

SPECIFIKACIJE PRESLIKAVANJA IZMEĐU HETEROGENIH ŠEMA PODATAKA KORISTEĆI PRISTUP ZASNOVAN NA MODELIMA

SPECIFICATIONS MAPPINGS BETWEEN HETEROGENEOUS DATA SCHEMAS USING MODEL DRIVEN APPROACH

Siniša Nešković, Milica Vučković, Nenad Aničić, *Fakultet Organizacionih Nauka, Beograd*

Sadržaj - U radu se razmatra originalni postupak za rešavanje problema specifikacije preslikavanja između heterogenih šema podataka koristeći razvoj zasnovan na modelima. Definisana je problem specifikacije preslikavanja, a zatim je dat opis postojećeg pristupa za njegovo rešavanje u Eclipse Modeling Framework okruženju i diskutovani su njegovi nedostaci. Rešenje koje se predlaže u ovom radu zasniva se na eksplicitnom korišćenju meta-šema podataka i odgovarajućeg modela za semantička pravila preslikavanja između njihovih koncepata. Ovaj model semantičkih preslikavanja se tretira kao meta-model za specifikaciju preslikavanja između konkretnih šema. U radu je razmatran i problem realizacije predloženog rešenja u EMF okruženju. Identifikovana su dva moguća načina realizacije i diskutovane se njihove prednosti i mane.

Abstract - The paper discusses an original approach for solving the problem of mapping specifications between heterogeneous schema using model driven development. Firstly, the problem of mapping specifications is defined and then a description of the existing approach for its solution in the Eclipse Modeling Framework environment as well as its shortcomings is given. The proposed solution is based on the explicit use of meta-schemas and the corresponding models expressing semantic rules for mappings between their concepts. Semantic rules for mappings are treated as a meta-model for specifying mappings between specific schemes. The paper discussed the problem of implementation of proposed solutions in the EMF environment. Two possible realizations are identified and their advantages and disadvantages are discussed.

1. UVOD

Specifikacija preslikavanja između različitih šema podataka danas je vrlo aktivno polje istraživanja u različitim istraživačkim kontekstima: bazama podataka [1], ontologija [2], skladišta podataka (data warehousing) [3], razvoja softvera zasnovan na modelima (model driven development - MDD) [4], itd. U ovom radu se problem preslikavanja razmatra prvenstveno u kontekstu razvoja zasnovanog na modelima.

Razvoj zasnovan na modelima (RZM) se bazira na ideji da najvažniji rezultat razvoja softvera nije izvorni kod, već modeli koji predstavljaju modifikovano znanje o sistemu koji se razvija. Glavni cilj RZM je da se automatizuje razvoj softvera kroz postupak uzastopnih transformacija modela, počev od modela koji predstavljaju specifikaciju sistema do modela koji predstavljaju detaljni opis fizičke realizacije iz kojih se onda može automatski generisati izvorni kod izvršnog sistema.

U skladu sa osnovnim motom da sve treba tretirati kao modele, u RZM se za specifikaciju preslikavanja između heterogenih šema koristi posebna vrsta modela tzv. modeli tkanja (weaving models - WM), [5, 6, 7]. Ovim modelima se zapravo definišu veze (korespondencije) između pojedinih elemenata drugih modela. U skladu sa OMG arhitekturom, definisan je i odgovarajući meta-model (meta-weaving-

model – MWM) sa kojim WM modeli moraju biti u skladu. Sam WMM meta-model je definisan na osnovu meta-meta-modela, odnosno tzv. ECore modela koji je jedna konkretna implementacija MOF meta-meta-modela iz OMG MDA standarda. WMM model zapravo definiše tipove veza (vrste korespondencija) koje se mogu javiti između pojedinih koncepata konkretnih šema. Sa druge strane, WM model definiše preslikavanja između šema preko skupa veza (korespondencija) određenog tipa definisanih u WMM.

Međutim, u definiciji WMM modela se ne koriste koncepti meta-modela šema podataka, tj. koncepti jezika za opis šema (npr. SQL DDL ili XSD), već samo koncepti Ecore meta-meta-modela. Stoga, odgovarajući WM modeli, tj. preslikavanja između konkretnih šema (npr. konkretna relacionalna šema i konkretna XML šema) nisu kontrolisani semantičkim pravilima preslikavanja između njihovih meta-modela.

