

Pregled sustava temeljenih na znanju za potporu procesu modeliranja podataka

Sabrina Šuman

Veleučilište u Rijeci

Doktorand Odjela Informatike Sveučilišta u Rijeci

ssuman@veleri.hr

Sažetak - Modeliranje podataka složen je proces koji ovisi o znanju i iskustvu dizajnera koji ga provodi. Kvaliteta izrađenog modela značajno utječe na kvalitetu daljnjih faza razvoja informacijskog sustava. Ovim radom je dan pregled i opis šireg i užeg područja istraživanja koje obuhvaća: proces modeliranja podataka, metodu entiteta i veze (EV), aktere procesa modeliranja podataka te sustave, alate i metode za potporu procesu modeliranja podataka. Pored toga su identificirani i problemi koji se javljaju tijekom faza razvoja informacijskog sustava. Oni su ujedno i motivacija provođenja istraživanja koje za cilj ima razvoj sustava temeljenoga na znanju za potporu procesu modeliranja podataka. Identificirane su i mogućnost primjene teorije formalnih jezika (posebno prevodenja) u procesu konceptualnoga modeliranja što je i pravac kojim će ići istraživanje razvoja novoga sustava. U zaključku rada opisane su temeljne osobine koje bi novi sustav koji se razvija trebao posjedovati, a koje su proizašle iz provedene analize postojećih sustava, alata i metoda za potporu modeliranju podataka.

Ključne riječi - modeliranje podataka, metoda entiteta i veza, sustav temeljen na znanju, alati i metode za potporu modeliranju podataka

I. UVOD

Tijekom razvoja informacijskog sustava moguće je identificirati sljedeće njegove razvojne faze: strateško planiranje, analiza, oblikovanje (dizajn), implementacija, uvođenje, održavanje i evaluacija. Područje istraživanja autorice je modeliranje podataka koje se događa u fazi oblikovanja razvoja informacijskog sustava.

U sklopu faza oblikovanja i implementacije informacijskog sustava razvija se struktura baze podataka koja služi zadovoljenju informacijskih potreba svojih korisnika. Taj razvoj uključuje konceptualni i logički dizajn baze podataka baziran na sakupljenim i analiziranim zahtjevima, fizički dizajn baze podataka i implementaciju baze podataka [1]. Upravo je proces konceptualnog i logičkog dizajna baze podataka (konceptualno i logičko modeliranje podataka) koji se odvija unutar faze dizajna informacijskog sustava uža domena istraživanja. Cilj ovoga rada je dati terminološki pregled područja interesa istraživanja te pregled nekih sustava temeljenih na znanju (engl. Knowledge based system-KBS) za potporu modeliranju podataka.

Konceptualni dizajn baze podataka (konceptualno modeliranje podataka) je aktivnost koja se izvodi u ranim fazama razvoja informacijskog sustava te je usmjerena na stvaranje konceptualne sheme koja predstavlja model podataka promatrane realnosti. Jednom odobrena od

strane korisnika, konceptualna se shema prevodi u odgovarajuće konstrukte baze podataka.

Cilj konceptualnog dizajna je iz specifikacije korisničkih zahtjeva koja opisuje promatranu realnost, uočiti temeljne koncepte i njihove međusobne odnose, te ih prikazati u obliku konceptualne sheme [2].

Upravo stvaranje konceptualne sheme koja je potpuna, jednostavna za korištenje, razumljiva i semantički ispravna, predstavlja problem koji se rješava primjenom različitih metoda u konceptualnom dizajniranju.

Svaka metoda modeliranja koristi mali skup konstrukata koji definiraju vokabular metode i sintaksnu pravila njihova povezivanja [3]. Konstrukti su grafički ili negrafički izraženi koncepti, specifični za danu metodu modeliranja, čija je svrha organiziranje i predstavljanje znanja iz interesne domene. Tako su u metodi entiteta i veza (metoda EV, engl. Entity relationship method-ER method), glavni konstrukti entiteta, veze i atributi [4]. Inače je metoda EV sa svojim varijacijama i proširenjima - obično pod nazivom proširena EV (engl. extended ER ili EER) - jedna od češće korištenih metoda u konceptualnom dizajniranju [5].

Glavni cilj istraživanja autorice je razvoj sustava temeljenog na znanju za potporu procesu modeliranja podataka koji se odvija unutar faze oblikovanja informacijskog sustava. Jedna od aktivnosti unutar toga istraživanja je i pregled šireg i užeg interesnog područja čiji su rezultati prikazani ovim radom. Osim prikaza temeljnoga terminološkog sustava interesnog područja, prikazani su i neki sustavi temeljeni na znanju kao potpora procesu modeliranja podataka, te neke metode koje olakšavaju izradu modela podataka. Do njih se došlo pretragom dostupnih radova unutar baza podataka EBSCO, ScienceDirect i Google Scholar, te društvene mreže Research Gate. Kao uvjet pretrage definirani su izrazi „Knowledge based systems for data modelling“, „formalization of data modelling“, „Entity relationship method“. Dobiveni rezultati su se dodatno filtrirali prema kriteriju relevantnosti s novim sustavom temeljenim na znanju koji se istraživanjem želi razviti. Kriterij relevantnosti je uključivao: metode interakcije s korisnikom, zaključivanje temeljeno na prošlim slučajevima, primjeni teorije formalnih jezika (gramatika, sintakсна analiza, parser), te pravila prevodenja u konstrukte modela entiteta i veza. Pretraživanje je rezultiralo s ukupno 12 sustava koji su potpora modeliranju podataka. Neki od njih imaju osobine sustava temeljenih na znanju dok su drugi na

