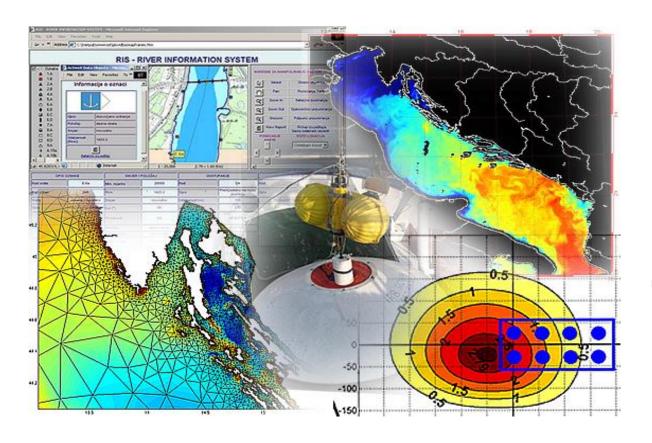
Obrada podataka - znanost



podaci prikupljeni tijekom istraživanja u fizici, biologiji, kemiji...

Obrada podataka - istraživanje i zaštita okoliša

- informatika o okolišu
- satelitska oceanografija



morski sustavi
 (plimna dinamika,
 temperatura, slanost,
 širenje tvari u moru)

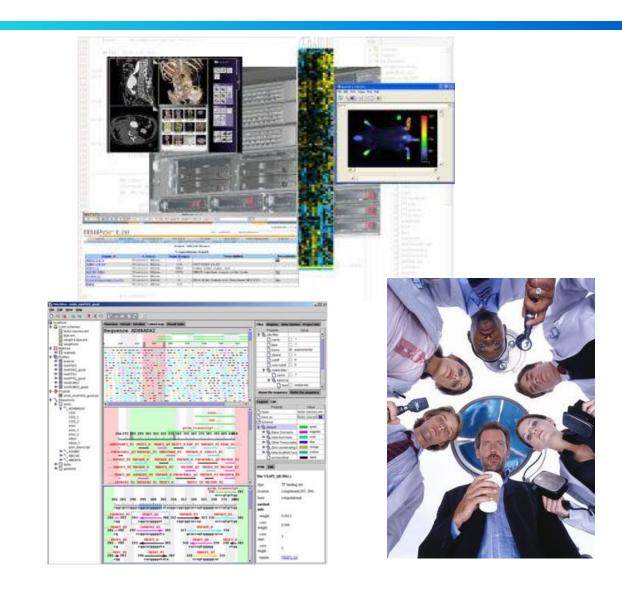
- rezultati mjerenja
- satelitski podaci
- računalne simulacije

ekološko modeliranje

Obrada podataka - medicina

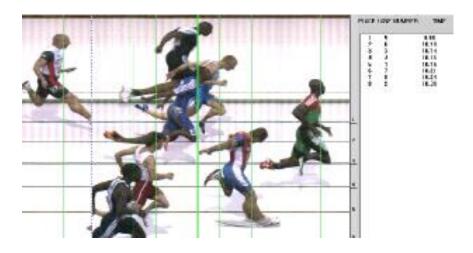


- podaci o pacijentu
- povijest bolesti
- rezultati testova
- podaci s bolničkih instrumenata i senzora



Obrada podataka – sport





0 0 0
22.0
0.0.0
0.0.0



- rezultati utakmica i utrka
- statistika



Obrada podataka – putovanja







- vozni red / red letenja
- rezervacija karata
- kupnja karata
- slobodna i zauzeta mjesta









