

UDC: 004.42:004.658

INFO M: str. 38-44

**PRIMENA FAZI LOGIKE U RELACIONIM BAZAMA PODATAKA  
APPLICATION OF FUZZY LOGIC IN RELATIONAL DATABASES**

Miljan Vučetić

**REZIME:** Fazi relacioni model baza podataka (FRDBM) predstavlja proširenje klasične relacione baze podataka, a proširenje se odnosi pre svega na mogućnost i sposobnost upravljanja nepreciznim i neodređenim informacijama. Zbog toga se u klasičnim relacionim bazama podataka inkorporiraju elementi Zadeh-ove teorije fazi logike. Pitanje nepreciznosti podataka i njihove reprezentacije je jako zastupljena tema i vrlo važna u raznim oblastima i aplikacijama, kao što su geografski informacijski sistemi (GIS) i reprezentacija prostornih podataka, Data mining sistemi, statističke baze podataka, pretraživanje fazi informacija itd. Fazi pristup se kontinualno proučava i neprekidno se javljaju nove mogućnosti primene ovog koncepta u analizi problema različite prirode. Menadžeri baza podataka, na prvom mestu, analiziraju mogućnost generisanja fazi upita, te dodavanje nepreciznih informacija u fazi relacionim modelima baza podataka. U tom smislu predloženi su različiti modeli. Sledeći korak može da predstavlja analiza, karakterizacija i generalizacija podataka kao što je utvrđivanje postojanja zavisnosti, redundantnosti podataka, pojava anomalija itd. U ovom radu prikazane su najznačajnije oblasti istraživanja u okviru modeliranja fazi baza podataka.

**KLJUČNE REČI:** Fazi relacioni modeli baza podataka, fazi skup, fazi upiti, zavisnost podataka, redukcija.

**ABSTRACT:** A fuzzy relational database model (FRDBM) is an extension of classical relational databases and this is related primarily to the possibility and ability to manipulate with imprecise and uncertain information. Consequently, classical relational databases incorporate elements of Zadeh's fuzzy logic and fuzzy sets. The issue of data uncertainty and their representation has been represented as very important topic in many areas and applications such as geographic information systems (GIS) and spatial data representations, Data Mining systems, statistical databases, fuzzy information retrieval etc. Fuzzy approach is continually studying and constantly appearing new possibilities for application of this concept in the different problem analysis. Firstly, database managers analyze the possibility of fuzzy queries generating and adding of imprecise information in fuzzy relational database models. In this sense, various fuzzy database models have been proposed. The next step is analysis, characterization and generalization of data such as functional dependency, data redundancy, data anomalies etc. This paper reviews the most significant research in the areas of fuzzy database modeling.

**KEY WORDS:** Fuzzy relational database models, fuzzy set, data dependencies, redundancy.

## 1. UVOD

Klasični modeli baza podataka često „pate“ od svoje nesposobnosti predstavljanja i manipulisanja nepreciznim i neodređenim informacijama koje se javljaju u stvarnom svetu brojnih aplikacija. Od ranih 80-tih godina, Zadeh-ova fazi logika se koristi za proširenje različitih modela podataka. Svrha uvođenja fazi logike u baze podataka jeste mogućnost predstavljanja i upravljanja neodređenim i nepreciznim informacijama. To je dovelo do brojnih doprinosa u računarskim aplikacijama. Fazi relacione baze podataka na prirodan način proširuju funkciju klasičnih fazi baza znanja, što obezbeđuje viši nivo adaptacije fazi sistema, kao jedne od osnovnih osobina (pored sistema planiranja, učenja, predikcije, sistema za pretragu znanja, robota) inteligentnih sistema. U samoj biti ovih modela podataka je činjenica da postoji veliki broj aktivnih istraživačkih oblasti koje direktno uključuju ili koriste ove baze znanja. Pitanje nepreciznosti podataka i njihove reprezentacije je jako zastupljena tema i vrlo važna u raznim oblastima. Navešćemo samo neke od njih: Geografski informacijski sistemi (GIS) i reprezentacija prostornih podataka, sistemi za otkrivanje znanja u podacima (Data Mining), statistički modeli baza podataka, pretraživanje fazi informacija.

Rasplintute, neprecizne informacije (fazi informacije) mogu biti uključene u sistem baza podataka na dva različita nivoa:

prvi nivo razmatra mogućnost definisanja nepreciznih upita u klasičnim relacionim bazama podataka, a drugi se odnosi na problem dodavanja nepreciznih informacija u bazu. U oba slučaja, teorija fazi skupova predstavlja moćan alat za upravljanje neodređenim i nepreciznim podacima. Razmatranje drugog pristupa nas dovodi do *Fazi Relacionih Modela Baza Podataka*.

## 2. NESAVRŠENE INFORMACIJE I TEORIJA FAZI SKUPOVA

### 2.1. *Neprecizne i neodređene informacije*

Nedoslednost, nepreciznost, neodređenost, neizvesnost i nejasnoća su osnovni tipovi nesavršenih informacija u sistemima baza podataka [1]:

1. Nedoslednost je vrsta semantičkog konflikta, što znači da je isti aspekt stvarnog sveta predstavljen različito u jednoj ili različitim bazama podataka. Npr. Starost jedne osobe je skladištena kao 34 i 37 godina istovremeno.
2. Intuitivno, nepreciznost i neodređenost su relevantni za sadržaj vrednosti atributa, a to znači da vrednost atributa mora da leži u datom opsegu (intervalu ili skupu) vrednosti, ali mi ne znamo koja vrednost će se tačno birati u datom trenutku. U principu, nejasne informa-

cije se predstavljaju lingvističkim varijablama. Npr. Mlad čovek je skup vrednosti {18, 19, 20, 21, 22, 23} što znači da mlad čovek može imati 18 godina, ali i 23 godine.

