Ontologijama podržano obogaćivanje sheme relacijskoga skladišta podataka višedimenzijskim konceptima iz dokumentno-orijentiranoga podatkovnoga izvora

Marina Ptiček, mag. ing. comp.
(marina.pticek@fer.hr)
mentor: prof. dr. sc. Boris Vrdoljak,
Sveučilište u Zagrebu Fakultet elektrotehnike i računarstva

1. Uvod

Sve popularnije dokumentno-orijentirane baze podataka obilježava **polustrukturiranost** i **odsutnost eksplicitne definicije sheme**, te moguća shematska nekonzistentnost podataka.

Prilikom multidimenzionalnog modeliranja skladišta podataka **analiziraju se i shematski integriraju** izvori podataka. Odsutnost sheme u dokumentno-orijentiranom izvoru podataka **otežava dizajneru integraciju** dokumentno-orijentiranog izvora podataka u postojeće skladište podataka.

3. Metodologija

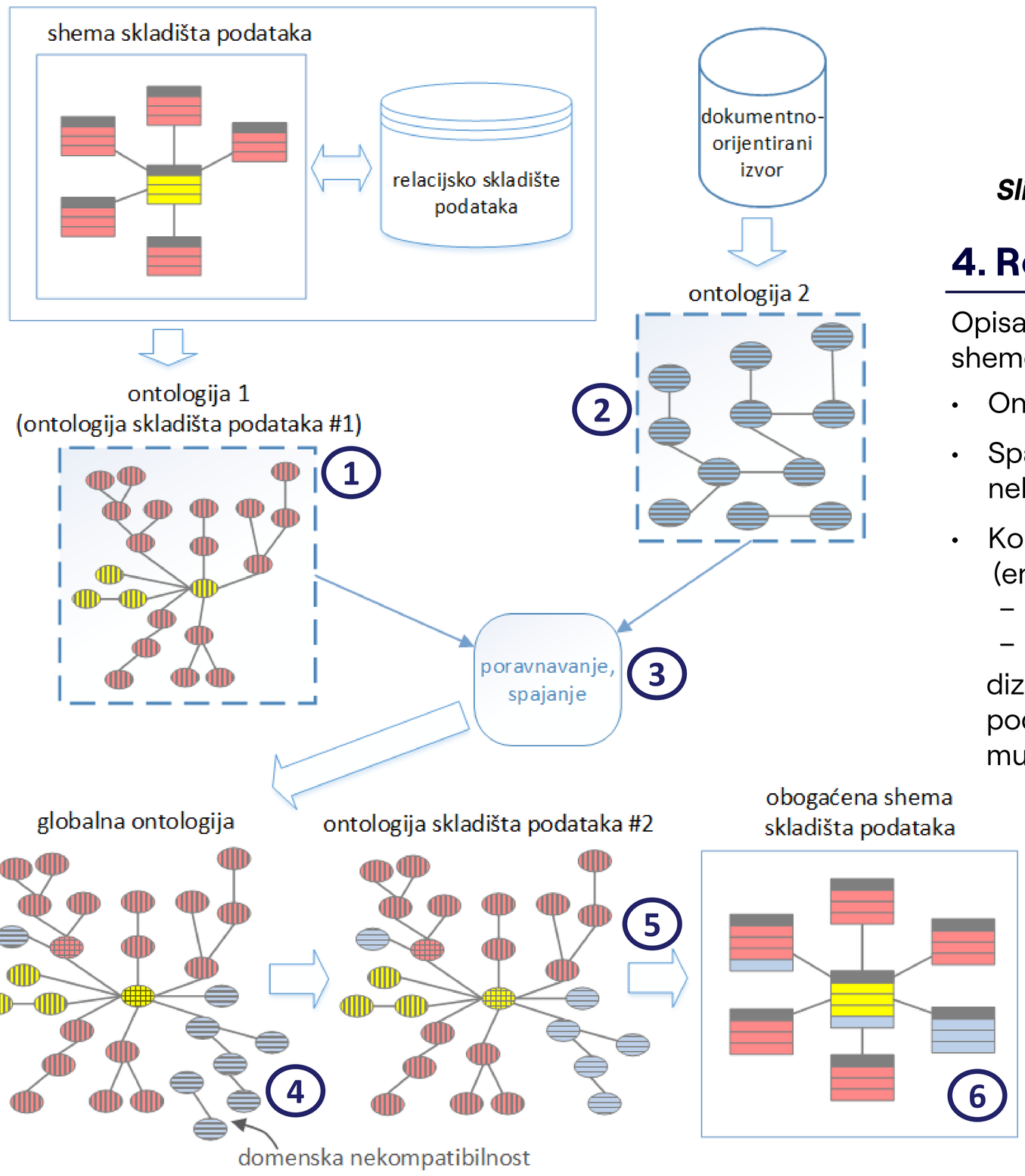
Sheme obiju baza se mogu **opisati na jednak način**, pomoću ontologija semantičkog weba (Slika 2).