razini pomoćnih alata. Pored navedenih sustava, pretraga je rezultirala i nizom metoda koje olakšavaju izradu modela podataka.

Uz uvod, rad se sastoji iz dodatnih pet dijelova. U drugom dijelu rada se opisuje motivacija koja je inicirala istraživanje. Treći dio rada prikazuje proces modeliranja podataka i metodu entiteta i veze. Četvrti dio rada opisuje ključne aktere u procesu modeliranja podataka koji će potencijalno biti korisnici sustava temeljenoga na znanju za potporu modeliranju podataka koji se razvija. Peti dio rada prikazuje postojeće sustave za potporu modeliranju podataka kao i metode koje olakšavaju proces modeliranja. Članak završava zaključkom u kojemu se navode daljnji pravci istraživanja.

II. MOTIVACIJA ZA ISTRAŽIVANJE

Motivacija autorice za istraživanje područja modeliranja podataka proizlazi iz niza problema koji se pojavljuju u fazama razvoja informacijskog sustava. Neki od tih problema, a koji se odnose na faze analize i oblikovanja su:

- složenost procesa prevođenja modela procesa u model podataka
- nedostatak znanja potrebnog za izradu modela podataka
- nepostojeća ili nedovoljna formalizacija znanja o izradi modela podataka neophodnog u računalno podržanom sustavu
- mogućnost pogrešaka u izradi modela podataka
- niska učinkovitost (vrijeme, novac i ostali resursi) prevođenja modela procesa u model podataka
- moguća niska djelotvornost izrađenog modela podataka u slučaju promjene poslovnih procesa

Jedan od prvih rezultata procesa modeliranja podataka jeste konceptualni model podataka nastao konceptualnim modeliranjem. Konceptualno modeliranje fokusira se na konceptualne aspekte neke domene i za razliku od modeliranja baze podataka izuzima razmatranja vezana za oblikovanje i implementaciju. Autori u [6] naglašavaju aktualnost istraživanja vezanih za evaluaciju gramatike konceptualnog modeliranja, kao npr. gramatika modeliranja entiteta i veza ili modeliranja poslovnih procesa [4]. U [6] smatraju da treba povećati utjecaj gramatike na stvaranje kvalitetnih skripti konceptualnog modeliranja. Autori skriptom nazivaju tekstualni opis konceptualnog modela koji je generiran gramatikom nekoga jezika (npr. skripte koje generira gramatika entiteta i veza su dijagrami entiteta i veza). Pojmovi koji se pojavljuju u ovom pravcu istraživanja su: gramatika konceptualnog modeliranja, skripte konceptualnog modeliranja te jezik konceptualnog modeliranja.

Upravo primjena procesa prevođenja iz teorije formalnih jezika, koja uključuje gore spomenute pojmove, bit će osnova razvoja sustava temeljenoga na znanju za potporu modeliranju podataka (engl. Knowledge Based System for Data Modeling – KBSforDM).

III. MODELIRANJE PODATAKA

Tijekom faze analize informacijskog sustava korisnički zahtjevi se specificiraju (otkrivaju i zapisuju). U analizi se doznaje što treba poslovnoj organizaciji. Nakon faze analize slijedi faza oblikovanja informacijskog sustava unutar koje se, pored ostaloga, provodi i modeliranje podataka. Oblikovanje informacijskog sustava je proces pomoću kojega se na osnovi zahtjeva iz faze analize definiraju osnovni dijelovi informacijskog sustava i prikazuju detaljnije veze među njima na način neovisan o implementaciji. Faza oblikovanja sastoji se od: oblikovanja odnosa među podacima, oblikovanja baze podataka i oblikovanja arhitekture programskog proizvoda. Prva dva oblikovanja nazivaju se i modeliranje podataka. Metoda za modeliranje podataka je definirani postupak nalaženja i prikazivanja informacijskih objekata i njihova međusobnog odnosa. [7]

Rezultat primjene neke metode (npr. metoda entiteta i veza) u procesu modeliranja podataka je model podataka (npr. model entiteta i veza). Model podataka konstruiraju dizajneri informacijskih sustava, a potvrdu njegove ispravnosti daju krajnji korisnici. On služi za prikaz podatkovnih dijelova sustava i njihovih odnosa i osnova je za izradu baze podataka i definiranje arhitekture programskog proizvoda. Model podataka sadrži tri važna koncepta: entitete, atribute entiteta (svojstva, karakteristike) i veze među entitetima. [7]