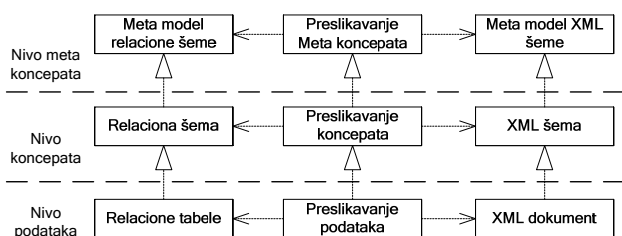
U ovom radu se razmatra prevazilaženje navedenog problema kroz eksplicitno uvođenje meta-modela šema podataka i eksplicitnog definisanja semantičkih pravila preslikavanja između njihovih koncepata. Najpre se u sledećem poglavlju detaljnije opisuje problem preslikavanja šema podataka. Zatim se u trećem poglavlju opisuje postojeći pristup u RZM koji se u okruženju EMF (Eclipse Modeling Framework) [8] koristi za specifikaciju preslikavanja šema. Predloženi novi pristup za specifikaciji šema i mogući načini njegove konkretne tehničke realizacije se razmatra

u četvrtom poglavlju. Na kraju, u zaključku se daju osnovne prednosti ovog pristupa i diskutuju dalji pravci istraživanja.

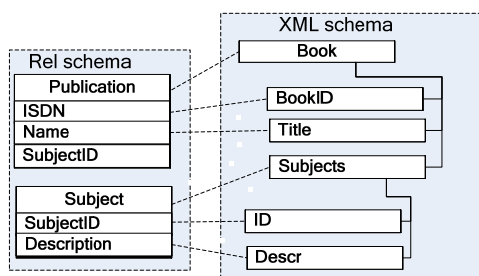
2. OPIS PROBLEMA PRESLIKAVANJA ŠEMA PODATAKA

Generalno govoreći, pod pojmom preslikavanja modela podrazumeva se definisanje korespondencije između elemenata različitih modela. U kontekstu integracije podataka specifikacije preslikavanja se mogu posmatrati na tri nivoa apstrakcije, kao što je to pokazano na Slici 1:

- Nivo meta konceptata na kome se definišu preslikavanja između konceptata jezika za definisanje šema podataka
- Nivo konceptata na kome se definišu preslikavanja između pojedinih konceptata šema podataka
- Nivo podataka na kome se definišu preslikavanja između instanci konceptata šeme podataka



Slika 1. Specifikacija preslikavanja na različitim apstraktnim nivoima



Slika 2. Preslikavanje između relacione i XML šeme

Očigledno je da navedeni nivoi preslikavanja odgovaraju apstraktnim nivoima odgovarajućih modela. Pri tome, slično kao što važi za modele sa različitih apstraktnih nivoa u OMG MDA arhitekturi [4], preslikavanje sa jednog nivoa apstrakcije definiše pravila za preslikavanja na nižem nivou apstrakcije, odnosno mora biti u skladu sa preslikavanjem sa višeg nivoa apstrakcije. Drugačije rečeno, preslikavanje na jednom nivou apstrakcije predstavlja ekstenziju skupa preslikavanja sa nižeg nivoa (tj. definiše način za izvođenje preslikavanja sa nižeg nivoa), odnosno predstavlja intenziju (eksplicitno navedene elemente) skupa preslikavanja definisanog na višem nivou.

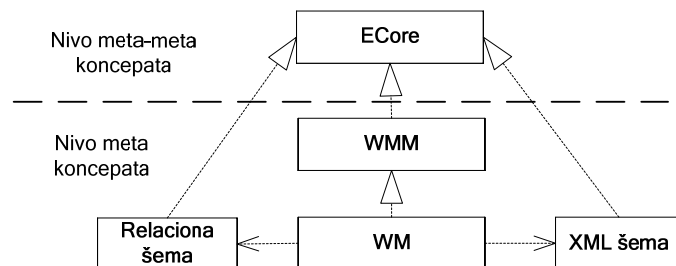
Na Slici 1. je dat primer za slučaj preslikavanja između relacionog modela i XML modela. Na nivou meta konceptata se definišu pravila kako se pojedini koncepti relacionog

modela (npr. table i kolone) preslikavaju u koncepte XML šeme (npr. XML elemente i attribute). Na nivou konceptata se definišu pravila za preslikavanje između konkretnih šema podataka. Ova preslikavanja su ilustrovana na Slici 2. Na nivou podataka su preslikavanja koja definišu preslikavanja između podataka smeštenih u relacionoj bazi podataka i podataka koji su zapamćeni u XML dokumentima.

