Sinteza relacionog modela - uvod

Relacioni model baze podataka se sastoji iz dva dela:

- **Logičkog**
- **Fizičkog modela**

Logički model baze podataka nastaje kao rezultat (uz prethodno definisanu strukturu i oblik podataka koji će činiti sadržaj baze)

- **Analize postojećih i/ili dobijenih novih podataka**
- **Definisanja tabela (relacija) i veza među njima**
- **Dovođenjem modela na relacioni oblik**

Fizički model baze podataka određuje kako su podaci memorisani na memoriji računara i kako se njima manipuliše – danas najčešće na disku.
Problemu sinteze logičkog modela pristupa se na dva načina:

- **Preko modela objekat-veze (MOV, Entity-Relationship Model, E-R model)**
- **Postupkom normalizacije tabela**

Treba naglasiti da jedan pristup ne isključuje drugi, nego se naprotiv, često i dopunjuju.

E-R model

Ovaj pristup sintezi relacione baze podataka prikazan je po prvi put u radu CHEN P.P The Entity Relationship (1976). Najkraća definicija ovog postupka glasi: ***dobijanje saznanja o objektima, vezama među njima, te njihovim svojstvima.***
Chen je predložio da se model zove E-R model (**E**ntity-**R**elationship , tj entiteti-relacije) tj objekti i veze među njima. U našoj literaturi ovaj model se naziva i MOV (skraćenica od **M**odel **O**bjekat-**V**eze).

Osnovne definicije i pojmovi E-R modela

Chen je uveo nekoliko novih pojmova, koji olakšavaju i ubrzavaju optimalnu sintezu Cood-ovog relacionog modela. Objekte opisane atributima Chen je podelio na:

- **Objekte (u užem smislu)**
- **Veze**
- **Vezne objekte**

Objekti

Status objekta u E-R modelu imaju oni entiteti koji pored identifikatora objekta (primarnog ključa) imaju još i neka svojstva koja se opisuju *atributima. Atribut* je svojstvo objekta koje se opisuje jednim podatkom. Ukoliko je za opisivanje atributa potrebno više podataka, onda taj atribut predstavlja novi objekat.

Neka je objekat PRODAVNICA opisan sledećim atributima:

PRODAVNICA ***<šifraprod#, naziv, delatnost, grad>***

Objekat PRODAVNICA, definisan na ovaj način, je entitet koji ima identifikator – ključ (šifraprod#) i tri atributa (koja opisuju njegova svojstva koja su od značaja za poslovanje). U sledećem koraku treba utvrditi da li svi atributi zadovoljavaju postavljeni kriterijum.
Ako pretpostavimo da je:

- Šifra prodavnice jednoznačna (što mora biti)
- Da svaka prodavnica ima samo jedno ime
- Ima definisanu delatnost

onda su i atributi “naziv” i “delatnost” atributi. Međutim, ako za atribut “grad” pored imena postoji i podatak o broju stanovnika u tom gradu, onda “grad” ne može biti atribut jer ga opisuju dva podatka (naziv i broj stanovnika). Grad predstavlja stoga objekat koji mora biti definisan, na primer kao:

GRAD ***<šifragrada#,naziv,brojstan>***

a u datoteci PRODAVNICA atribut “grad” se briše. Na kraju, treba još ustanoviti i definisati vezu između tih objekata (PRODAVNICA, GRAD).