Model podataka opisuje statičnu informacijsku strukturu preko entiteta i njihovih veza. Konceptualni, logički i fizički model tri su evolucijske etape modeliranja podataka od kojih svaka pojedina faza nadograđuje prethodni model podataka novim detaljima, optimizirajući ga za implementaciju. Ova struktura podataka često je predstavljena grafički s dijagramom entiteta i veza (DEV) ili UML dijagramom klasa. [8]

A. Metoda entiteta i veze

Metoda entiteta i veze (EV metoda) prikazuje međusobno povezane podatke promatranog sustava. Ona grafički prikazuje skupine podataka. Semantički je bogata i raspolaže ljudski bliskim konceptima kojima se prikazuju entiteta i njihove veze. Metoda prirodno opisuje poslovnu organizaciju, a njezin rezultat - dijagram modela podataka - lagan je za razumijevanje te omogućuje komunikaciju projektanta i korisnika. Metoda EV se koristi u mnogim metodologijama (npr. CASE*Method, MIRIS, SSADM, IEM). Dijagram modela podataka, odnosno konceptualna shema baze podataka, načinjena metodom EV naziva se dijagram entiteta i veza – DEV (engl. Entity-Relationship Diagram-ERD) [9].

Dijagram entiteta i veza prvotno je osmišljen kao sredstvo brzog dobivanja ideje o strukturi baze podataka. Koristi se za modeliranje podataka sustava i planiranje i oblikovanje njegove baze podataka. Budući da on implicira procese koji djeluju nad podacima ali ih ne pojašnjava, zajedno s dijagramima toka podataka (rezultatom modeliranja procesa), EV model daje projektantu mogućnost alternativnog logičkog pogleda na sustav [10].

Autori literature vezane uz područje razvoja softvera i oblikovanja baze podataka i CASE alata koriste različite DEV notacije, što stvara poteškoće kod usvajanja znanja i vještina potrebnih projektantima početnicima. Iz istog je razloga otežan prijenos i razumljivost pojedinih DEV-a između autora, a moguće i članova dislociranih timova. Modeliranje nekom metodom koristeći postojeći CASE alat ovisno je o znanju i vještini projektanta, što izravno utječe na probleme istraživanja identificirane u uvodu ovoga rada.

Istraživanjem se u [11], nakon analize 10 metoda s različitim konceptima i notacijama, ustanovilo da se DEV notacije razlikuju po sljedećim točkama:

1. dozvoljavaju li se n-arne veze ili ne
2. način na koji se predstavljaju ograničenja entiteta u vezi - kardinalnost i opcionalnost; notacije za minimalan i maksimalan broj pojava entiteta u nekoj vezi
3. mjesto gdje se označavaju ograničenja (specificiranje ograničenja kardinalnosti i opcionalnost tako da se ona čita od suprotnoga kraja veze – engl. Look Across – ili prema suprotnom kraju veze – engl. Look Here)
4. postoji li atribut veze
5. postoji li vanjski ključ na DEV-u
6. kako se prikazuju preklapajući i disjunktni podtipovi entiteta
7. kako se prikazuje potpuna ili djelomična specijalizacija

Istraživanje opisano u [11] je pokazalo niz jednostavnijih i složenijih metoda modeliranja podataka. To je poslužilo autorici u odabiru temeljne metode i konstrukata za modeliranje podataka koji će se koristiti u sustavu temeljenom na znanju koji se razvija: entiteti, veze, atributi i kardinalnost. Izabrana metoda i konstrukti opisani su u [12], osim opcionalnosti koja u originalnoj verziji ER modela nije bila navedena.

IV. KLJUČNI AKTERI U PROCESU MODELIRANJA PODATAKA

Planer, analitičar, dizajner i izvođač su stručnjaci koji s različitih perspektiva pristupaju problemu podataka informacijskog sustava. Dok planer identificira glavne podsustave poslovne organizacije i njihove veze, analitičar detaljno (i u suradnji s korisnikom) analizira poslovanje utvrđujući koji se poslovni procesi izvode u uočenim podsustavima, koje ulazne/izlazne podatke koriste, te što je okolina analiziranog sustava. Prema ulazno/izlaznim podacima koje koriste poslovni procesi, dizajner utvrđuje entitete (objekte), njihove attribute i strukturu potrebnu za opis podataka koji kolaju poslovnim sustavom. Izvođač (u slučaju podataka je to administrator baze podataka) razrađuje strukturu zapisa podataka u prostoru baze podataka [13].

U početnim fazama razvoja informacijskog sustava, planer i analitičar utvrđuju poslovne podsustave, njihove poslovne procese i tokove podataka kojima su povezani. Dizajner utvrđuje kako će podaci iz tokova podataka biti logički pohranjeni, dok njihovu fizičku pohranu razrađuje izvođač. Ovo pokazuje da se kompleksan problem izrade baze podataka informacijskog sustava raščlanjuje na manje i logički odvojene probleme (problem

prepoznavanja podataka, problem logičkog rasporeda podataka, problem fizičke pohrane podataka) koje je lakše pojedinačno rješavati [13].