11

Obrada podataka – (elektronička) trgovina

- Trgovina
 - računi
 - količina prodanih proizvoda
 - zarada



- Elektronička trgovina
 - narudžbe i prodaja putem Interneta

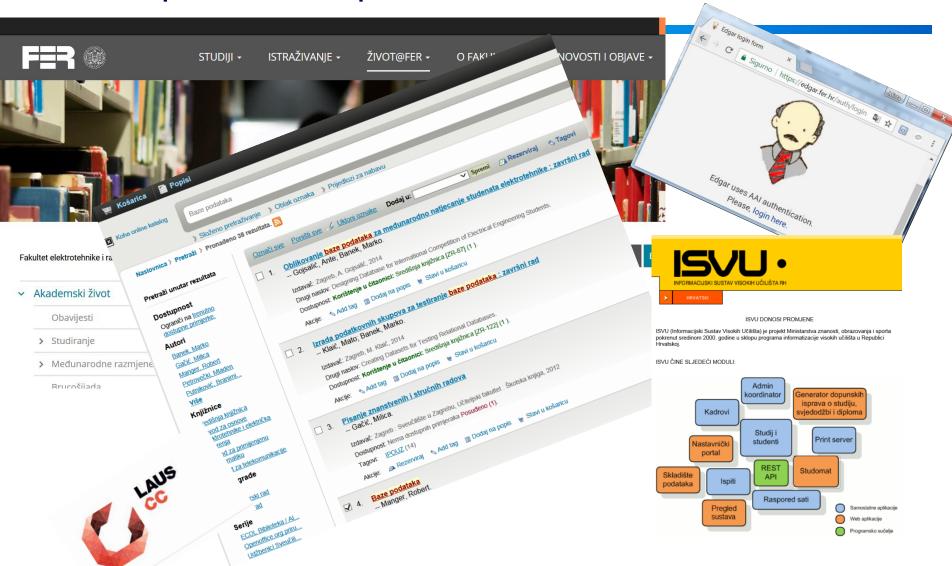








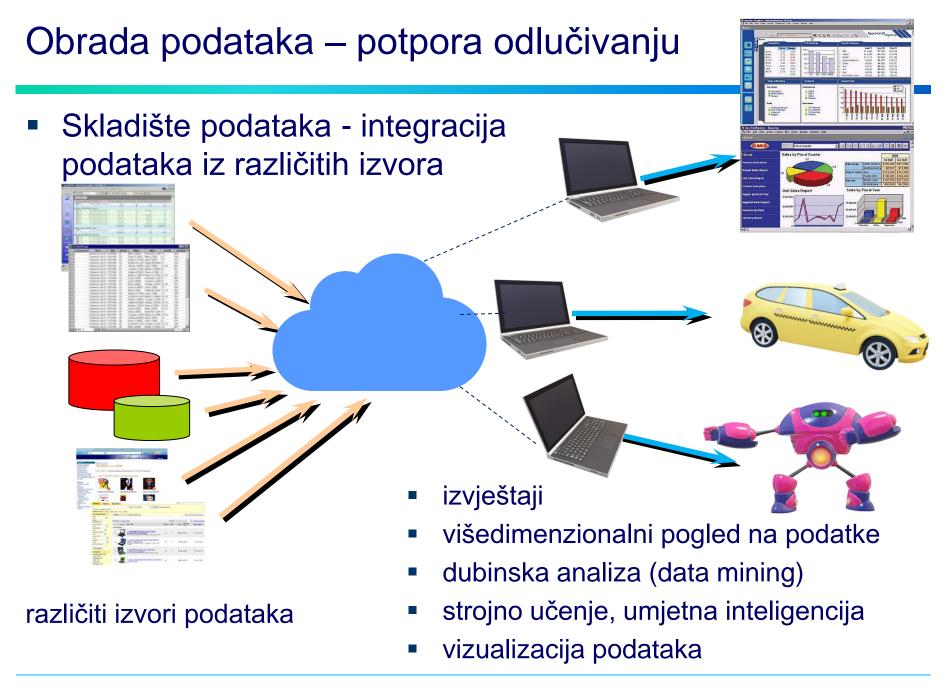
Obrada podataka – portali



prikaz podataka iz različitih izvora

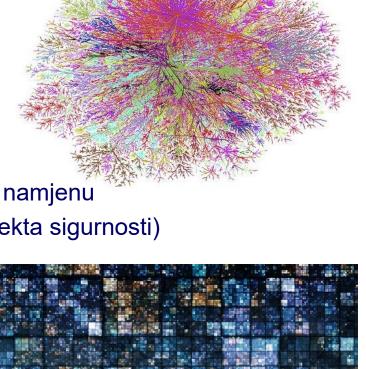


■ Facebook – 2,7 mlrd. korisnika, 60 mil. upita i 4 mil. promjena u sec



Big Data / NoSQL sustavi

- Big Data (3, 4, 5, 6, 7, 10) V's:
 - Volume = volumen, veličina
 - Velocity = brzina
 - Variety = različitost, heterogenost
 - Variability = različitost, nekonzistentnost
 - Veracity = istinitost, točnost
 - Validity = točnost i ispravnost u odnosu na namjenu
 - Vulnerability = osjetljivost naročito s aspekta sigurnosti)
 - Volatility = promjenjivost
 - Visualisation = vizualizacija
 - **Value** = vrijednost
- NoSQL nerelacijski sustavi



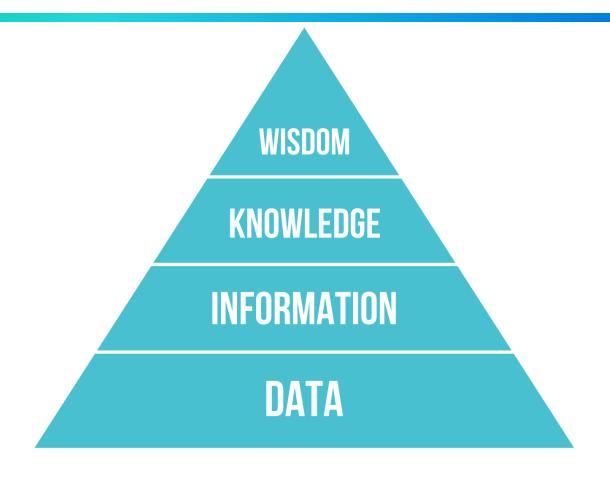
Uvod u baze podataka

Stvarni svijet Informacije **Podaci**

Informacija, podatak (Information, Data)

- INFORMACIJA je sadržaj koji primatelju opisuje nove činjenice.
- Taj sadržaj se materijalizira u obliku PODATAKA koji služe za prikaz informacija u svrhu spremanja, prijenosa i obrade.
- Podatak je skup simbola (znakova).
- Informacija je i obrađeni podatak koji za primatelja ima karakter novosti, otklanja neizvjesnost i služi kao podloga za odlučivanje.
 - Podatak izvan konteksta nema značenja
 - podatak: 4.62
 - Podatak koji interpretiramo i primjereno povežemo predstavlja informaciju
 - informacija: prosjek svih ocjena studenta Marka Horvata na studiju na FER-u u ovom trenutku je 4.62

Podatak, informacija, znanje, mudrost



Može se definirati preslikavanje pojmova iz piramide: know-nothing, know-what, know-how, and know-why (Milan Zeleny, 1987)

https://en.wikipedia.org/wiki/DIKW_pyramid

20

Informacijski sustav (Information System)

Definicija:

- Ukupna infrastruktura, organizacija, osoblje i komponente koje služe za prikupljanje, obradu, pohranu, prijenos, prikaz, širenje i raspolaganje informacijama
- The entire infrastructure, organization, personnel, and components for the collection, processing, storage, transmission, display, dissemination, and disposition of information [INFOSEC-99].