3. Neizvesnost se odnosi na stepen istine vrednosti atributa, a to znači da možemo da procenimo verodostojnost pojedinih atributa. Na primer, mogućost da je Marko star 35 godina je 97%. Slučajne neizvesnosti koje se opisuju teorijom verovatnoće se ne uzimaju u obzir ovde.
4. Nejasnoća – dvosmislenost znači da nekim elementima modela podataka nedostaje potpuna semantika što doводи do nekoliko mogućih tumačenja.

Generalno, nekoliko različitih nesavršenosti može da koegzistira u odnosu na isti podatak u bazi. Na primer, starost osobe je podatak iz skupa vrednosti {18,19,20,21} i njihove mogućnosti (stepen pripadnosti) je 0.70, 0.85, 0.95 i 0.98 respektivno. Nepreciznost, neizvesnost i nejasnoća su tri glavne vrste nesavršenih informacija sa kojima se susrećemo u tradicionalnim bazama podataka.

## 2.2. Teorija fazi skupova i mogućnost distribucije

Mnogi od postojećih pristupa koji se bave nepreciznošću i neodređenošću informacija se baziraju na teoriji fazi skupova i mogućnosti distribucije. Fazi skup {0.7/18, 0.85/19, 0.95/20, 0.98/21} za starost neke osobe sadrži neprecizne informacije (starost osobe može biti 18, 19, 20 ili 21 godinu) i stepen pripadnosti (0.7, 0.85, 0.95 i 0.98) istovremeno. Jedna od najvažnijih osobina fazi skupova jeste njihova mogućnost da izraze stepen neodređenosti u čovekovom razmišljanju i njegovu subjektivnost. Takva osnovna ideja skupa sa funkcijom pripadnosti ili ponderisanim elementima pokazala se veoma korisnom u analizi znanja i predavljanju informacija.

Neka je  $U$  domen. Fazi vrednosti na domenu  $U$  okarakterisane su sa fazi skupom  $A$  na  $U$  i funkcijom pripadnosti

$$\mu_A: U \rightarrow [0,1] \quad (1)$$

Na ovaj način svaki element  $x \in U$  ima stepen pripadnosti fazi skupu  $\mu_A(x) \in [0,1]$ . Dakle, fazi skup  $A$  kompletno je uređen sa skupom uređenih parova:

$$A = \{(x, \mu_A(x)) : x \in U\} \quad (2)$$

pričemu je  $\mu_A(x)$  stepen pripadnosti elementa  $x$  skupu  $A$ . Ukoliko je  $\mu_A(x)$  veće utoliko ima više istine u tvrdnji da element  $x$  pripada skupu  $A$ .

Analogno za kontinualne fazi skupove važi:

$$A = \int_U \frac{\mu_A(x)}{x} \quad (3)$$

Tri glavna značenja za funkciju pripadnosti, koja se susreću u literaturi su: sličnost, preferencija i neizvesnost. [2]. Svaka od navedenih semantika se nalazi u osnovi posebnih klasa aplikacija. Funkcija pripadnosti fazi skupova ponekad predstavlja i funkciju korisnosti koja predstavlja fleksibilno ograničenje u problemima donošenja odluka.

U narednim odeljcima definisana su referentna tumačenja primene funkcije pripadnosti:

**Stepen sličnosti:** Pomoću funkcije pripadnosti može da se definiše stepen bliskosti i sličnosti posmatranih elemenata. Ovo je, ujedno i najstarija semantika koju uvode Bellman i ostali. Ovaj prikaz je naročito značajan u klaster analizi i regresionoj analizi, gde se susrećemo sa problemom predavljanja podataka i određivanja bliskosti između informacija.

**Stepen preferencije:** Fazi skup  $A$  predstavlja skup manje ili više željenih objekata ili vrednosti varijable na osnovu koje se donose odluke. U tom slučaju,  $A(x)$  predstavlja intenzitet prednosti u korist objekta  $x$ , odnosno preferenciju izbora objekta (vrednosti)  $x$ . Fazi skup je tada kriterijum ili fleksibilno ograničenje. Ovaj pristup koji su izneli Bellman i Zadeh, predstavlja danas osnovu problema fazi optimizacije, fazi linearnog programiranja i analize odluka. Aproksimativno rezonovanje na bazi varijabli i ograničenja koja mogu biti fazi naročito je pogodno za primenu ovakvog pristupa.

**Stepen neizvesnosti:** Ovo tumačenje je predložio Zadeh u svojoj teoriji fazi skupova.  $\mu_A(x)$  predstavlja stepen pripadnosti elementa  $x$  fazi skupu  $A$ . Funkcija pripadnosti rangira vrednosti u smislu njihove verodostojnosti. Ovakav pristup se koristi u ekspertnim sistemima.

Jedna od najvažnijih osobina fazi logike jeste njena mogućnost da izrazi stepen neodređenosti u čovekovom razmišljanju i njegovu subjektivnost. Fazi logika se često pogrešno poistovećuje sa verovatnoćom. Osnovna razlika između fazi logike i teorije verovatnoće sastoji se u tome da fazi logika operiše sa determinističkim nedorečenostima i neodređenostima, dok se verovatnoća bavi fenomenom ponavljanja i simbolizuje se slučajnošću (slučajnim promenljivim ili slučajnim procesima). Fazi logika prosto ima za cilj prevazilaženje problema u komunikaciji vezanim za razlike između pravila koja nameću formalne teorije i načina razmišljanja koji opisuju ponašanje ljudskog uma. Drugim rečima, fazi i slučajni su dva atributa koji se razlikuju u svojoj prirodi, odnosno, oni opisuju drugačiji aspekt neodređenosti. Dakle, fazi logika pokriva subjektivnost ljudskog mišljenja, osećanja, jezika, dok verovatnoća pokriva objektivnu statistiku u prirodnim naukama. Analogno, fazi modeli i modeli formirani na bazi verovatnoće nose drugačiji vid informacija: fazi funkcija pripadnosti predstavlja sličnost objekata u kontekstu neprecizne definicije osobina, dok verovatnoća daje informaciju o frekvenciji ponavljanja.