Upravo je djelovanje aktera koji su usko povezani s procesom modeliranja podataka potrebno detaljno analizirati jer se očekuje da će oni biti glavni korisnici sustava temeljenoga na znanju za potporu modeliranju podataka koji se razvija.

V. PREGLED SUSTAVA, ALATA I METODA ZA POTPORU PROCESU MODELIRANJA PODATAKA

U ovom poglavlju su prikazani neki postojeći sustavi temeljeni na znanju za potporu modeliranju podataka. Posebno su istaknute sljedeće njihove karakteristike: vrsta ulaza, način ekstrahiranja entiteta, atributa i veza, tip korisnika kojima je sustav namijenjen te uključuje li sustav konzultiranje prošlih slučajeva (engl. case based reasoning). Pored sustava temeljenih na znanju, ovo poglavlje donosi i opis nekih alata i metoda za potporu procesu modeliranja podataka.

U [14] se daje detaljna analiza mnogih alata i sustava za pomoć u dizajnu baze podataka u rasponu od 1982. do 1998. Kao jedna od smjernica za daljnja istraživanja navodi se potreba za razvojem novih načina prikaza znanja, zaključivanja i tehnika strojnoga učenja koji će biti pogodni za razvoj inteligentnih alata za oblikovanje baza podataka.

Sustav CODASYS je razvijen s restriktivnim korisničkim sučeljem i sučeljem koje vodi korisnika (engl. restrictive and guidance interface). Dok je kod restriktivnog sučelja korisniku ograničen broj putanja kojima se može kretati u procesu izrade modela podataka, kod sučelja koje ga vodi, sustav savjetuje korisnika u procesu izrade modela podataka čime broj putanja nije ograničen. Empirijskom studijom primjene ovoga sustava se pokazalo značajno poboljšanje kod dizajnera početnika u izradi modela podataka (više kod sučelja koje vodi korisnika, manje kod restriktivnog sučelja). U provedenoj studiji su svi ispitanici prije pokusa primili istu količinu treninga vezanoga za modeliranje podataka. Ti rezultati ukazuju na korisnost razvoja sustava temeljenoga na znanju za potporu modeliranju podataka. Budući su veze ključan aspekt modeliranja podataka, CODASYS daje najveću potporu tom području. Posebno i stoga što su dizajneri početnici sposobni dobro identificirati i modelirati entitete ali imaju poteškoća s modeliranjem veza među njima. [15], [16]

Autori u [17] u predstavljanju sustava temeljena na znanju za konceptualno dizajniranje naziva CABSYYDD tvrde da je sustav koji se temelji na prošlim slučajevima modeliranja učinkovitiji i produktivniji od sustava koji pomaže da se iz početka konstruira model podataka. Navode da dotadašnji sustavi za potporu dizajnu baze podataka pomažu dizajneru ali ne sadrže spremišta prošlih modela. Autori CABSYYDD-a su vođeni idejom da će dobro osmišljen alat za potporu dizajnu baze podataka koji se temelji na prošlim slučajevima moći: pohraniti i dohvatiti prethodne slučajeve dizajniranja modela podataka, pomoći dizajneru u pronalaženju pohranjenog slučaja koji je sličan njegovom problemu i pomoći dizajneru u prilagodbi rješenja iz sličnih pohranjenih

slučajeva a kako bi se došlo do rješenja novog problema. CABSYYDD modeliranje podataka temelji na proširenom modelu entiteta i veza iz [1], uz ograničavanje veza na binarne. Sastavljen je iz dva sustava gdje prvi - SYDD pomaže u oblikovanju modela podataka koji se stvara prvi puta, odnosno u slučaju kada nisu poznati nikakvi njemu slični modeli u bazi prošlih slučajeva. Drugi dio sustava naziva CABSYYDD je zadužen za modeliranje podataka koje se temelji na nekim postojećim prošlim slučajevima. Ovaj sustav poziva SYDD onda kada za neke dijelove modela podataka nema odgovarajućih rješenja u bazi prošlih slučajeva. Dakle SYDD posjeduje znanje potrebno za oblikovanje baze podataka i nema mogućnost učenja. S druge strane CABSYYDD posjeduje mehanizam zaključivanja koji se temelji na prošlim slučajevima modeliranja podataka, bazu prošlih slučajeva te funkcionalnost učenja iz prijašnjih slučajeva [17].