Informacijski sustav (Information System)

- Informacijski sustav je dio svakog organizacijskog sustava
- Svrha mu je prikupljanje, obrada, pohranjivanje i distribucija informacija, koje su potrebne za praćenje rada i upravljanje organizacijskim sustavom ili nekim njegovim podsustavom.
 - Informacijski sustav je aktivni sustav koji može (ali ne mora) koristiti suvremenu informacijsku tehnologiju
 - Središnji dio informacijskog sustava je BAZA PODATAKA

Informacijski sustav, 1868.



Organizacijski sustav

- ORGANIZACIJSKI SUSTAV je složeni sustav koji sadrži tehničke i humane podsustave
 - poduzeće, ustanova, djelatnost, društvena organizacija, tehnički sustav kao npr. telekomunikacijska mreža i sl.
 - često se organizacijski sustav naziva i POSLOVNIM SUSTAVOM, iako pojam organizacijski sustav ima nešto šire značenje.
- Primjeri organizacijskih sustava:
 - Knjižnica
 - Sveučilište
 - Zračna luka

Baza podataka (Database)

 BAZA PODATAKA je skup podataka koji su pohranjeni i organizirani tako da mogu zadovoljiti zahtjeve korisnika.
 (M. Vetter, 1981.)

BAZA PODATAKA je skup međusobno povezanih podataka, pohranjenih zajedno, uz isključenje bespotrebne zalihosti (redundancije), koji mogu zadovoljiti različite primjene. Podaci su pohranjeni na način neovisan o programima koji ih koriste. Prilikom dodavanja novih podataka, mijenjanja i pretraživanja postojećih podataka primjenjuje se zajednički i kontrolirani pristup. Podaci su strukturirani tako da služe kao osnova za razvoj budućih primjena.

(J. Martin, 1979.).

Entitet (*Entity*)

- bilo što, što ima suštinu ili bit i posjeduje značajke s pomoću kojih se može razlučiti od svoje okoline
- osobe: studenti, radnici, građani, ...
 - student Horvat Ivan (0036123456)
- ostala bića
 - Gupčeva lipa, Policijski pas As
- objekti: vozila, strojevi, uređaji, ulice, zgrade, mjesta, ...
 - Eiffelov toranj, Koncertna dvorana "Vatroslav Lisinski"
- apstraktni pojmovi: boje, predmeti nastavnog programa, ...
 - predmet Baze podataka u studijskom programu FER-3
- događaji (nešto se desilo, dešava se ili se planira da će se desiti)
 - dana 20.07.1969. Ljudi su se prvi put spustili na Mjesec
- povezanost među objektima, osobama, događajima, ...
 - student Horvat Ivan (0036123456) stanuje u Zagrebu
- nešto o čemu želimo prikupljati i pohranjivati podatke

Atribut (Attribute)

- Entitet posjeduje neka SVOJSTVA ili ATRIBUTE koji ga karakteriziraju.
 - Za Informacijski sustav visokih učilišta (ISVU) važna svojstva studenta Horvat Ivana su:
 - Matični broj studenta (JMBAG)
 - Ime
 - Prezime
 - Datum rođenja, ...
- Izbor svojstava (atributa) koje ćemo pratiti ovisi o namjeni informacijskog sustava
 - Horvat Ivan u informacijskom sustavu MUP-a bit će karakteriziran i atributima:
 - Boja kose
 - Boja očiju
 - Otisak prsta, ...

Skup entiteta (Entity Set)

- Slični entiteti se svrstavaju u skupove entiteta
- Slični su oni entiteti kojima se promatraju ista svojstva
- Svi entiteti koji su članovi istog skupa entiteta imaju iste atribute
- "atributi entiteta" ⇔ "atributi skupa entiteta"
 - atributi skupa entiteta PREDMET
 - sifPred
 - nazPred
 - ectsBod
 - nastProg
 - atributi skupa entiteta STUDENT
 - jmbag
 - ime
 - prezime
 - datRod

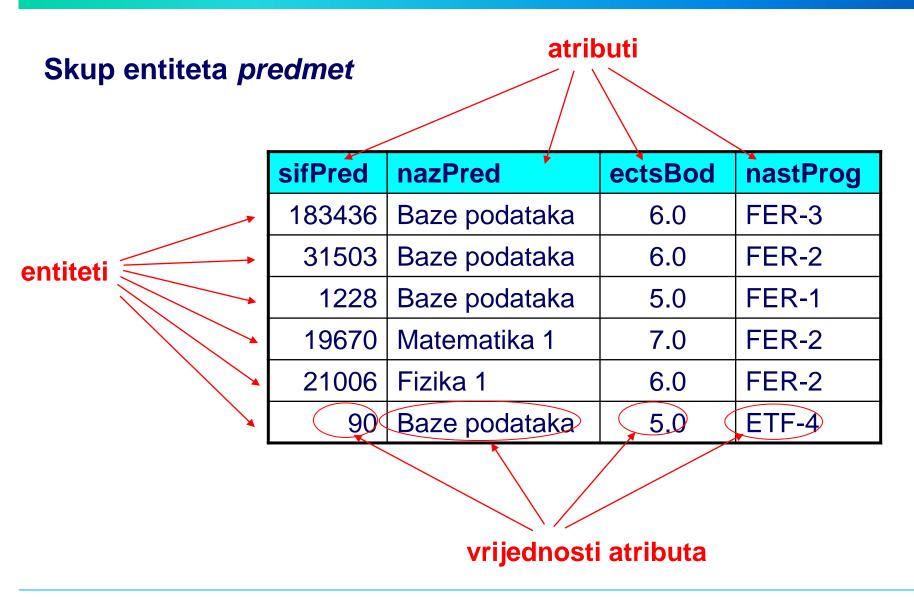
Skupovi entiteta - primjer



Domena i vrijednost atributa (Domain, Attribute Value)