Generalno, objekti u E-R modelu mogu se podeliti na:

- Čvrste objekte – objekti u punom smislu te reči
- Slabe objekte koji na neki način (egzistencijalno ili identifikaciono) zavise od jednog ili više čvrstih objekata

Pod čvrstim objektom smatra se onaj objekat koji se može u potpunosti definisati primarnim ključem i nizom atributa. To su, npr., objekti:

STUDENT ***<brind#, prez, ime, datrodj, adresa, tel, ……>***
RAČUN ***<brojračuna#, datum, iznos, ……..>***
HOTEL ***<šifrahotela#, naziv, adresa, tel, ……..>***

Pod slabim, egzistencijalno i/ili identifikaciono zavisnim objektom, smatra se onaj koji egzistencijalno i/ili identifikaciono zavisi od nekog drugog, čvrstog, objekta.
Na primer, objekat STAVKA nekog računa:

STAVKA ***<brojstavke#, količina, cena, ……..>***

može bit prisutan samo onda ako pripada nekom računu. RAČUN je čvrst objekat, a STAVKA je slabi objekat koji može postojati samo ako postoji odgovarajući RAČUN (egzistencijalna zavisnost).

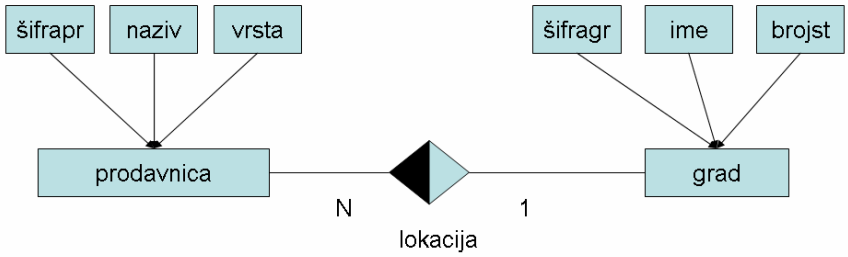
Na ovaj način definisan objekat STAVKA nije samo egzistencijlano, nego i identifikaciono zavisan od čvrstog objekta RAČUN, jer stavke iz raznih računa mogu imati isti broj (istu šifru), pa ih prema tome ne bi bilo moguće jednoznačno identifikovati i međusobno razlikovati bez navođenja i broja računa kojem ta stavka pripada. Korektno se objekat STAVKA stoga mora definisati uz RAČUN kao:

RAČUN ***<brojračuna#, iznos, …….>***
STAVKA ***<brojstavke#, brojračuna#, kolličina, cena, …..>***

Zaključak: ***slabi objekti koji identifikacijski zavise od čvrstih objekata moraju imati složen primarni ključ koji se sastoji od ključa slabog objekta i primarnih ključeva jakih objekata od kojih zavise.***

Veze

Datoteke PRODAVNICA i GRAD nisu nezavisne, jer se svaka prodavnica odnosi na grad u kojem se nalazi. Neophodno je ova dva objekta međusobno povezati. Nazovimo tu vezu među njima “lokacija”.



Slika 1

Vezu između GRADA i PRODAVNICE čini "lokacija", nalazi se između GRADA i PRODAVNICE a tip uspostavljanja veze (moguće varijante su 1:1, 1:N, M:N) mora biti poznat. U konkretnom slučaju, ako u gradu može biti više prodavnica, a određena prodavnica može biti samo u jednom gradu, tip veze je 1:N.

Ako je veza „lokacija" još i opcionalna, dakle neobavezna,(svaki grad može, ali ne mora, imati prodavnicu posmatranog lanca trgovina), onda tu činjenicu označavamo na strani veze koja je neobavezna, kružićem. Vezu, da bi je u ER modelu razlikovali od objekta, predstavljaćemo grafički rombom. ***Veza tipa 1:1 i 1:N, ne mogu imati vlastite atribute kojima se pobliže opisuju.***

Vezni objekti

Da bi se veze tipa N:M u E-R modelu eliminisale, prevode se pod određenim uslovima u novi objekat, daju im se svojstva objekta, čime takve veze postaju vezni objekti koji onda mogu imati svoja svojstva objekta , imaju prema tome i identifikator (ključ), a mogu, ali i ne moraju, imati atribute. Definisanje tipa veze je stoga osnovni preduslov da se ER modelom realno prikaže stanje stvari.