Predmet interesa ovog rada su primarno nivoi meta konceptata i konceptata, a posebno način njihove upotrebe u kontekstu razvoja zasnovanog na modelima.

3. POSTOJEĆI PRISTUP ZA PRESLIKAVANJA ŠEMA U RAZVOJU ZASNOVANOG NA MODELIMA

Kao što je već rečeno u uvodu, u okviru RZM se preslikivanje šema zasniva se na definisanju eksplicitnog modela tkanja (weaving model) zasnovanog na ATLAS Model Weaver (AMW) metamodelu [5,6,7]. U ovom pristupu weaving model (WM) se koristi za reprezentovanje preslikavanja između elemenata različitih modela. WM je zapravo u saglasnosti sa weaving meta modelom (WMM), koji je proširenje generičkog WMM tzv. Core WMM. Generički weaving meta model sadrži informacije o tipovima preslikavanja (weaving linkovi) između elemenata modela koji se povezuju. Njegovom proširenjem se mogu se definisati novi tipovi preslikavanja, specifični za određeni domen.



Slika 3. Postojeći pristup za preslikavanje šema u razvoju zasnovanom na modelima

Karakteristike opisnog pristupa je da se ne definišu preslikavanja meta konceptata, na primer preslikavanja između konceptata relacione šeme i XML šeme, već samo preslikavanja konceptata. Stoga da bi se iskazala mapiranja konkretnih šema, one se proglašavaju meta modelima, tj. za definisanje šema se koristi Ecore, meta-meta model [8]. Sa druge strane, preslikavanje konceptata iskazano preko WM modela se zasniva na WMM meta-modelu. Ovaj meta-model je, kao i svaki drugi, definisan preko Ecore meta-meta modela. To znači da on se ne zasniva na konceptima iz jezika za opisivanje šema podataka, već isključivo na konceptima ECore-a.

Ovo predstavlja suštinski problem ovog pristupa, jer WMM meta-model ne sadrži potrebna semantička pravila za preslikavanje između šema. Posledica ovoga je da preslikavanje između koncepta konkretnih šema nije semantički kontrolisano već samo sintaksno. Drugačije rečeno, na projektantu je da poznaje ova semantička pravila i da prilikom preslikavanja sam vodi računa o njima.

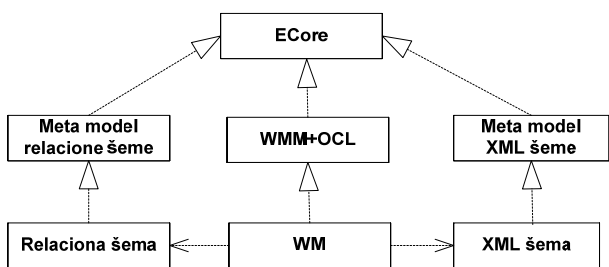
4. PREDLOG NOVOG PRISTUPA

Kao rešenje navedenog problema se predlaže eksplicitno uvođenje meta šema podataka i odgovarajućeg meta modela za semantička preslikavanja između njih, kao što je to pokazano na Slici 1.

Međutim, ovo predloženo logičko rešenje se ne može direktno realizovati u EMF implementacionom okruženju. Indirektno, ovo rešenje se može realizovati na jedan od sledeća dva načina:

1. Eksplicitnim definisanjem meta modela za semantička preslikavanja
2. Definisanjem WM modela za semantička preslikavanja i njegovim transformisanjem u WMM

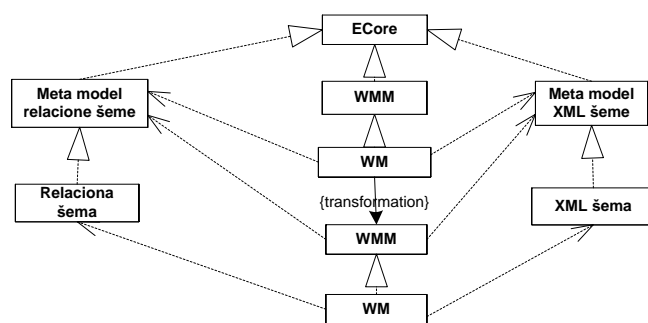
Prvi način podrazumeva da se meta modela za semantička preslikavanja iskaže kao proširenje kroz proširenje postojećeg WMM sa odgovarajućim OCL ograničenjima [9]. Ovo rešenje je prikazano na Slici 4.