- Za svaki entitet, atribut poprima vrijednosti iz određenog skupa vrijednosti koji predstavlja domenu tog atributa
 - domene atributa za skup entiteta PREDMET
 - sifPred: skup šifara predmeta cijelih brojeva iz intervala [1,999999]
 - nazPred: skup naziva predmeta nizova znakova duljine do 80 znakova
 - ectsBod: skup vrijednosti ECTS bodova realnih brojeva iz intervala
 [0.5, 30.0] (s jednom znamenkom iza decimalne točke)
 - nastProg: skup oznaka nastavnih programa nizova znakova duljine do 5 znakova
 - vrijednosti atributa za entitet Baze podataka (183436)
 - sifPred: 183436
 - nazPred: Baze podataka
 - ectsBod: 6.0
 - nastProg: FER-3

Skup entiteta - prikaz u obliku tablice



Identifikatori entiteta, ključevi

- Skupove atributa čije vrijednosti jednoznačno određuju entitet u promatranom skupu entiteta (dakle ne postoje dva entiteta s posve istim vrijednostima tih atributa) nazivamo IDENTIFIKATORIMA ili KLJUČEVIMA SKUPA ENTITETA.
- Primjer: u skupu entiteta PREDMET prikazanom na slici:

sifPred	nazPred	ectsBod	nastProg
183436	Baze podataka	6.0	FER-3
31503	Baze podataka	6.0	FER-2
1228	Baze podataka	5.0	FER-1
19670	Matematika 1	7.0	FER-2
21006	Fizika 1	6.0	FER-2
90	Baze podataka	5.0	ETF-4

- skup atributa { sifPred } jest ključ skupa entiteta
- skup atributa { nazPred } nije ključ skupa entiteta
- skup atributa { nazPred, nastProg } jest ključ skupa entiteta

Modeliranje stvarnog svijeta

- Modeliranje stvarnog svijeta predstavlja preslikavanje stvarnog svijeta u oblik pogodan za računalnu obradu;
- Baza podataka nekog informacijskog sustava predstavlja sliku stvarnog organizacijskog sustava;
- Stvarni svijet, zbog njegove složenosti, ne možemo prikazati sa svim detaljima;
- Stvarni svijet predstavlja se pojednostavnjenim, nadomjesnim modelom;
- Model stvarnog svijeta predstavlja se uz pomoć nekog formalnog sustava;
- Model podataka je formalni sustav koji koristimo kod modeliranja baza podataka

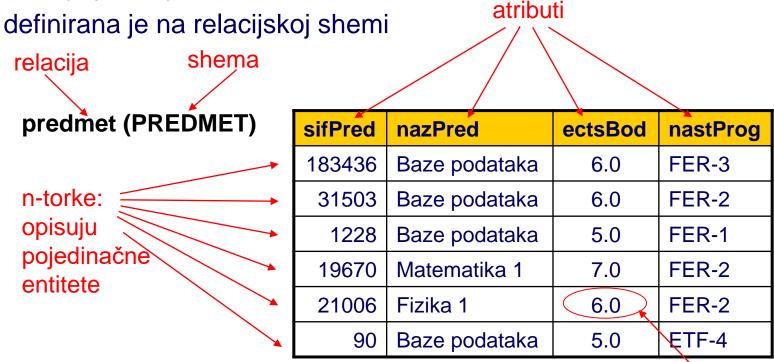
Model podataka (Data Model)

- Model podataka je formalni sustav koji se sastoji od:
 - skupa objekata osnovnih elemenata (koncepata) baze podataka
 - skupa operacija koje se provode nad objektima
 - skupa integritetskih ograničenja (integrity constraints)
 - implicitno ili eksplicitno definiraju skup konzistentnih stanja podataka, promjena stanja, ili oboje
 - Povijesni razvoj modela podataka:
 - Hijerarhijski model
 - Mrežni model
 - Relacijski model
 - ER model
 - Objektni model
 - Objektno-relacijski model

Relacijski model podataka – objekti

elementi skupa objekata u relacijskom modelu podataka su relacije





elementarni podatak

- shema relacije obuhvaća naziv relacijske sheme (PREDMET) i skup atributa: (sifPred, nazPred, ectsBod, nastProg)
- PREDMET = sifPred, nazPred, ectsBod, nastProg

Relacijski model podataka – operacije

operacija "selekcija" u relacijskom modelu podataka:

predmet

sifPred	nazPred	ectsBod	nastProg
183436	Baze podataka	6.0	FER-3
31503	Baze podataka	6.0	FER-2
1228	Baze podataka	5.0	FER-1
19670	Matematika 1	7.0	FER-2
21006	Fizika 1	6.0	FER-2
90	Baze podataka	5.0	ETF-4

 $\sigma_{\text{ectsBod=6.0}}$ (predmet)

sifPred	nazPred	ectsBod	nastProg
183436	Baze podataka	6.0	FER-3
31503	Baze podataka	6.0	FER-2
21006	Fizika 1	6.0	FER-2

- ostale operacije u relacijskom modelu podataka
 - unija, razlika, presjek, projekcija, ...