## 3. FAZI RELACIONI MODELI BAZA PODATAKA

U prethodnoj deceniji veliki broj autora [3-8] postavio je pitanje uvođenja nepreciznosti i neodređenosti u relacionim bazama podataka, a to nas dovodi do sistema baze podataka koji leže u okviru sistema veštačke inteligencije, jer nam se omogućava da upravljamo informacijama koje su veoma slične prirodnom jeziku. Codd uvodi organizaciju relacione baze podataka koja je zasnovana na teoriji skupova i relacionoj teoriji. Zadehova teorija fazi skupova predstavlja generalizaciju opšte teorije skupova, dok je koncept fazi relacije generaliza-

cija relacije teorije. Dakle, za proširenje klasičnog modela podataka u fazi model najprije definišimo okvire za predstavljanje podataka koji generišu gore spomenute pretpostavke. U ovom poglavlju prikazujemo četiri okvira za predstavljanje podataka u fazi modelu [9-12] i tofazi model na bazi sličnosti, fazi model na bazi mogućnosti, fazi model na bazi fazi relacije i fazi model sa proširenom mogućnosti.

Naka je data relacija  $R(\text{lme, Adresa, Godine, Produktivnost, Zarada})$ . Posmatrajmo različite tipove n-torki koje u relaciji mogu da se pojave u jednoj od sledećih formi:

- $T_1$ : (Marko, Bulevar revolucije, {21,22,23}, dobra, visoka ili srednja)
- $T_2$ : (Milan, Knez Mihajlova 10, {0.7/22, 1/25, 0.8/28}, odlična, {niska, srednja})
- $T_3$ : (Saša, Kneza Miloša 2, 27, zadovoljavajuća, visoka):  $\mu$ , gde je  $\mu$  pripadnost n-torkerelaciji i  $\mu \in [0,1]$ .

U datom primeru primećujemo da se kao vrednosti atributa pojavljuju lingvistički izrazi (niska, srednja, visoka, zadovoljavajuća...), fazi skup (0.7/22, 1/25, 0.8/28), podskup datoga domena (niska, srednja). Zatim, primetimo da n - torke mogu pripadati relaciji sa nekim parcijalnim stepenom pripadnosti.

U fazi modelu na bazi sličnosti svaki domen skupa atributa u fazi relacionim bazama podataka, kao prvo asociiran je sa relacijom sličnosti, umesto relacijom identiteta, i drugo domen vrednosti je definisan kao podskup osnovnog baznoga skupa, umesto jednoga elementa domena kao u klasičnim relacionim bazama podataka. Tako imamo sledeću definiciju:

**Definicija 3.1.** Fazi relacija R je podskup skupa Dekartovog proizvoda  $2^{D_1} \times 2^{D_2} \times \dots \times 2^{D_n}$  gde je  $D_i$  konačan domen i  $2^{D_i}$  je partitivni skup skupa  $D_i$ . Bilo koji član relacije naziva se jednostavno n - torkom.

Prema tome, fazi n - torca  $t_i$  je bilo koji član od  $R$  i  $2^{D_1} \times 2^{D_2} \times \dots \times 2^{D_n}$  koja je data u formi  $t_i = (d_{i1}, d_{i2}, \dots, d_{ij}, \dots, d_{in})$ , gde je  $d_{ij} \subset D_j, j=1..n, d_{ij} \neq \emptyset$ .

Relacija sličnosti  $s_j$  na  $D_j$  je preslikavanje  $D_j \times D_j \rightarrow [0,1]$ , tako da  $x, y, z \in D_j$  važi:

- a)  $s_j(x,x) = 1$  (refleksivna)
- b)  $s_j(x,y) = s_j(y,x)$  (simetrična)
- c)  $s_j(x,z) > \max(\min(s_j(x,y), s_j(y,z)), \min(s_j(x,y), s_j(y,z)))$  (max-min tranzitivna).

Primer ovoga modela moguće je predstaviti n-torkom  $T_1$ :

- $T_1$ : (Marko, Bulevar revolucije, {21,22,23}, dobra, visoka ili srednja).

Fazi model na bazi mogućnosti proširuje teoriju klasičnih relacionih baza podataka dozvoljavajući upotrebu fazi vrednosti za attribute. Fundamentalni koncept fazi informacija jeste da promenljiva (atribut) nije vezana za konkretnu vrednost. U tom kontekstu se koristi termin mogućnost distribucije, koja svaku moguću vrednost promenljive povezuje sa vrednostima iz intervala [0,1]. Uopšteno, mogućnost distribucije se identifikuje sa funkcijom stepena pripadnosti. Zato kažemo da element x pripada fazi skupu A ("Visina osobe") sa stepenom pripadnosti 0.9. Onda je mogućnost distribucije elementa  $x \in U$  na fazi skupa A, koji je definisan na domenu U,  $\pi_A(x) = 0.9$ .

**Definicija 3.2.** Fazi relacija R je podskup domena  $\Pi(D_1) \times \dots \times \Pi(D_n)$ , gde je:  $\Pi(D_i) = \{\pi_{Ai} | \pi_{Ai} \text{ je moguća distribucija od } A_i \text{ na } D_i\}$ . Odgovarajuća n - torca je data u formi  $t_i = (\pi_{A1}, \pi_{A2}, \dots, \pi_{An})$ .