U [18] su opisani problemi vezani za ekstrahiranje entiteta iz poslovnih specifikacija i potrebu razvoja formalnih metodologija ekstrahiranja. Autori predlažu novu metodologiju ekstrahiranja entiteta koja osposobljava zaposlenika iz danog poslovnog okruženja da sam izvede konceptualno modeliranje podataka bez konzultiranja dizajnera eksperta. U procesu ekstrahiranja entiteta postoji pretprocesiranje za otklanjanje semantičke redundancije budući da inicijalni poslovni opisi pisani od strane zaposlenika potencijalno sadrže više redundantnih opisa istih podataka. Taj proces izvodi se alatom naziva Semantic Association Model - SAM. Nakon faze pretprocesiranja, slijedi primjena formalne metodologije imena REEM za ekstrahiranje entiteta iz danih poslovnih opisa. Očekivani rezultat primjene metodologije REEM je odgovarajući DEV. Razvijen je i sustav za automatiziranje dizajna baze podataka naziva BizData a koji se temelji na REEM metodologiji. Jedini ulaz potreban za ovaj sustav je poslovni opis u kojemu se entiteti i atributi pojavljuju zajedno. Sustav provodi proces modeliranja podataka samo na temelju informacija dostupnih u poslovnome opisu - ne koristi prošla znanja. Isti autori u drugom članku zaključuju da pretprocesiranje koje zahtjeva metodologija REEM (uklanjanje semantičkih redundancija iz poslovnoga opisa) opterećuje korisnika s previše pitanja te se umanjuje korisnikova percepcija funkcionalnosti i korisnosti samoga sustava. Također, autori zaključuju da se posebna pažnja treba usmjeriti na format ulaza u sustav (poslovni opis). [19]

Jedan od alata za automatsku izradu DEV-a koji omogućava zaposleniku iz neke poslovne domene, koji nema iskustva u modeliranju podataka, izradu DEV-a prema poslovnom opisu je ABCM (engl. Association-based Conceptual Model) [19]. Glavna motivacija autora za stvaranje ovakvog sustava je mogućnost da dizajner baze podataka pogrešno interpretira semantiku poslovnog opisa koju je sastavio zaposlenik iz neke poslovne domene. Autori su analizirali i neke druge sustave za automatiziranje dizajniranja modela podataka te zaključili da se oni slabo koriste iz tri razloga: od zaposlenika iz neke poslovne domene zahtijevaju previše znanja iz područja modeliranja, zahtijevaju previše interakcija s korisnikom i modeliranje provode preko krutih pravila (npr. određeni se objekt poslovne domene stalno pretvara u entitet, iako negdje treba biti atribut). ABCM sustav se

sastoji iz ABCD generatora (engl. Association-Based Conceptual Diagram) i DEV generatora. ABCD generator prema ABCD pravilima ekstrahira objekte i relacije iz poslovnog opisa kojega sastavlja zaposlenik iz neke poslovne domene, a DEV generator rezultate ABCD generatora prevodi u EV konstrukte. Cjelokupan proces kojim ABCM sustav izvodi DEV se sastoji iz sljedećih koraka:

1. zaposlenik iz neke poslovne domene stvara poslovni opis u prirodnom jeziku.
2. parser iz poslovnog opisa ekstrahira objekte te ih bilježi u meta riječnik.
3. ABCD generator preko ABCD pravila i ekstrahiranih objekata izvodi ABC dijagram. U procesu izvođenja se uočavaju i rješavaju redundantne veze.
4. DEV generator iz ABC dijagrama preko DEV pravila izvodi DEV.
5. DEV se na pogodan način prikazuje zaposleniku iz neke poslovne domene.

Niz radova o primjeni ontologija u konceptualnom modeliranju podupiru važnost razvoja sustava temeljenih na znanju za potporu modeliranju podataka [20], [21], [22], [23], [24].

U [25], opisana je primjena procesiranja prirodnoga jezika (engl. Natural Language Processing - NLP) u procesu konceptualnog modeliranja. Pritom opisuju razvijeni alat DBDT (engl. Data Base Development Tool) namijenjen dizajnerima baza podataka. Alat iz tekstualnog poslovnog opisa na prirodnom jeziku na poluautomatizirani način (uz interakciju korisnika) izvodi model podataka po metodi EV. Proces izvođenje modela podataka ima osam koraka:

1. morfološka analiza poslovnoga opisa u prirodnome jeziku (detektiranje osnovnih objekata i njihovo svodenje na jedninu).
2. sintaktička analiza poslovnoga opisa (dobivanje sintaktičke kategorije svake riječi).
3. semantička analiza (detektiranje EV konstrukata iz sintaktičkih kategorija).
4. integracija konstrukata koji ne slijede direktno iz poslovnog opisa.
5. izbacivanje redundantnih elemenata, sinonima te revizija imenovanja konstrukata.
6. konstrukcija preliminarnog modela.
7. ljudska intervencija i revizija.
8. finaliziranje modela.

U [26] se kao posebni izazov u domeni NLP-a navodi prevođenje semantike iskazane prirodnim jezikom u neki formalni zapis. Kao jedan od jezika za formalizaciju semantike iskazane prirodnim jezikom je jezik SBVR (engl. Semantics of Business Vocabulary and Business Rules). U [27] predstavljen je pristup koji poslovni opis u prirodnome jeziku prevodi u SBVR poslovna pravila.