Relacijski model podataka – integritetska ograničenja

pravilo domenskog integriteta u relacijskom modelu podataka:

predmet

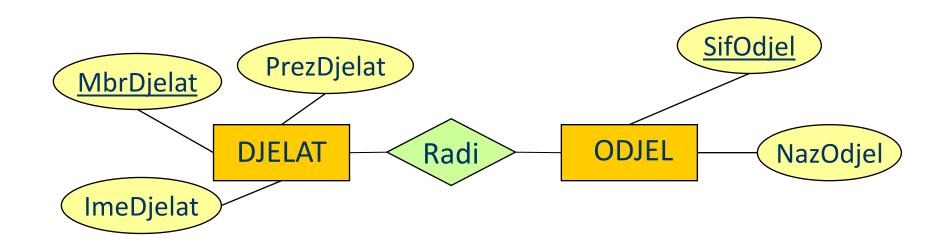
sifPred	nazPred	ectsBod	nastProg
183436	Baze podataka	6.0	FER-3
31503	Baze podataka	6.0	FER-2
1228	Baze podataka	5.0	FER-1
19670	Matematika 1	7.0	FER-2
21006	Fizika 1	6.0	FER-2
90	Baze podataka	5.0	ETF-4

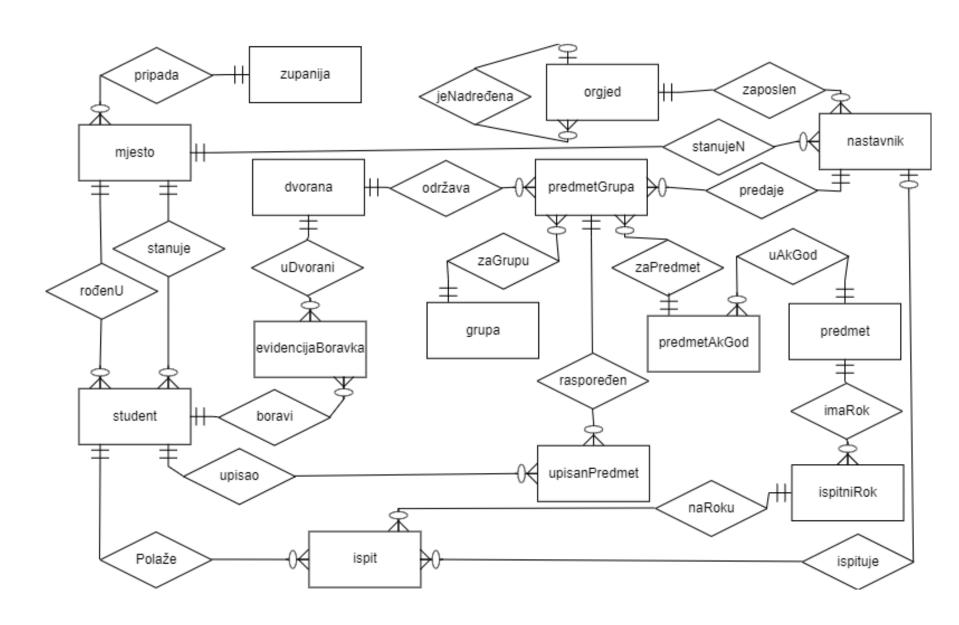
- domenski integritet:
 - vrijednost atributa sifPred mora biti iz intervala [1, 999999]
 - vrijednost atributa ectsBod mora biti iz intervala [0.5, 30.0]
 - ...
- ostala integritetska ograničenja u relacijskom modelu podataka:
 - entitetski, referencijski, ...

ER model podataka

Model entiteti-veze Entity-Relationship Model

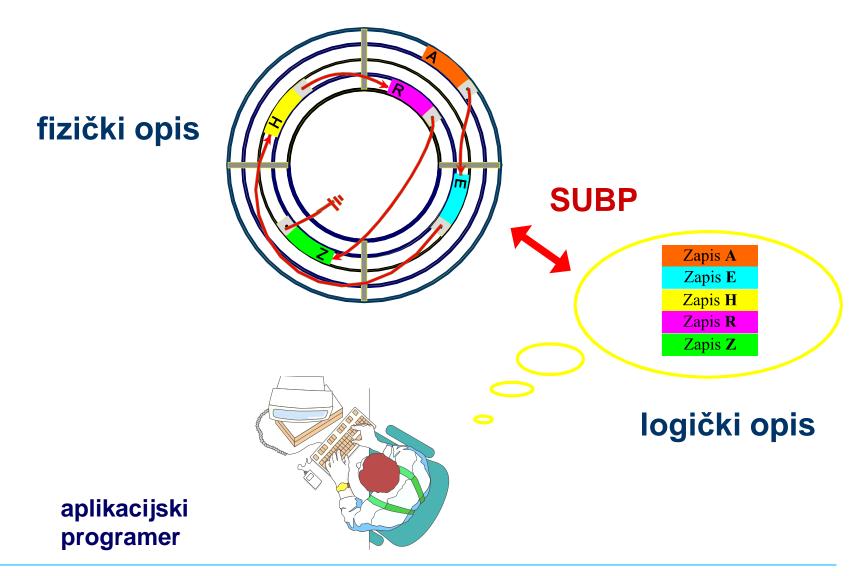
- Postrelacijski model
- Zadržava dobra svojstva relacijskog modela
- Objekti ER modela su entiteti i njihove međusobne veze