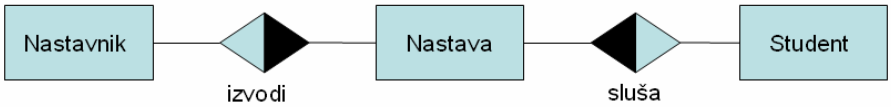
Kada se pri sintezi ER modela pokaže da su veze tipa N:M neminovne (jer egzistiraju u realnom svetu), „eliminacija" ovih veza iz modela izvodi se na sledeći način:

- ***Svaka veza tipa N:M zamenjuje se novim objektom (najčešće istog naziva kao i veza N:M koja se zamenjuje) sa složenim ključem – identifikatorom – koji se sastoji od ključeva objekata koje je veza tipa N:M povezivala***
- ***Nakom toga se novi objekat povezuje vezama 1:N sa postojećim***

Na primer, objekti NASTAVNIK i STUDENT vezani su N:M jer jedan nastavnik komunicira sa više (N) studenata, a svaki student komunicira sa više (M) nastavnika.
ER model sadrži tu vezu (nazovimo je nastava) koju treba eliminisati, tačnije zameniti vezama tipa 1:N.

Slika 2:ER model veze N:M

Zamena veze *„nastava"*, veznim objektom NASTAVA, izvodi se uvođenjem dve nove veze tipa 1:N između NASTAVNIKA i NASTAVE (nazovimo je „izvodi") i STUDENT i NASTAVE (nazovimo je *„sluša"*). Tako dobijamo nov ER model (bez veze N:M) – slika 3.



Slika 3: Transformisani ER model veze N:M

Novouvedeni objekat (NASTAVA) mora imati složen ključ koji se sastoji od ključeva objekata koje povezuje. Pored toga, vezni objekat može, ali ne mora, imati i još neke atribute koji ga bliže opisuju. Na primer, objekat NASTAVA može imati atribut koji daje termin kada se nastava (nastavnika sa šifrom šifran# i studenta sa šifrom broj_indeksa#) izvodi. Prema tome relacioni model može imati sledeći oblik:

NASTAVNIK ***<šifran#, ime, prezime, adresa, telefon,....>***
STUDENT ***<broj_indeksa#, ime, prezime, adresa, telefon, ...>***
NASTAVA ***<šifran#, broj_indeksa#, termin_predavanja, ...>***

Objektu NASTAVA može se dodeliti (po želji) i vlastiti ključ, ali dva spoljna ključa moraju i tada ostati atributi. Na primer:

NASTAVA ***<šifranastave#, šifran, broj_indeksa, termin_predavanja, ...>***

Osobine veza

Pred tipa veze, svaku vezu karakterišu još i sledeće osobine:

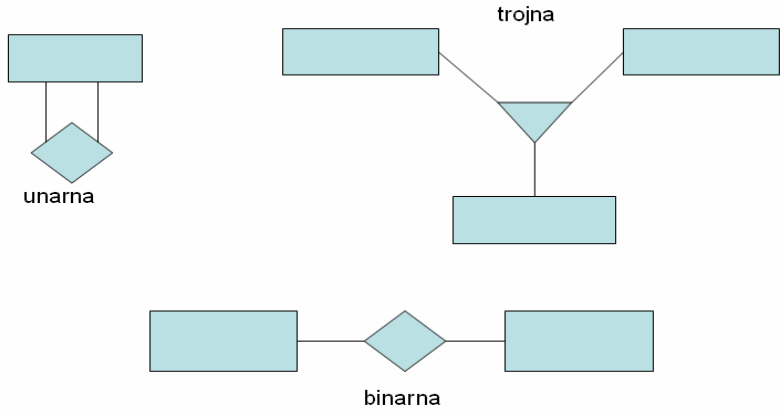
- Red veze
- Kardinalnost veze
- Način uspostavljanja veze

Red veze

Red, odnosno stepen veze, određuje broj objekata koji čine neku vezu. U praksi se najčešće javljaju:

- Unarne
- Binarne
- Trojne veze

Mogu se javiti i veze vreda višeg od tri (n>3) koje se zbog nepreglednosti ER modela u praksi izbegavaju uvek kada je to moguće svođenjem na veze nižeg reda.