Ontologija 1 opisuje skladište podataka **shematski** (multidimenzionalni model) te ujedno i **domenski** (prirodna domena *tematike* skladišta podataka). (1)

Polustrukturirani dokumentno-orijentirani podatci sastoje se od parova *ključ-vrijednost*, stoga bi moglo vrijediti sljedeće:

1. dokument sadrži **implicitnu shemu** u svom skupu *ključeva*
2. izvučena implicitna shema se može **opisati ontologijom**
3. dobivena *ontologija 2* opisuje **domenu** novog izvora. (2)

Navedene ontologije se mogu **usporediti i spojiti** gotovim algoritmom za spajanje i poravnavanje ontologija u globalnu ontologiju. (3)



Slika 2. Predloženi postupak

2. Opis problema

Problem je:

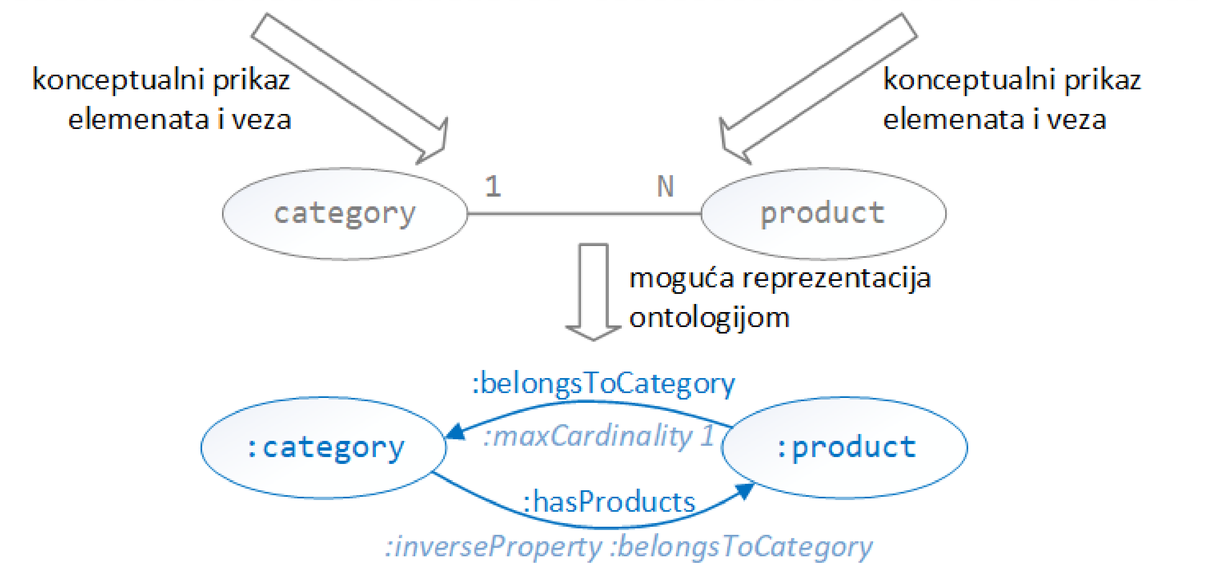
- **kako shematski opisati** polustrukturirani, shematski nekonzistentni dokumentno-orijentirani izvor podataka? (Slika 1. primjer 1 i 2)
- **kako usporediti i integrirati shemu** dobivenu u prethodnom koraku s uređenom shemom relacijskog skladišta podataka?

primjer 1.

```
product:
[{ prodID: 1000,
   prodName: "Ice tea",
   category: { catID: 298,
               catName: "drinks"
             }
 },
 { prodID: 2000,
   prodName: "Orange juice",
   category: { catID: 298,
               catName: "drinks"
             }
  },
  ...
]
```

primjer 2.

```
categories:
 [{ catID: 298,
    catName: "drinks",
    products:[
              {
               prodID: 1000,
               prodName: "Ice tea"
              },
              {
               prodID: 2000,
               prodName: "Juice"
              }
             ]
  },
  ...]
```



Slika 1. Reprezentacija sheme dokumenata ontologijom

4. Rezultati

Opisani postupak **djelomično automatizira** proces obogaćivanja sheme relacijskog skladišta podataka (***olakšava* posao dizajneru**):

- Ontologija **sažeto i čitljivo** opisuje dokumentno-orijentirani izvor.
- Spajanjem ontologija **automatski se ustanovljavaju i odbacuju** nekompatibilni koncepti iz novog izvora. (4)
- Korištenjem **algoritama semantičkog zaključivanja** (5) (eng. *semantic reasoning*) nad:
  - **ontološkim shematskim opisom** skladišta podataka, te
  - **domenskom ontologijom** novog izvora,

  dizajneru se može **predložiti obogaćivanje sheme** skladišta podataka prijedlogom smještaja dijelova sheme novog izvora u multidimenzionalni model. (6)

5. Zaključak

- Ontologije su pogodne za reprezentaciju sheme i sadržaja neke domene, čitljive i čovjeku i računalu.
- Spajanje ontologija i zaključivanje iz ontologija odvija se uz visok stupanj automatizacije.
- Predloženi postupak bi mogao ***olakšati i djelomično automatizirati*** obogaćivanje sheme postojećeg relacijskog skladišta podataka multidimenzionalnim konceptima iz dokumentno-orijentiranog izvora.

Dan doktorata Fakulteta elektrotehnike i računarstva, 6. srpnja 2021.