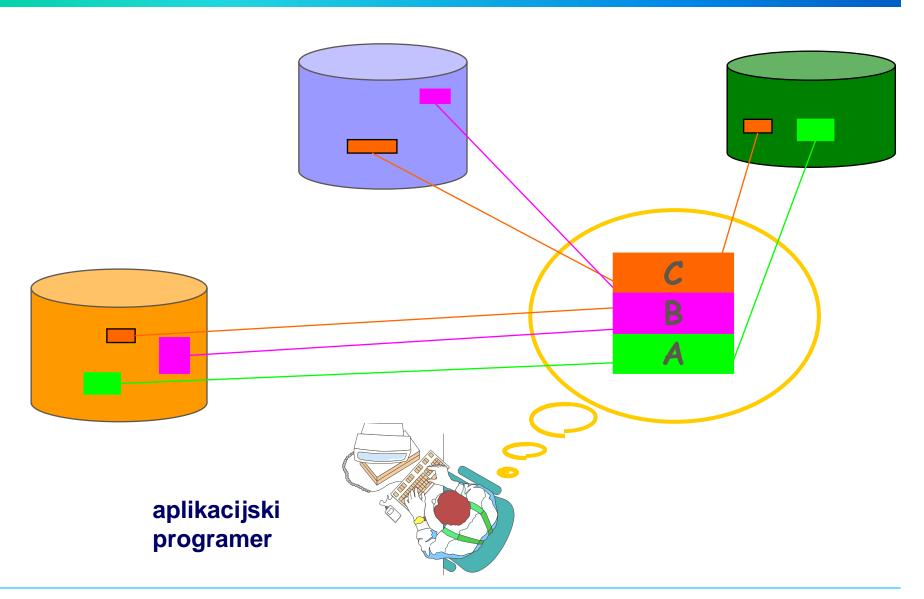


Arhitektura baze podataka

Fizička i logička organizacija podataka



Fizička i logička organizacija podataka



Arhitektura baze podataka

korisnici / aplikacije

Vanjska razina

EKSTERNA
SHEMA 1

EKSTERNA
SHEMA 2

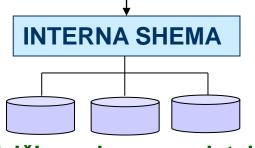
Preslikavanja: konceptualna ↔ eksterne sheme

Konceptualna razina

preslikavanje: konceptualna ↔ interna shema

KONCEPTUALNA SHEMA

Unutarnja razina



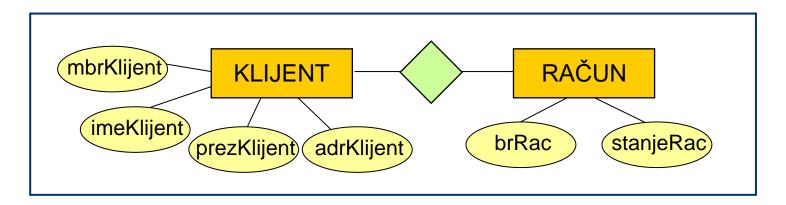
Fizička pohrana podataka

Arhitektura baze podataka

- Shema (struktura) baze podataka se opisuje na tri razine apstrakcije:
 - Na konceptualnoj razini opisuje se
 - KONCEPTUALNA SHEMA
 - Na unutarnjoj razini opisuje se
 - INTERNA SHEMA
 - Na vanjskoj razini opisuju se
 - EKSTERNE SHEME
- Jedna baza podataka ima jednu konceptualnu, jednu internu i (najčešće) više eksternih shema
- Shema baze podataka se relativno rijetko mijenja
- Sadržaj ili instanca baze podataka (skup svih podataka baze podataka u određenom trenutku) se ČESTO mijenja

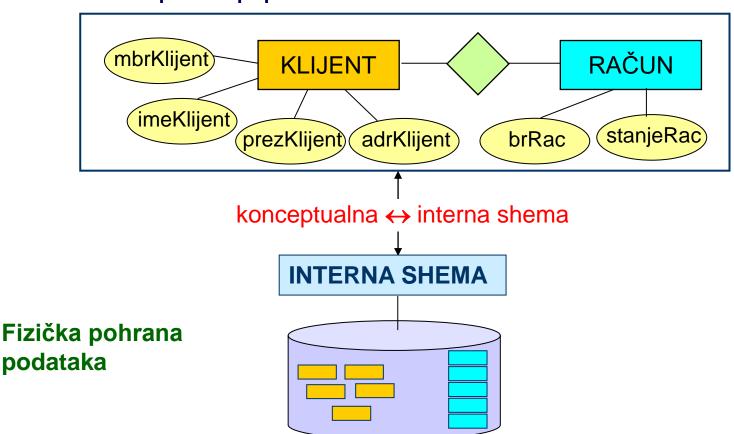
Konceptualna shema

- često se koristi i naziv LOGIČKA SHEMA
- sadrži opis svih entiteta i veza, atributa, domena i integritetska ograničenja
- konceptualna shema se može opisati korištenjem modela podataka, npr. relacijskog ili ER modela



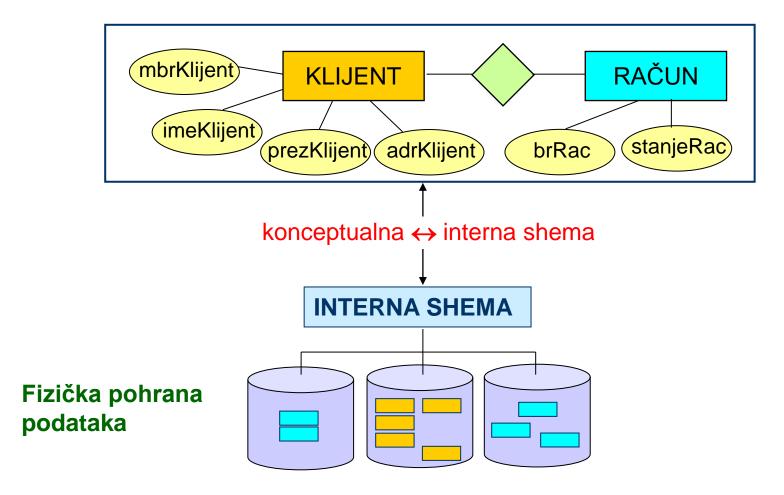
Interna shema

 opisuje detalje fizičke strukture pohrane i metode pristupa podacima: kako su podaci pohranjeni i koje se metode koriste za pristup podacima