Dalje u ovome modelu uveden je jedan poseban element e za takozvane neprimenljive situacije koje mogu poprimiti odgovarajuće vrednosti atributa. U tu svrhu definišemo sledeću funkciju:

$$\pi_{Ai}: D \cup e \rightarrow [0,1] \quad (4)$$

Primer ovog modela moguće je predstaviti n-torkom  $T_2$ :

- $T_2$ : (Milan, Knez Mihajlova 10, {0.7/22, 1/25, 0.8/28}, odlična, {niska, srednja}).

U fazi modelu na bazi fazi relacije (Baldwin), fazi relacija definisana je na sledeći način:

**Definicija 3.3.** Fazi relacija R na  $D_1 \times \dots \times D_n$  određena je sa funkcijom pripadnosti:  $\mu_R: D_1 \times \dots \times D_n \rightarrow [0,1]$ , gde je  $D_i$  domen atributa  $A_i$ . Generalna forma binarne relacije R na  $D_1 \times D_2$  predstavljena je kao:  $R = \{\mu_R(u_1, v_1) / (u_1, v_1), \dots, \mu_R(u_m, v_m) / (u_m, v_m)\}$  u n - torci date sa  $R = \{u_1, v_1, \mu_R(u_1, v_1), \dots, u_m, v_m, \mu_R(u_m, v_m)\}$  gde je  $u_j \in D_1, j = 1, 2, \dots, m$  i  $v_k \in D_2, k = 1, 2, \dots, n$ .

Ovaj model određuje da je pripadnost neke n - torke datoj relaciji jeste fazi, dok individualne vrednosti atributa ne moraju biti fazi, ili mogu biti lingvističke varijable, ali tada se tretiraju kao atomske ili kao jedno - vrednosne varijable. Primer ovog modela moguće je predstaviti n-torkom  $T_3$  kao:

- $T_3$ : (Saša, Kneza Miloša 2, 27, zadovoljavajuća, visoka):  $\mu$

Fazi model sa proširenom mogućnošću proširuje prethodno opisan modelna bazi mogućnosti, ne samo dozvoljavajući moguću distribuciju vrednostima atributa, već dozvoljavajući i relaciju bliskosti koja je u vezi sa datim domenom. Ovo proširenje generalizuje klasični relacioni model baza podataka. Naznačimo da je relacija sličnosti samo specijalni slučaj relacije bliskosti kod koje se zahteva da bude refleksivna i simetrična.

**Definicija 3.4.** Relacija bliskosti  $c_i$  na domenu  $D_i$  je preslikavanje  $D_i \times D_i \rightarrow [0,1]$  tako da za svako  $x, y \in D_i$ :

- a)  $c_i(x,x) = 1$  (refleksivna)
- b)  $c_i(x,y) = c_i(y,x)$  (simetrična)

Na taj način gore spomenuti okviri postaju specijalni slučajevi osnovnog radnog okvir sa proširenom mogućnošću.

**Definicija 3.5.** Fazi relacija R je podskup domena  $\Pi(D_1) \times \dots \times \Pi(D_n)$ , gde je:  $\Pi(D_i) = \{\pi_{Ai} | \pi_{Ai} \text{ je moguća distribucija od } A_i \text{ na } D_i\}$ . Odgovarajuća n - torca je data u formi  $t_i = (\pi_{A1}, \pi_{A2}, \dots, \pi_{An})$ .

Na taj način pomoću relacije bliskosti određujemo bliskost elemenata odgovarajućih atributa na datom domenu. Primer ovoga modela moguće je predstaviti informacijom:

- $T_2$ : (Milan, Knez Mihajlova 10, {0.7/22, 1/25, 0.8/28}, odlična, {niska, srednja})

3.1. GEFRED

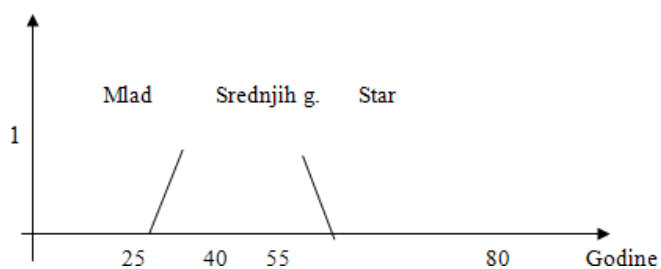
Grupa autora [13,14] uvodi model nazvan GEFRED (General Fuzzy Relational Database). To je generalizovani fazi relacioni model baza podataka koji inkorporira elemente prethodnih studija u jedinstveni model. U ovom delu uvodimo osnovne elemente fazi proširenja ovog relacionog modela baza podataka. Struktura GEFRED modela može da se prikaže na sledeći način:

$$R_{FG} \in (D_{G_1}, C_1) \times \dots \times (D_{G_n}, C_n),$$

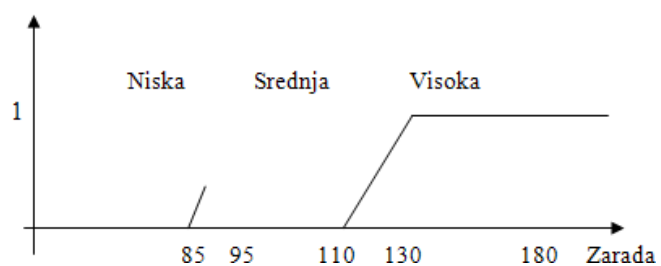
gde je  $D_{G_i}$  domen atributa, a  $C_i$  "kompatibilnost atributa" koja uzima vrednost iz intervala [0,1]. U Fazi Relacionom Modelu kompatibilnost atributa se ne prikazuje, ali u svakoj n-torci vrednost atributa povezana je sa odgovarajućom vrednošću  $C_i$ . Posmatrajmo sledeći primer koji opisuje proširenje klasične relacije baze podataka u GEFRED model:

Ime	Adresa	Godine	Produktivnost	Zarada
Marko	Bulevar revolucije	31	Dobra	Visoka
Aleksandar	Medaković	Srednjih g.	Zadovoljavajuća	100 000
Nenad	Karaburma	Mlad	Loša	90 .000
Saša	Novi Beograd	Star	Odlična	Niska
Vlada	Cerak	Mlad	Dobra	Srednja
Goran	Rakovica	Oko 28	Dobra	125 000
Milan	Žarkovo	Između 30 i 35	Zadovoljavajuća	105 000

Tabela 1. – Generalizovani fazi relacioni model baza podataka



Grafik 1. – Funkcija pripadnosti atributa Godine [13]



Grafik 2. – Funkcija pripadnosti atributa Zarada [13]

Atributi Ime i Adresa sadrže crisp informacije, pri čemu je primarni ključ atribut Ime. Sa druge strane, atributi Godine i Zarada sadrže fazi informacije, a odgovarajuće funkcije pripadnosti za linvističke promenljive iz relacije date su na graficima 1. i 2. Atribut Produktivnost sadrži informacije iz diskretnog domena (4 tipa vrednosti), pa u tom slučaju moramo da definišemo relaciju bliskosti među elementima ovog domena, a to je prikazano u sledećoj tabeli:

S ( $d_i, d_j$ )	Loša	Zadovoljav.	Dobra	Odlična
Loša	1	0.8	0.5	0.1
Zadovoljav.	0.8	1	0.7	0.5
Dobra	0.5	0.7	1	0.8
Odlična	0.1	0.5	0.8	1

Tabela 2. – Relacija bliskosti za atribut Produktivnost [13]

Ovaj teorijski model uključuje sve neophodne elemente za definisanje Fazi Relacionih Modele Baza Podataka (FRDBM).

4. FAZI UPITI

Fazi relacioni model baza podataka predložen od strane Baklsa i Petrija, zasnovan na teoriji sličnosti, predstavlja generalizaciju klasične relacije baze podataka. Sheno i Melton zamenjuju relaciju sličnosti sa relacijom bliskosti i na taj način proširuju fazi relacioni model baza podataka [15-17]. Uvođenjem relacije bliskosti značajno se pojednostavljuje generisanje fazi upita nad fazi modelima baza podataka. Primer: Posmatrajmo FRDBM u kojoj je prisutan fazi atribut BOJA KOSE kao fizička karakteristika neke individue. Relacija bliskosti između elementa domena BOJA KOSE prikazana je u sledećoj tabeli:

Boja Kose	C	TB	K	C	S	SB	P
Crna	1.0	0.8	0.6	0.5	0.4	0.3	0.1
Tamno braon	0.8	1	0.7	0.6	0.6	0.5	0.2
Kestenjasta	0.6	0.7	1	0.8	0.7	0.4	0.3
Crvenkasta	0.5	0.6	0.8	1	0.7	0.5	0.4
Svetla	0.4	0.6	0.7	0.7	1	0.7	0.5
Svetlo braon	0.3	0.5	0.4	0.5	0.7	1	0.8
Plava	0.1	0.2	0.3	0.4	0.5	0.8	1

Tabela 3. – Relacija bliskosti za Atribut BOJA KOSE [17]

Pretpostavimo sada situaciju da su pojedinci iz baze osumnjičeni za podmetanje požara. Jedini dokaz koji je dostupan u istrazi je boja kosa počinioca (svetlo braon), koju je uspeo da zapazi jedan od čuvara. Sada smo u mogućnosti da formulišemo fazi upit:

```
SELECT *
FROM Fizicke karaktersitike
```

WHERE Boja\_Kose = 'SB'  
 WITH LEVEL (Boja\_Kose) = 0.8  
 Pogledajmo sada bazu podataka:

Ime	Boja kose	Grada
Marko	Crna	Veliki
Miloš	Kestenjasta	Srednje razvijen
Petar	Plava	Veoma mali
Milorad	Svetlo braon	Mali
Rade	Tamno braon	Veliki
Momčilo	Crna	Veoma veliki

Tabela 4. – Baza podataka

Na osnovu formulisanog upita, dobijamo iz baze podataka da su osumnjičene dve osobe:

Petar	Plava	Veoma mali
Milorad	Svetlo braon	Mali

Tabela 5. – Rezultat fazi upita

### 5. ZAVISNOST PODATAKA

U logičkom dizajnu baza podataka, ograničenja integriteta imaju ključnu ulogu. Među tim ograničenjima, zavisnost između podataka predstavlja važnu i možda najšire ispitivanu klasu jer pruža direktnu mogućnost normalizacije relacionog modela baze podataka. Zbog toga se posebna pažnja pruža izučavanju funkcionalnih zavisnosti [1]. One dovode u relaciju vrednosti jednoga skupa atributa sa vrednostima drugog skupa atributa. Razne vrste zavisnosti, kao što su funkcionalne zavisnosti, višeznačne zavisnosti, zavisnosti pridruživanja itd. se koriste prilikom projektovanja baza podataka. Druga, vrlo važna pretpostavka u relacionim bazama podataka jeste činjenica da neki podaci u potpunosti određuju druge podatke ili su podaci potpuno nezavisni. Međutim, u mnogim realnim situacijama, postoje podaci koji nisu 100% zavisni, ali ipak zavisnost postoji (npr. "Nivo inteligencije osobe manje ili više određuje uspeh u poslu.") U ovom primeru i stepen uspeha i nivo inteligencije su fazi atributi, pa govorimo o fazi zavisnostima. Ako znamo da je čovek inteligentan, mi znamo da će biti uspešan, ali njegov uspeh nije precizno određen i definisan.

Razlikujemo dve vrste pristupa u kojima se izučava navedena tematika: u prvu grupu spadaju radovi u kojima se definiše pojam fazi funkcionalnih zavisnosti, a drugu grupu sačinjavaju radovi u kojima se primenju koncept funkcionalnih zavisnosti za dekompoziciju podataka i smanjenje redundancije i aproksimizaciju upita u bazi.