Poslovna pravila su osnova svakog poslovanja. Ona određuju način izvođenja akcija kako bi se postigao cilj. Ideja svih SBVR formalizacija jest svodenje poslovnog opisa na niz pravila strukturiranih po nekoj SBVR metodi. Autori u [28] naglašavaju da je velik problem u formaliziranju poslovnih opisa to što zaposlenik iz neke poslovne domene nedovoljno poznaje metodu

formalizacije, a s druge strane dizajner ekspert, koji ima dovoljna znanja o metodi formalizacije, pak nedovoljno poznaje poslovnu domenu. Stoga autori razvijaju alat za formalizaciju poslovnoga opisa preko SBVR metode kojega će moći koristiti zaposlenici iz neke poslovne domene koji posjeduju minimalna znanja o metodi formalizacije. Formalizacija izvedena preko alata može dalje poslužiti za automatsko izvođenje modela podataka.

U [29] je također opisana primjena SBVR. Autori su razvili RuleCNL - sustav za formalno definiranje i zapisivanje poslovnih pravila preko posebnog kontroliranog prirodnog jezika. Za njega su autori razvili posebni RuleCNL vokabular i gramatiku te parser koji ih koristi. Nakon što se sastavi poslovni opis koji je prema RuleCNL sustavu sintaktički ispravan (odnosno sastavljen je u kontroliranom prirodnom jeziku) on se dalje automatski formalizira u SBVR notaciju.

Vezano za SBVR, u domeni formalizacije semantike kod poslovnih opisa, treba spomenuti i pojam kontroliranog prirodnog jezika (engl. Control Natural Language - CNL) koji je podskup prirodnog jezika s restriktivnom gramatikom i vokabularom kako bi se reducirala dvosmislenost i složenost prirodnog jezika [30], [31]. To je vrsta prirodnog jezika koja ima formalnu logičku bazu - formalnu sintaksu i semantiku - te je prikladan za prikaz znanja i za proces zaključivanja. CNL je dobar posrednik između prirodnog i formalnog jezika te se stoga i koristi u formalizaciji poslovnoga opisa SBVR jezikom (tekst poslovnoga opisa je u kontroliranom jeziku). [29], [32]

Autor u [33] je razvio parser temeljen na proširenoj mreži prijelaza (engl. Augmented Transition Network – ATN) koji iz poslovnih opisa na prirodnome jeziku stvara dijagram zavisnosti koncepta. Temeljem toga parsera razvijen je i prototip sustava koji prima poslovni opis u prirodnome jeziku i prema njemu daje konceptualnu shemu.

Jedan od načina ekstrahiranja relevantnih konstrukata metode entiteti i veze iz poslovnoga opisa se temelji na semantičkim odnosima među riječima rečenica poslovnog opisa. Dakle, semantičko znanje iz poslovnoga opisa u prirodnome jeziku transformira se u model entiteta i veza. [34]

Osim opisanoga načina ekstrahiranja konstrukata metode entiteta i veza iz poslovnoga opisa, postoji i način u kojemu se primjenjuju posebna pravila. Jedan od alata koji se temelji na ovom pristupu je E-R Generator. To je alat koji pruža mogućnost poluautomatskog generiranja modela podataka metodom entiteta i veza. [35]

Autori u [36] opisuju sustav temeljen na znanju koji se sastoji iz tri modula: modula za čitanje i parsiranje ulaznoga teksta u prirodnom jeziku kojega daje korisnik, modula za heurističko klasificiranje teksta koji će poslužiti kao ulaz sljedećem modulu i modula za generiranje DEV-a i finalnoga izlaza u obliku grafičkog dijagrama. Treći je modul zadužen za generiranje tekstualnog oblika modela koji sadrži sve informacije potrebne za generiranje DEV-a .

Jezik za formalizaciju semantike iskazane prirodnim jezikom je i jezik FMTEK (Formalisation Method for the

Text Expressed Knowledge) opisan u [37], [38], [39], [40], [41]. Autorica vjeruje da ovaj jezik ima potencijala za primjenu u sustavima temeljenim na znanju za modeliranje podataka.

VI. ZAKLJUČAK

U radu je opisan proces modeliranja podataka, te posebno metoda entiteti i veze. Također su prikazani i glavni akteri koji sudjeluju u ovome procesu. Dan je i pregled sustava, alata i metoda za potporu procesa modeliranja podataka.

Svrha istraživanja koje autorica provodi nad procesom modeliranja podataka je razvoj sustava temeljenoga na znanju za potporu modeliranju podataka, a koji će se temeljiti na verbalizaciji – tekstualnom prikazu znanja.

Osnovna ideja autorice je problem modeliranja podataka koje se temelji na poslovnom opisu svesti na problem prevođenja iz kontroliranog prirodnoga jezika u kojemu je poslovni opis verbaliziran u posebno definirani formalni jezik u kojemu je model podataka verbaliziran.