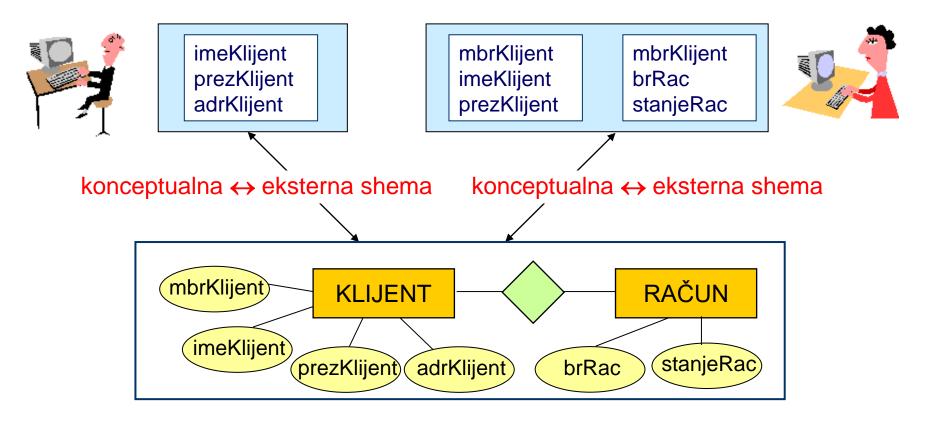
Fizička nezavisnost podataka

izmjena interne sheme ne utječe na konceptualnu shemu



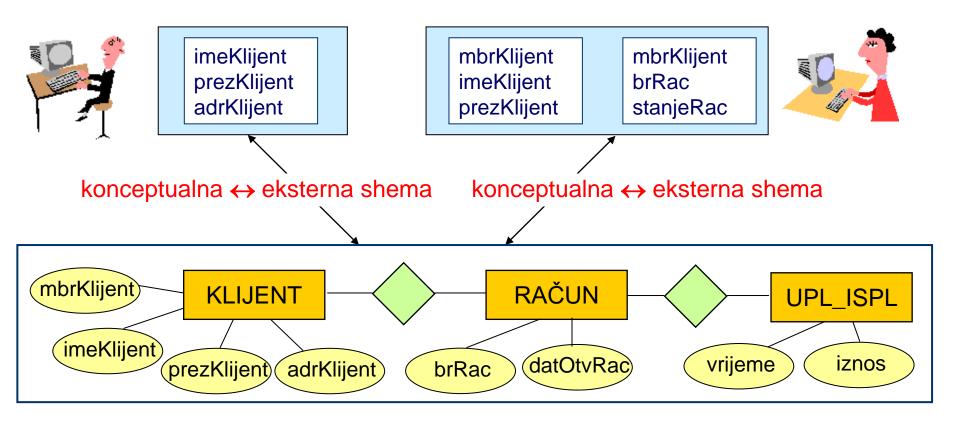
Eksterna shema

- eksterna shema opisuje "pogled" na dio baze podataka koji je namijenjen specifičnoj grupi korisnika
- osnova za opis eksternih shema je konceptualna shema



Logička nezavisnost podataka

 izmjena konceptualne sheme ne mora izazvati izmjenu eksternih shema → izmjena konceptualne sheme ne utječe na korisnike i aplikacijske programe koji ih koriste



Sustav za upravljanje bazom podataka - SUBP

Database Management System - DBMS

Sustav za upravljanje bazom podataka

Sustav za upravljanje bazom podataka - SUBP (Database Management System - DBMS) je **programski sustav** koji omogućava upravljanje podacima u bazi podataka. SUBP se temelji na odabranom modelu podataka.

Prema modelu podataka na kojem se temelje, **SUBP**-ove dijelimo na:

- · hijerarhijske,
- mrežne,
- relacijske,
- objektno-relacijske,
- objektno-orijentirane.

Sustav za upravljanje bazom podataka

- sakriva od korisnika detalje fizičke pohrane podataka
- osigurava logičku i fizičku nezavisnost podataka
- omogućuje definiciju i rukovanje podacima
 - DDL Data Definition Language
 - DML Data Manipulation Language
- obavlja funkciju zaštite podataka
 - integritet podataka
 - pristup podacima autorizacija, sigurnost
 - kontrola paralelnog pristupa
 - obnova u slučaju razrušenja
- obavlja optimiranje upita

Jezici baze podataka

- DDL (Data Definition Language)
 - omogućava imenovanje i opis entiteta, atributa, veza i pripadnih ograničenja integriteta i pravila sigurnosti
 - koristi se za definiranje nove sheme baze podataka ili modificiranje postojeće
 - obavljanje DDL operacija rezultira izmjenom sadržaja rječnika podataka (metapodataka)
- DML (Data Manipulation Language)
 - omogućava korištenje skupa operacija za rukovanje podacima u bazi podataka
 - upitni jezik (query language)
 - ne sasvim korektno, pojam se koristi ne samo za operacije dohvata podataka, već i za operacije izmjene, brisanja i unosa podataka
 - proceduralni jezici, neproceduralni jezici

Zašto koristimo sustave za upravljanje bazama podataka?

Zašto ne bismo koristili samo datotečne sustave?