Značaj funkcionalnih zavisnosti se ogleda pre svega u mogućnosti da prilikom dizajniranja naše baze, eliminišemo redundanciju koja je svakako prisutna. Za par atributa X i Y kažemo da postoji funkcionalna zavisnost  $X \rightarrow Y$ , ako je  $t[X]=t'[X]$  onda važi i  $t[Y]=t'[Y]$ . tj:

$$t[X]=t'[X] \Rightarrow t[Y]=t'[Y] \quad (5)$$

pri čemu je  $t[X]$  projekcija n-torke (vrste) t na attribute X.

Kod fazi funkcionalnih zavisnosti se ne možemo osloniti na koncept jednakosti Sozat i Yazici, Bosc i ostali u analizi fazi funkcionalnih zavisnosti koriste relaciju sličnosti [2,3]. U tom slučaju kažemo: Ako je  $t[X]$  slično sa  $t'[X]$ ,  $t[Y]$  je takođe slično sa  $t'[Y]$ , i pišemo:

$$X \xrightarrow[F]{\theta} Y \quad (6)$$

Tipičan primer FFD: "Zaposleni sa sličnim iskustvom imaju slične plate". Iskustvo i plata su atributi koji mogu da imaju vrednosti u vidu lingvističkih izraza. Jedan od načina da izrazimo stepen zavisnosti je koristeći prag vrednosti. Npr. "Zaposleni sa sličnim iskustvom imaju slične plate" – stepen zavisnosti možemo da iskažemo sa vrednošću 0.6. "Nivo inteligencije osobe manje ili više određuje uspeh u poslu" – lingvistička snaga 0.6. Ova vrednost se zove snaga, odnosno stepen fazi funkcionalne zavisnosti i označavamo je sa  $\theta$ .

### 6. FUNKCIONALNE ZAVISNOSTI

Funkcionalne zavisnosti predstavljaju semantičko ograničenje nad podacima. Formalno, funkcionalna zavisnost je izraz oblika  $X \rightarrow Y$  i kažemo da je sadržana u relaciji R, ako za sve n-torke t i t'  $\subseteq R$  važi  $t[X]=t'[X] \Rightarrow t[Y]=t'[Y]$  [4].

**Definicija 6.1.** Neka je  $R(A_1, A_2, \dots, A_n)$  relacijska šema na domenima  $D_1, D_2, \dots, D_n$  tako da je domen od  $A_i$   $Dom(A_i) = D_i$  i neka su X i Y podskupovi skupa atributa  $U = \{A_1, A_2, \dots, A_n\}$  tj.  $X, Y \subseteq U$  i neka je r relacija od R,  $r \subseteq D_1 \times \dots \times D_n$ . Kažemo da relacija r zadovoljava funkcionalnu zavisnost  $X \rightarrow Y$ , ako za svake dve n - torke t i t'  $\subseteq r$ , za koje vredi  $t(X) = t'(X)$  implicira da vredi i  $t(Y) = t'(Y)$ .

Navedena definicija označava da kad god suparovi (x,y) i (x,y') elemente relacije R[XY] tada je  $y=y'$ . Ovo je upravo slovkoj razlikuje funkciju od relacije. Ako funkcionalna zavisnost  $X \rightarrow Y$  nepostojondareliacija R[X,Y] može sadržavati više elemenata koji imaju istu vrednost atributa iz X, a različitu vrednost atributa iz Y. Drugo,  $X \rightarrow Y$  je vremenskinepromenljivosvojestvo. Skup n - torko koje opisuju  $R(A_1, \dots, A_n)$  se menja u vremenu, a to ondavaži i za R[X,Y]. Definicija funkcionalne zavisnosti zahteva da sutepromenetakve da u svakom trenutku vremena R[X,Y] predstavlja ne samorelaciju nego funkciju  $R[X] \rightarrow R[Y]$ . Značaj funkcionalnih zavisnosti se ogleda u činjenici da pomoću njih možemo da odredimo primarni ključ relacije i što je još važnije da definišemologički model baze podataka.

#### 6.1. Fazi funkcionalne zavisnosti

Ukoliko funkcionalne zavisnosti uključimo u fazi model baze podataka, tada prethodnu definiciju nije moguće direktno primeniti u tom model jer se ona temelji na konceptu jednakosti. Pošto ne postoji jasan način potvrđivanja kada su dve neprecizne vrednosti jednake, tada definicija funkcionalnih zavisnosti mora da se proširi odnosno generalizuje. Ta prošir-

rena/generalizovana verzija funkcionalnih zavisnosti naziva se fazi funkcionalna zavisnost [18-22]. Postoji više različitih definicija fazi funkcionalnih zavisnosti, koje su dobivene kao rezultat upotrebe fazi logike u klasičnoj funkcionalnoj zavisnosti i sve takve definicije fazi funkcionalnih zavisnosti asocirane su sa datim okvirom u fazi bazama podataka. Zbog toga su jedino primenljive unutar datoga okvira, iako postoje osnovne odnosno generalne osobine i karakteristike koje su potrebne da fazi funkcionalne zavisnosti poseduju.