Analizom postojećih sustava temeljenih na znanju, alata i metoda za potporu modeliranju podataka uočeni su elementi izvedbe koji su funkcionalni i efikasni te oni koji su se pokazali neučinkoviti i opterećujući za korisnika. Elementi izvedbe koje autori koriste u svojim sustavima su: restriktivno sučelje i ono sučelje koje vodi korisnika kroz proces modeliranja, preprocesiranje (formaliziranje) poslovnog opisa, otklanjanje semantičkih redundancija, sintaksna i semantička analiza, posebni formalni jezici, CNL i/ili SBVR jezici, pravila, te zaključivanje prema prošlim slučajevima. Sustavi su namijenjeni ili zaposlenicima poslovne domene koji nemaju iskustva u modeliranju podataka ili dizajnerima početnicima.

Novi sustav temeljen na znanju za potporu modeliranju podataka koji se razvija bi trebao primati poslovni opis u formatu koji nije potpuno u prirodnom jeziku već je u nekom podskupu prirodnog jezika (po uzoru na SBVR uz CNL ili FMTEK). Sustav bi također trebao imati mogućnost vođenja korisnika kroz proces izrade potpuno novog modela podataka te isto tako konzultiranja s prošlim slučajevima. Težnja je minimizirati broj pitanja koje sustav postavlja korisniku, budući da veći broj postavljenih pitanja negativno djeluje na korisnički doživljaj. Izlazni format modela podataka kojega sustav daje treba biti formalizirani tekstualni opis dostatan da se iz njega izradi DEV.

LITERATURA

- [1] R. Elmasri and S.B. Navathe, "Fundamentals of Database Systems", 6-th Edition, Addison-Wesley, 2011.
- [2] C. Batini, S. Ceri and S. B. Navathe, "Conceptual Database Design An Entity-Relationship Approach", The Benjamin/Cummings Publishing Company, Inc., 1992
- [3] K. Siau, Y.Wand and I.Benbasat, "When parents need not have children – cognitive biases in information modeling", Lecture Notes in Computer Science – Advanced Information Systems Engineering, 1080, Constantopoulos, P., Mylopoulos, J. & Vassiliou, Y. (eds), pp. 402–420. Springer-Verlag, London, 1996.
- [4] K. Siau and M. Rossi, "Evaluation techniques for systems analysis and design modelling methods – a review and