Slika 4. Specifikacija modela preslikavanja zasnovana na OCL ograničenjima

Osnovna svrha OCL ograničenja je da se njima definišu semantička pravila preslikavanja. Pošto se WMM ne definiše nad konceptima Ecore modela, OCL ograničenja se mogu samo eksplicitno referencirati na koncepte meta šema podataka koristeći njihove nazive. To predstavlja i osnovni nedostatak ovog pristupa, jer se OCL ograničenja uvode na nivou meta konceptata i ne koriste se koncepti šema direktno u definisanju mogućih preslikavanja.

Drugo rešenje se zasniva na ideji da se WM model jednoznačno može transformisati u WMM meta model za opis specifikacije preslikavanja na nižem apstraktnom nivou. Ovde se rešenje problema preslikavanja najpre svodi na definisanje semantički pravila preslikavanja između konceptata meta-šema preko odgovarajućeg WM modela, a zatim transformisanjem ovog modela u WMM meta-model koji služi kao meta model za WM model na nižem nivou apstrakcije kojim se definišu preslikavanja između konceptata konkretnih šema podataka. Na ovaj način se definiše hijerarhija modela preslikavanja koji su konkretizacija prethodno definisanog modela ili specifikacija preslikavanja na nižem hijerarhijskom nivou. Ovo je ilustrovano na Slici 5.



Slika 5. Specifikacija preslikavanja zasnovana na transformaciji WM modela

Prednost ovog pristupa u odnosu na prethodne je u tome što se veze između elemenata modela uvek definišu kao ekstenzija, i dobijena ekstenzija se može transformisati i predstavljati intenziju (specifikaciju pravila) za opis elementa na nižem apstraktnom nivou.

5. ZAKLJUČAK

U radu je dat originalni predlog za rešavanje problema specifikacije preslikavanja između heterogenih šema podataka u kontekstu razvoja zasnovanog na modelima. Najpre je opisan problem specifikacije, a zatim je dat opis postojećeg pristupa u EMF okruženju i diskutovana su njegove karakteristike i nedostaci.

Predloženo rešenje podrazumeva eksplicitno korišćenje meta-šema podataka i odgovarajućeg modela za semantička pravila preslikavanja između konceptata meta-šema podataka. Ovaj model semantičkih preslikavanja treba da bude meta-model za model preslikavanja između konkretnih šema.

U radu je razmatran i problem realizacije ovog rešenja u EMF okruženju. Identifikovana su dva moguća načina realizacije i diskutovane su njihove prednosti i mane.

Dalji pravci istraživanja su usmereni na realizaciju ovih predloženih načina realizacije.

LITERATURA

[1] M. Lenzerini, *Data integration: a theoretical perspective*, Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, June 03-05, 2002, Madison, Wisconsin.

[2] M. Ehrig, York Sure, *Ontology Mapping - An Integrated Approach*, ESWS 2004.

[3] Y. Cui, and J. Widom, *Lineage Tracing for General Data Warehouse Transformations*, VLDB Journal, 12(1):41-58, May 2003.

- [4] J. Miller, J. Mukerji, *Model Driven Architecture (MDA)*, OMG Document available at <http://www.omg.org>; 2001.
- [5] M. Didonet Del Fabro, J. Bézivin, F. Jouault, P. Valduriez. *Applying Generic Model Management to Data Mapping*, In V. Benzaken, editor, Proc 21emes Journees Bases de Donne BDA 2005, Saint Malo, Actes, 2005.
- [6] M. D. Del Fabro, P. Valduriez, *Semi-automatic Model Integration using Matching Transformations and Weaving Models*, SAC'07, March 11-15, 2007, Seoul, Korea.
- [7] M. D. Del Fabro, P. Valduriez, *Towards the Efficient Development of Model Transformations Using Model Weaving and Matching Transformations*, Software and System Modeling 8(3): 305-324 (2009).
- [8] *Eclipse Modeling Framework Project (EMF)*, <http://www.eclipse.org/modeling/emf/>
- [9] Object Management Group (OMG), *Object Constraint Language OMG Available Specification Version 2.0*, May 2006.
- [10] F. Jouaultand, I. Kurtev, *Transforming Models with ATL*. MoDELS 2005 Conference, Jamaica, 2006.
- [11] F. Jouaultand, F. Allilaire, J. Bézivin, I. Kurtev, *ATL: A Model Transformation Tool*, Science of Computer Programming 2008 journal homepage: www.elsevier.com/locate/scico
- [12] *XSL Transformations (XSLT)*, <http://www.w3.org/TR/xslt>