U opštem slučaju kada vrednosti atributa ne poprimaju samo atomske elemente domena, već i distribuciju mogućnosti, tada stepen  $X \rightarrow Y$  nije nužno potrebno da bude 1, već može da bude u jediničnom intervalu  $[0,1]$ . Tako da se prirodno nameću sledeća pitanja. Kako odrediti  $t(X) = t'(X)$  i  $t(Y) = t'(Y)$  kada su  $t(X)$ ,  $t'(X)$ ,  $t(Y)$  i  $t'(Y)$  sve neprecizne vrednosti definisane kao distribucije mogućnosti. Kao drugo, kako odrediti stepen propozicije Ako je  $t(X) = t'(X)$  tada je  $t(Y) = t'(Y)$  kada su  $t(X) = t'(X)$  i  $t(Y) = t'(Y)$  delimično istiniti sa stepenom pripadnosti iz intervala  $[0,1]$ . I na kraju, kako proceniti stepen od  $X \rightarrow Y$  ako različiti parovi  $n$ -torki daju različite istinitne vrednosti za propoziciju Ako ... tada.

Otuda proizilazi da su ova pitanja povezana sa problemima fazi bliskosti podataka, fazi logičke implikacije i fazi i (and) operatora. Sada definišimo fazi funkcionalne zavisnosti na sledeći način.

**Definicija 6.2.** Neka je  $R(A_1, A_2, \dots, A_n)$  relacijska šema na domenima  $D_1, D_2, \dots, D_n$  sa  $\text{Dom}(A_i) = D_i$  i neka su  $X$  i  $Y$  podskupovi skupa atributa  $U = \{A_1, A_2, \dots, A_n\}$  tj.  $X, Y \subseteq U$  i neka je  $r$  relacija od  $R$ ,  $r \subseteq \Pi(D_1) \times \dots \times \Pi(D_n)$ , gde je  $\Pi(D_i) = \{\pi \mid \pi \text{ je moguća distribucija od } A_i \text{ na } D_i, i=1,2,\dots,n\}$ . Kažemo da  $X$  fazi funkcionalno određuje  $Y$  sa stepenom  $\theta$ , označeno kao  $X \xrightarrow{\theta} Y$  ako i samo ako za  $\forall r \in R$ :

$$\min I(t(X) =_c t'(X), t(Y) =_c t'(Y)) \geq \theta, t, t' \in r. \quad (7)$$

gde je  $\theta \in [0,1] =_c [0,1] \times [0,1] \rightarrow [0,1]$  mera bliskosti i  $I: [0,1] \times [0,1] \rightarrow [0,1]$  je fazi operator implikacije.

## 7. REDUNDANTNOST PODATAKA

Redundancija podataka je jedan od primarnih problema u klasičnim bazama podataka. Kao prvo, redundancija označava postojanje duplih  $n$ -torki, kojima nije dozvoljeno da se pojavljuju u pojedinačnim tabelama. To znači da neće postojati  $n$ -torke  $t$  i  $t'$  u  $R$ , takve da su  $t$  i  $t'$  identične u odnosu na odgovarajuće vrednosti atributa.

**Definicija 7.1.** Za  $n$ -torku  $t = (t(A_1), t(A_2), \dots, t(A_n))$  i  $n$ -torku  $t' = (t'(A_1), t'(A_2), \dots, t'(A_n))$  u relaciji  $R$ , kažemo da su jednake ako i samo ako je  $t(A_i) = t'(A_i)$ ,  $i = 1, 2, \dots, n$ .

Alternativno, redundancija  $n$ -torki može se takođe definisati na osnovu primarnog ključa  $K$ , na osnovu koncepta zavisnosti podataka. Primarni ključ  $K$  relacije  $R$  na relacionoj šemi  $R(U)$  je podskup od  $U$  takav da za jednake vrednosti  $n$ -torki na  $K$  impliciraju jednake vrednosti  $n$ -torki za svaki atribut  $A_i \in U$ , pa su tada dve  $n$ -torke  $t$  i  $t'$  redundantne ako

i samo ako su njihove odgovarajuće vrednosti  $n$ -torki na  $K$  jednake. Drugim rečima, budući da iz  $t(K) = t'(K)$  implicira  $t(A_i) = t'(A_i)$ ,  $n$ -torke  $t$  i  $t'$  su redundantne ako i samo ako je  $t(K) = t'(K)$ .

Osim toga, redundancija takođe znači dupliranje vrednosti u odnosu na neke attribute u skupu. U fazi relacionim bazama podataka ne postoji jasan način potvrđivanja da su dve vrednosti atributa jednake ili različite, već se posmatra kada su one približno jednake. Zbog toga klasična razmatranja ne važe i problem redundancije nije više trivijalan i zahteva dodatna razmatranja.

Prvi problem nastaje zbog toga što vrednost nekog atributa  $n$ -torke može biti podskup domena ili može biti i moguća distribucija. Na primer,  $\text{Visina} = \{170, 180, 190\}$  ili  $\text{Visina} = \{0.3/150, 0.9/175, 0.1/200\}$ .

Drugi problem nastaje kada se neke fazi relacije definišu na datim domenima, na primer relacijasličnosti ili relacija bliskosti. Postoje situacije kada je potrebno uvesti takve relacije nad domenima. U klasičnom slučaju na domenu  $D = \{\text{crn, plav, braon, siv, zelen, crven, ljubičast, žut, bel}\}$  atribut *Boja* na  $D$  može uzeti bilo koju vrednost iz domena  $D$ . Ali, kada kažemo da je nešto tamno? Da li je jasno tada određena granica? Dalje, ako je dozvoljeno da vrednosti atributa budu podskupovi domena  $D$ , sa ciljem da predstavimo neodređenosti kao  $\{\text{crn, zelen}\}$ ,  $\{\text{braon, crven}\}$ ,  $\{\text{crn, siv, plav}\}$  itd. kako onda odrediti jednakost ili sličnost za neki par vrednosti.

Na kraju veći problemi nastaju kada su vrednosti atributa distribucije mogućnosti uzete zajedno sa nekom fazi relacijom definisanom na domenu atributa.