Slika 4: unarna, binarna, trojna veza

Unarne (ili unutrašnje) veze su relativno retke u praksi, a uspostavljaju se unutar jedne tabele, jednog objekta. Na primer:

- Neka osoba u opštinskoj datoteci GRAĐANIN (koja sadrži uobičajene podatke kao što su ime, prezime, godina rođenja, mesto rođenja, školska sprema, bračno stanje) može (veza je opcionalna – nije obavezna) da se nalazi u bračnoj vezi sa drugom osobom u istoj datoteci
- Neki takmičar, u disciplini takmičenja parova (na primer, u umetničkom klizanju ili tenisu) u datoteci TAKMIČAR, nalazi se u unarnoj vezi (ova veza je obavezna jer u parovima na može da nastupi jedan takmičar, sam) sa partnerom iz iste datoteke.

Binarna veza je veza između dva objekta i najčešće se susreće u praksi. Takva veza postoji, na primer, između objekata:

- GRAD i PRODAVNICA
- NASTAVNIK i PREDMET
- VOZAČ i VOZILO

Trojna veza nastaje proširenjem binarne veze. Na primer, binarna veza NASTAVNIK i PREDMET mora se proširiti i literaturom koju nastavnik koristi na svom predmetu. U tom slučaju potrebno je uvesti i treći objekat LITERATURA. Novonastala trojna veza (nazovimo je DISCIPLINA) iskazana rečima glasi: nastavnu disciplinu čine:

- PREDMET koji se predaje
- LITERATURA po kojoj se predaje
- NASTAVNIK koji je predaje

Veza reda većeg od 3 iskazuje na sličan način odnos između više od tri entiteta. Analiza im je identična trojnim vezama, ali zbog nepreglednosti modela u praksi se izbegavaju.

Način učestovanja u vezi

Način učestvovanja (prisustva) objekta u vezi može biti dvojak i to:

- Obavezan
- Neobavezan (opcionalan)

Obanezno prisustvo smatramo da postoji onda kada jedan objekat mora biti povezan sa drugim objektom. Na primer, GRAD se mora nalaziti na nekoj LOKACIJI.

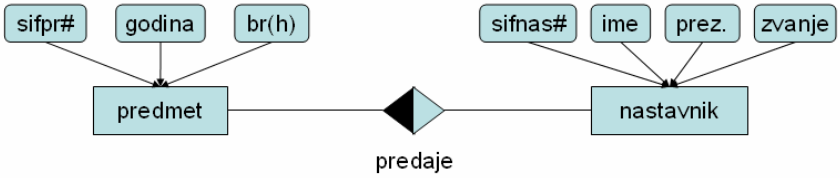
Opcionalna, neobavezna, veza postoji onda kada prethodni uslov nije zadovoljen. Na primer, svaki RADNIK može biti ZADUŽEN nekim POSLOM, ali mogu postojati i takvi poslovi za koje trenutno nije ZADUŽEN nijedan RADNIK. Veza ZADUŽEN između objekta RADNIK i POSAO prema tome bila bi opcionalnog karaktera.

Posmatrajmo sada detaljnije ponašanje i karakteristike objekata u binarnim i trojnim vezama.

Uzmimo za primer bazu podataka obrazovne ustanove u kojoj treba povezati objekte NASTAVNIK i PREDMET a pod uslovom koji je definisan pravilnikom škole i postojećim statutom koji glasi:

- Jedan nastavnik može da predaje više predmeta (NN može da predaje matematiku i fiziku), ali jedan, određeni predmet (PP) u nekom odeljenju može da predaje samo jedan, određeni, nastavnik (na primer fiziku predaje i odeljenju Vv samo NN iako škola ima više profesora fizike). Tip veze je 1:N

Grafički se ER model može prikazati na razne načine. Chen je u njegovom radu koristio simbole kako je to prikazano na slici 5.