1. Jednostavan pristup podacima

Datotečni sustav - podaci su raspršeni po datotekama koje mogu biti u različitim formatima. Za dobivanje odgovora na netipična pitanja potrebno je ili dotjerivati stare ili pisati nove programe. Primjer:

Potrebno je naći sve korisnike banke koji žive u području s određenim poštanskim brojem. Pretpostavimo da ne postoji gotovi program koji može izdvojiti tražene korisnike, ali postoji program koji vraća sve korisnike. Dva su moguća rješenja:

- a) službenik će iz popisa svih korisnika ručno izdvojiti one s određenim poštanskim brojem;
- b) tražit će se od programera da napiše odgovarajući program. Niti jedno od ova dva rješenja nije zadovoljavajuće. Svaki put kad se pojavi nova vrsta upita, opet će se morati ponoviti postupak
- SUBP omogućuje fleksibilniji dohvat i izmjenu podataka.

2. Upravljanje zalihostima i nekonzistentnošću Isti podaci mogu biti pohranjeni u više datoteka, što može voditi ka nekonzistentnosti.

Pretpostavimo da se adrese korisnika bankovnog računa pohranjuju u dvije datoteke (jedna za tekući, a druga za devizni račun). Nakon što je korisnik A promijenio adresu stanovanja, bankovni službenik promijenio je podatak o adresi u jednoj datoteci, ali je zaboravio promijeniti adresu u drugoj.

Ako se mjesečni izvještaji o stanju računa šalju na kućnu adresu korisnika na temelju adrese pohranjene u drugoj datoteci, korisnik A neće dobiti izvještaj na svoju novu adresu.

SUBP omogućuje bolju kontrolu zalihosti od datotečnog sustava.

3. Transakcijska obrada

Treba prebaciti 1000 kuna s računa A na račun B (to je jedna transakcija).

Ako se dogodi hardverska ili softverska pogreška za vrijeme izvođenja programa, moguće je da se 1000 kuna skine s računa A, ali se ne uspije prebaciti na račun B, što dovodi do nekonzistentnosti u bazi podataka. Prebacivanje se mora obaviti u potpunosti.

 SUBP obično ima ugrađenu podršku za transakcijsku obradu, što je vrlo teško ostvariti u datotečnom sustavu.

4. Složeni odnosi među podacima

 SUBP omogućuje predstavljanje različitih odnosa među podacima, definiranje novih odnosa kad se oni pojave te jednostavan dohvat i izmjenu međusobno povezanih podataka

5. Istovremeni pristup više korisnika

SUBP omogućuje većem broju korisnika pristup bazi podataka u isto vrijeme. Pri tome treba osigurati da u slučaju kad više korisnika koji nastoje promijeniti isti podatak, to se čini na kontrolirani način tako da rezultat izmjene bude točan i jednoznačan.

Pretpostavimo da službenici dvije turističke agencije nastoje istovremeno rezervirati isto sjedalo u zrakoplovu. SUBP mora osigurati da rezervaciju određenog sjedala u određenom trenutku može obavljati najviše jedan službenik.

U datotečnom sustavu je kontrolu istovremenog pristupa daleko teže provesti.

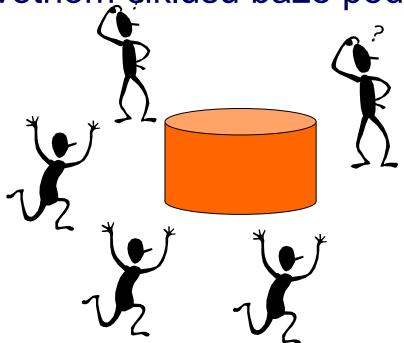
6. Autorizirani pristup

 Kad više korisnika koristi veću bazu podataka, obično neće svim korisnicima biti omogućen pristup svim podacima u bazi.

Na primjer, kako su podaci o financijama povjerljivi, samo će se autoriziranim osobama omogućiti pristup tim podacima.

- Neki korisnici će moći samo pregledavati podatke, dok će ih drugi moći i pregledavati i mijenjati.
- SUBP omogućava određivanje načina na koji će različiti korisnici pristupati podacima.
- U datotečnom sustavu je jednostavno odrediti je li neka datoteka namijenjena za čitanje, pisanje ili oboje, no nije jednostavno definirati različit način pristupa podacima za različite korisnike.

Uloge osoba u životnom ciklusu baze podataka



Baze podataka - uloge

1. Projektanti baze podataka

- razgovaraju s korisnicima da bi saznali njihove zahtjeve
- oblikuju bazu podataka prema zahtjevima korisnika definirajući strukturu za pohranu podataka (tj. model baze podataka)

2. Analitičari sustava i programeri aplikacija

- Analitičari sustava prikupljaju zahtjeve korisnika (npr. bankovnih službenika) i pišu specifikacije za razvoj aplikacija za pristup bazi podataka
- Programeri na temelju tih specifikacija izrađuju programe te ih testiraju, dokumentiraju i održavaju; moraju biti upoznati sa svojstvima i mogućnostima SUBP-a

60

Baze podataka - uloge

3. Administratori baze podataka

- instaliraju i nadograđuju SUBP
- odgovorni su za autorizaciju pristupa bazi podataka
- organiziraju, nadziru i optimiziraju korištenje baze

4. Korisnici

Pristupaju bazi tako da postavljaju upite, mijenjaju podatke i izrađuju izvještaje

Razlikujemo nekoliko skupina korisnika:

- korisnici koji povremeno pristupaju bazi koristeći upitni jezik
- korisnici koji često koriste bazu postavljajući standardne upite i radeći standardne promjene koristeći programirana sučelja (npr. službenici u banci, turističkim agencijama...)
- sofisticirani korisnici koji su dobro upoznati s bazom podataka i koriste je na složeniji način (npr. inženjeri, znanstvenici, poslovni analitičari)