## 8. ZAKLJUČAK

Fazi relacije baze podataka na prirodan način proširuju funkciju klasičnih fazi baza znanja, što obezbeđuje viši nivo adaptacije fazi sistema, kao jedne od osnovnih osobina (pored sistema planiranja, učenja, predikcije, sistema za pretragu znanja, robota) inteligentnih sistema. U oblasti veštačke inteligencije, fazi skupovi i fazi logika, izučavaju se sa aspekta dizajniranja fazi relacionih modela baza podataka, koji takođe pripadaju oblasti inteligentnih sistema. Teorija fazi skupova i fazi logika predstavljaju moćan alat za manipulisanje nepreciznim i neodređenim informacijama. U samoj biti ovih modela podataka je činjenica da postoji veliki broj aktivnih istraživačkih oblasti koje direktno uključuju ili koriste ove baze znanja. Pitanje nepreciznosti podataka i njihove reprezentacije ja jako zastupljena tema i vrlo važna u raznim oblastima. kao što su: geografski informacioni sistemi (GIS) i reprezentacija prostornih podataka, data mining tehnike, pretraživanje fazi informacija, statističke baze podataka, biznis registri, prikupljanje i selekcija podataka... Razmatranje sa lingvističke tačke gledišta je veoma važan aspekt koji se koristi u ovim sistemima.

U ovom radu opisan je i pristup projektovanju fazi relacionih modela baza podataka jer projektovanje predstavlja osnovni poligon za primenu rezultata i analiza iz oblasti funkcionalnih zavisnosti, tj. funkcionalne i fazi funkcionalne zavi-

snosti igraju veoma važnu ulogu u dizajniranju baza podataka. Problemi redundatnosti podataka i pojave anomalija u bazama podataka navode nas na zaključak da baza nije dobro projektovana. Upravo iz tog razloga, data je kraća analiza i smernice koje obezbeđuju teorijsku osnovu u dizajniranju istih.

## 9. LITERATURA

- [1] Z. M. Ma, L. Yan, "A literature overview of fuzzy database models", *Information Science and Engineering*, 2008, 172-180.
- [2] D. Dubois, H. Prade, "The three semantics of fuzzy sets", *Fuzzy Sets and Systems*, 90 (1997) 141-150.
- [3] P. Bosc, D. Kraft, F. Petry, "Fuzzy sets in database and information systems: status and opportunities", *Fuzzy Sets and Systems*, 156 (2005) 418-426.
- [4] D. Kraft, F. Petry, "Fuzzy information systems: managing uncertainty in databases and information retrieval systems", *Fuzzy Sets and Systems*, 90 (1997) 183-191.
- [5] T. Radecki, "A theoretical background for applying fuzzy set theory in information retrieval", *Fuzzy Sets and Systems*, 10 (1983) 169-183.
- [6] E. A. Rundensteiner, L. W. Hawkes, Wyllis Bandler, "On nearness measures in fuzzy relational data models", *International Journal of Approximate Reasoning*, 3 (1989) 267-298.
- [7] P. Bosc, M. Galibourg, "Indexing principles for a fuzzy database", *Information Systems*, 14 (1989) 493-499.
- [8] D. Chiang, L. R. Chow, N. Hsien, "Fuzzy information in extended fuzzy relational databases", *Fuzzy Sets and Systems*, 92 (1997) 1-20.
- [9] B. P. Buckles, F. E. Petry, "Extending the fuzzy database with fuzzy numbers", *Information Sciences*, 2(1984) 145-15.
- [10] J. Kacprzyk, B. P. Buckles, F. E. Petry, "Fuzzy information and database systems", *Fuzzy Sets and Systems*, 38 (1990) 133-135.
- [11] B. P. Buckles, F. E. Petry, "A fuzzy representation of data for relational databases", *Fuzzy Sets and Systems*, 7 (1982) 213-226.
- [12] G. Chen, *Fuzzy logic in data modeling: semantics, constraints and database design*, Kluwer Academic Publisher, 1998.
- [13] J.M. Medina, M. A. Vila, J. C. Cubero, O. Pons, "Towards the implementation of a generalized fuzzy relational database model", *Fuzzy Sets and Systems*, 75(1995) 273-289.
- [14] J. M. Medina, O. Pons, M. A. Vila, "GEFRED: A generalized model of Fuzzy Relational Databases", *Information Sciences*, 76 (1994) 87-109.
- [15] A. Melton, S. Shenoï "Fazi relations and fazi relational databases", *Computers & Mathematics with Applications*, 21 (1991), 129-138
- [16] S. Shenoï, A. Melton, "An extended version of the fazi relational database model", *Information Sciences*, 52 (1990), 35-52,
- [17] S. Shenoï, A. Melton, "Proximity relations in the fazi relational database model", *Fuzzy Sets and Systems*, 100 (1999), 51-62
- [18] M. Sözat, A. Yazici, "A complete axiomatization for fuzzy functional and multivalued dependencies in fuzzy database relations", *Fuzzy Sets and Systems*, 117 (2001) 161-181
- [19] P. C. Saxena, B. K. Tyagi, "Fuzzy functional dependencies and independencies in extended fuzzy relational database models", *Fuzzy Sets and Systems*, 69 (1995) 5-89
- [20] P. Bosc, O. Pivert, "On the impact of regular functional dependencies when moving to a possibilistic database framework", *Fuzzy Sets and Systems*, 140 (2003) 207-227.
- [21] W. Liu, "Fuzzy data dependencies and implication of fuzzy data dependencies", *Fuzzy Sets and Systems*, 92 (1997) 341-348.
- [22] S. Shenoï, A. Melton, L.T. Fan, "Functional dependencies and normal forms in the fuzzy relational database model", *Information Sciences*, 60 (1992) 1-28.



**Miljan Vučetić**, Fakultet organizacionih nauka  
 Kontakt: miljanvucetic@gmail.com  
 Oblast intresovanja: informacioni sistemi, elektronsko poslovanje, fazi logika, veštacka inteligencija.