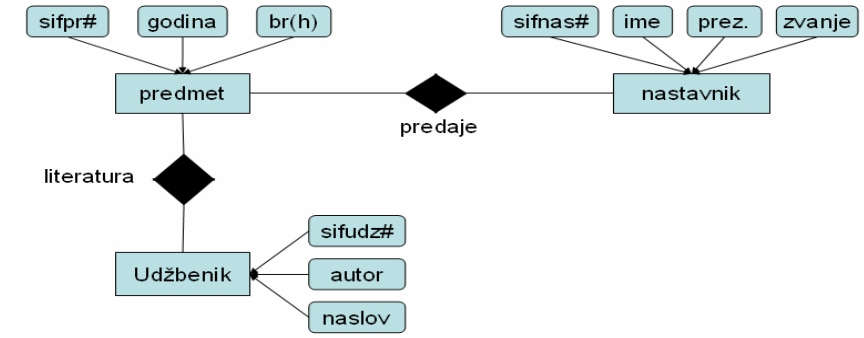
Slika 5: ER model dela informacionog sistema obrazovne ustanove

Dopunimo sada ovaj model i spiskom literature koju na pojedinim predmetima koriste nastavnici, uz uslov da:

- Jedan udžbenik može da se koristi za više predmeta, ali na jednom predmetu može, fakultativno, da se kotisti više različitih udžbenika

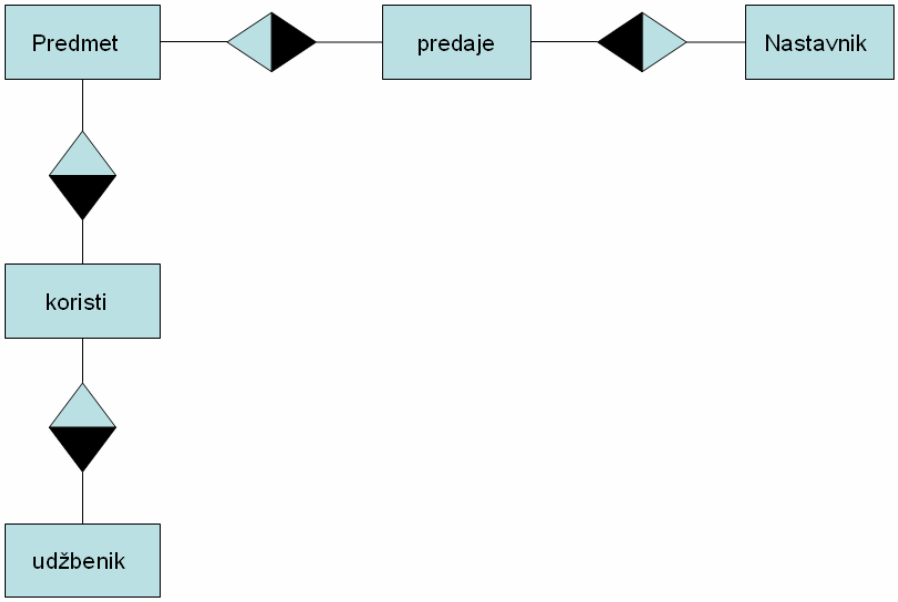
Postojećim datotekama NASTAVNIK i PREDMET te vezi *"predaje"* moramo pridodati novu datoteku UDŽBENIK, kao i novu vezu koja će je povezati sa relacijom PREDMET. Nazovimo tu novu vezu *"literatura"* i definišimo nova pravila poslovanja. Neka su to:

- Jedan predmet može predavati više nastavnika
- Jedan nastavnik može predavati više predmeta
- Za jedan predmet može se koristiti više udžbenika
- Jedan udžbenik može se koristiti za više predmeta



Slika 6: proširen model informacionig sistema obrazovne ustanove

Eliminacija veza tipa N:M izvodi se na isti način i u složenijiim strukturama. Tako, na primer, proširen model informacionog sistema obrazovne ustanove sa slike 6, sa eliminisanim vezama tipa N:M, vidi se na slici 7.