- comparative analysis”, *Information Systems Journal* 21, 249-268, Blackwell Publishing Ltd, 2011
- [5] P. Shoval and S. Shiran, “Entity-relationship and object-oriented data modeling - an experimental comparison of design quality”, *Data Knowl Eng* 21(3), pp. 297-315, 1997
- [6] A. Burton-Jones, Y. Wand and R. Weber, “Guidelines for Empirical Evaluations of Conceptual Modeling Grammars”, *Journal of the Association for Information Systems*, Volume 10, Issue 6, pp. 495-532, June 2009
- [7] M. Pavlič, “Oblikovanje baza podataka”, Odjel za informatiku Sveučilišta u Rijeci, 2011
- [8] P. Merson, “Data Model as an Architectural View”, technical note, Carnegie Mellon University, 2009.
- [9] M. Pavlič, “Informacijski sustavi”, Odjel za informatiku Sveučilišta u Rijeci, 2009
- [10] Dana Modeling, <https://www.liberty.edu/media/1414/%5B6330%5DERDDataModeling.pdf> (13.10.2015)
- [11] I.Y. Song, M. Evans and E.K. Park, “A Comparative Analysis of Entity-Relationship Diagrams”, *Journal of Computer and Software Engineering*, Vol. 3, No.4, pp. 427-459, 1995.
- [12] P. Chen, “The Entity-Relationship Model—Toward a Unified View of Data,” *ACM Transactions on Database Systems*, 1, pp. 9-36, 1976.
- [13] Ž. Panian, K. Ćurko, “Poslovni informacijski sustavi”, Sveučilište u Zagrebu, 2010
- [14] W.A. Lo and J. Choobineh, “Knowledge-based systems as database design tools: A comparative survey”, *Journal of Database Management*, 10, 3, pp. 26-40, 1999.
- [15] S.R. Antony and D. Batra, “CODASYS: a consulting tool for novice database designers”, *Data Base for Advances in Information Systems* 33 (3) pp. 54- 88, 2002.
- [16] D. Batra, and S.R. Antony, “Novice errors in conceptual database design”, *European Journal of Information Systems* 3 (1), 57- 69, 1994.
- [17] J. Choobineh and W. A. Lo, “CABSYYDD: Case-Based System for Database Design,” *Journal of Management Information Systems*, Vol. 21, No. 3, pp. 281-314, Winter 2004-5.
- [18] K. Namgyu, S. Lee and S. Moon, “Formalized entity extraction methodology for changeable business requirements”, *Journal of information science and engineering* 24, pp. 649-671, 2008.
- [19] S. Lee K. Namgyu, and S. Moon SANGWON, “Context-Adaptive Approach for Automated Entity Relationship Modeling”, *Journal of information science and engineering* 26, pp. 2229-2247, 2010.
- [20] V. C. Storey, R. C. Goldstein, and H. Ullrich, “Naive semantics to support automated database design”, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 14, , pp. 1-12, 2002.
- [21] V. Sugumaran and V. C. Storey, “Ontologies for conceptual modeling: Their creation, use, and management,” *Data and Knowledge Engineering*, Vol. 42, , pp. 251-271., 2002.
- [22] V. Sugumaran and V. C. Storey, “Supporting database designers in entity-relationship modeling: An ontology-based approach,” in *Proceedings of the 24th International Conference on Information Systems*, , pp. 59-71., 2003.
- [23] Y. Wand, V. C. Storey, and R. Weber, “An ontological analysis of the relationship construct in conceptual modeling,” *ACM Transactions on Database Systems*, Vol. 24, pp. 494-528., 1999.
- [24] R. Weber, “Conceptual modeling and ontology: Possibilities and pitfalls,” *Journal of Database Management*, Vol. 14, pp. 1-20, 2003.
- [25] L.A.E. Al-Safadi, “Natural Language Processing for Conceptual Modeling”, *International Journal of Digital Content Technology and its Applications*. 3, 2009.
- [26] M. Kleiner, P. Albert and J. Bézivin, “Parsing SBVR-Based Controlled Languages”, *Model Driven Engineering Languages and Systems*, LNCS Vol. 5795 pp.122-136, 2009
- [27] S. I. Bajwa, M. G. Lee and B. Bordbar, “SBVR Business Rules Generation from Natural Language Specification”, *Association for the Advancement of Artificial Intelligence (www.aaai.org)*, 2010.
- [28] M. Selway, G. Grossmann, W. Mayer and M. Stumptner, “Formalising Natural Language Specifications using a Cognitive Linguistics/Configuration Based Approach”, *IEEE* 2013.
- [29] P. B. F. Njonko, S. Cardey, P. Greenfield and W. El Abed, “RuleCNL: A Controlled Natural Language for Business Rule Specifications”, *arXiv:1406.2096v1 [cs.SE]* 9 Jun 2014.
- [30] G. Hart, M. Johnson and C. Dolbear, “Rabbit: Developing a Control Natural Language for Authoring Ontologies”, *5th European Semantic Web Conference (ESWC'08)* pp. 348-360, 2008.
- [31] S. Rolf, “Controlled Natural Languages for Knowledge Representation”, *COLING '10 Proceedings of the 23rd International Conference on Computational Linguistics*, pp. 1113-1121, 2010.
- [32] [32] N.E. Fuchs, K. Kaljurand, and T. Kuhn, “Attempto Controlled English for Knowledge Representation”, *Reasoning Web, Lecture Notes in Computer Science*, Volume 5224, pp. 104-124, 2008.
- [33] A. Gangopadhyay, “Conceptual modeling from natural language functional specifications”, *Artificial Intelligence in Engineering* Volume 15, Issue 2, pp. 207-218, April 2001.
- [34] N. Omar, P. Hanna and P. Kevitt, “Semantic Analysis in the Automation of ER Modelling through Natural Language Processing”, *International Conference Computing & Informatics*, pp.1-5, June 2006.
- [35] F. Gomez, C. Segami and C. Delaune, “A system for the semiautomatic generation of E-R Models from Natural Language Specifications”, *Data and Knowledge Engineering* 29 (1) pp. 57-81, 1999.
- [36] M. Shahbaz S. Ahsan, M. Shaheen, R.M.A. Nawab and S.A. Masood, “Automatic Generation of Extended ER Diagram Using Natural Language Processing”, *Journal of American Science*, 7(8), 2011.
- [37] A. Jakupović, M. Pavlič and H. Z. Dovedan, “Formalisation method for the text expressed knowledge”, *Expert systems with applications*. 41 (11) pp. 5308-5322, 2014.
- [38] A. Jakupović, M. Pavlič, A. Meštrović and V. Jovanović, “Comparison of the Nodes of Knowledge method with other graphical methods for knowledge representation”, *Proceedings of the 36th international convention. Rijeka: Croatian Society for Information and Communication Technology, Electronics and Microelectronics – MIPRO*, pp.1276-1280, 2013.
- [39] M. Pavlič, A. Jakupović, and A. Meštrović, “Nodes of knowledge method for knowledge representation”, *Informatologia*. 46 (3), pp. 206-214, 2013a.
- [40] M. Pavlič, A. Meštrović and A. Jakupović, “Graph-Based Formalisms for Knowledge Representation”, *Proceedings of the 17th World Multi-Conference on Systemics Cybernetics and Informatics (WMSCI 2013)*, Volume 2. Orlando: IIS, pp. 200-204, 2013b.
- [41] M. Pavlič, H. Z. Dovedan, and A. Jakupović, “Question Answering with a Conceptual Framework for Knowledge-Based System Development "Node of Knowledge"”, *Expert systems with applications*. 42 (12), pp.5264-5286, 2015.