Slika 7: transformisani ER model sa slike 6 sa dva nova vezna objekta

U prethodnom primeru, dvema binarnim vezama N:M nije, na žalost, eksplicitno definisan trojni odnos između PREDMETA, NASTAVNIKA i UDŽBENIKA, jer iz premisa:

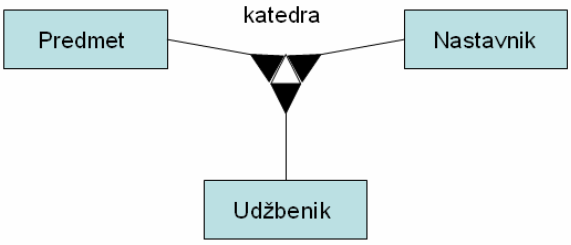
- Jedan predmet izvodi više nastavnika
- Za jedan predmet koristi se više udžbenika

ne sledi obavezno i zaključak da

- Svaki nastavnik koristi za svoj predmet sve udžbenike

ili negacija iste tvrdnje, jer se polazne pretpostavke ne baziraju na jednoj trojnoj, nego na dve biinarne veze.

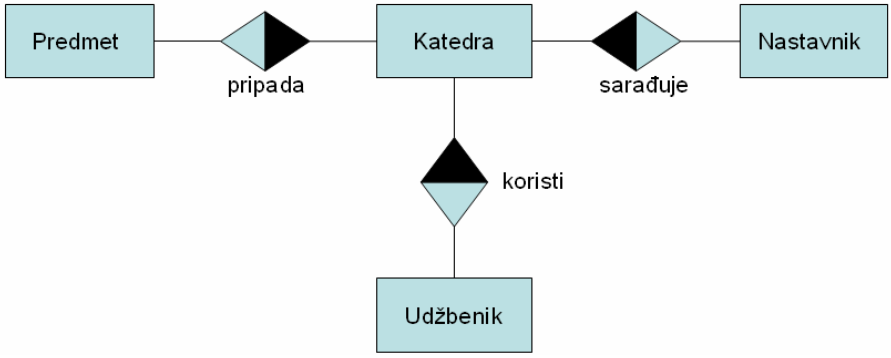
Odgovor na pitanje, "koje udžbenike za svoj predmet koristi neki nastavnik?", ovako koncipiran model ne može dati. Da bi ovo postalo moguće, binarne veze moraju se predefinisati u trojnu vezu kao što je to predstvljeno na modelu trojne veze KATEDRA, na slici 8.



Slika 8: trojna veza KATEDRA među objektima

Ima autora koji se bave ovom problematikom koji smatraju da trojne i veze višeg reda, "zamagljuju" prirodu odnosa među objektima i time smanjuju preglednost modela. Zbog toga se višestruka veza često ponovo zamenjuje sa binarnim vezama. Ovaj postupak zamene može se izvesti na sledeći način:

- Višestruka veza (u poslednjem primeru trojne veze to je veza KATEDRA) predstvlja se kao nov objekat kome se pridružuje njegov identifikator – ključ
- Nakon toga se svaki objekat višestruke (trojne) veze binarno povezuje sa vezom tipa 1:N sa tim novim objektom



Slika 9: model sa tri binarne veze

Vezni objekat KATERDA preveden je u objekat KATEDRA, a uvedene su tri nove veze i to:

- “pripada” – povezuje PREDMET i KATEDRA tipom veze 1:N
- “sarađuje” – povezuje NASTAVNIK i KATEDRA vezom 1:N
- “koristi” – povezuje UDŽBENIK i KATEDRA tipom veze 1:N