
Formal model for transforming relational data into RDF based on schema mapping (RFC)

Martin Švihla
svihlm1@fel.cvut.cz

KEG 7.12.2006



Agenda

- Úvod
 - Motivácia a ciele
 - Základné princípy
 - Nasadenie, testy
 - Rozšírenia
- **Formalizácia**
 - Formálne ciele a definície
 - Formálny model transformácie
 - Úplnosť – relačná, RDF, RDFS
 - Rozšírenie na OWL
 - Diskusia



KEG 7.12.2006

(2)



Motivácia

- The goal of the work is to enable a transformation of relational data into an RDF document. The transformation is based on a mapping that maps a relational database schema into an ontology.
- a schema formalism for a source schema is a relational model,
- source data are stored in an RDBMS,
- a schema formalism for a target schema is an RDFS and
- target data are stored in an RDF document

Ciele

- Data transformation
- Ďalšie ciele:
 - Flexibility
 - Performance
 - Reliability
 - Usability

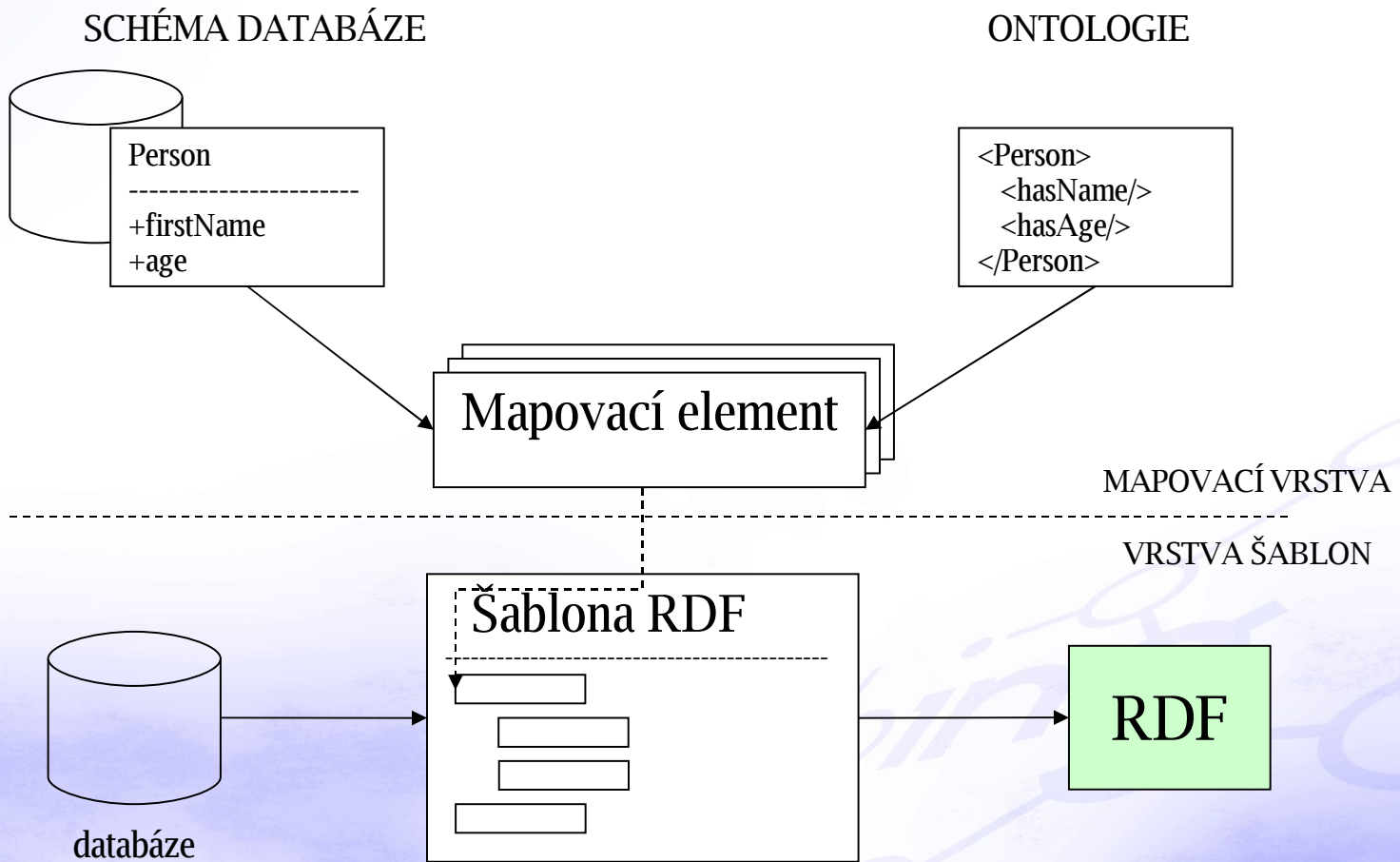


KEG 7.12.2006

(4)



Základný princíp: dve vrstvy



Proces transformácie

1. Požadavek:
person.rdf?id=novak

3. Požadavek na data
triedy Person a vlastnosti
name a surname kde id
= novak

6.
DATA

DATABÁZE

5. SQL dotaz na DB:
SELECT * FROM
people WHERE
username LIKE 'novak'

DB/RDF
Mapping

empla
layer

RDF

8. RDF:
<Person>
<name>
<surname>
</Person>

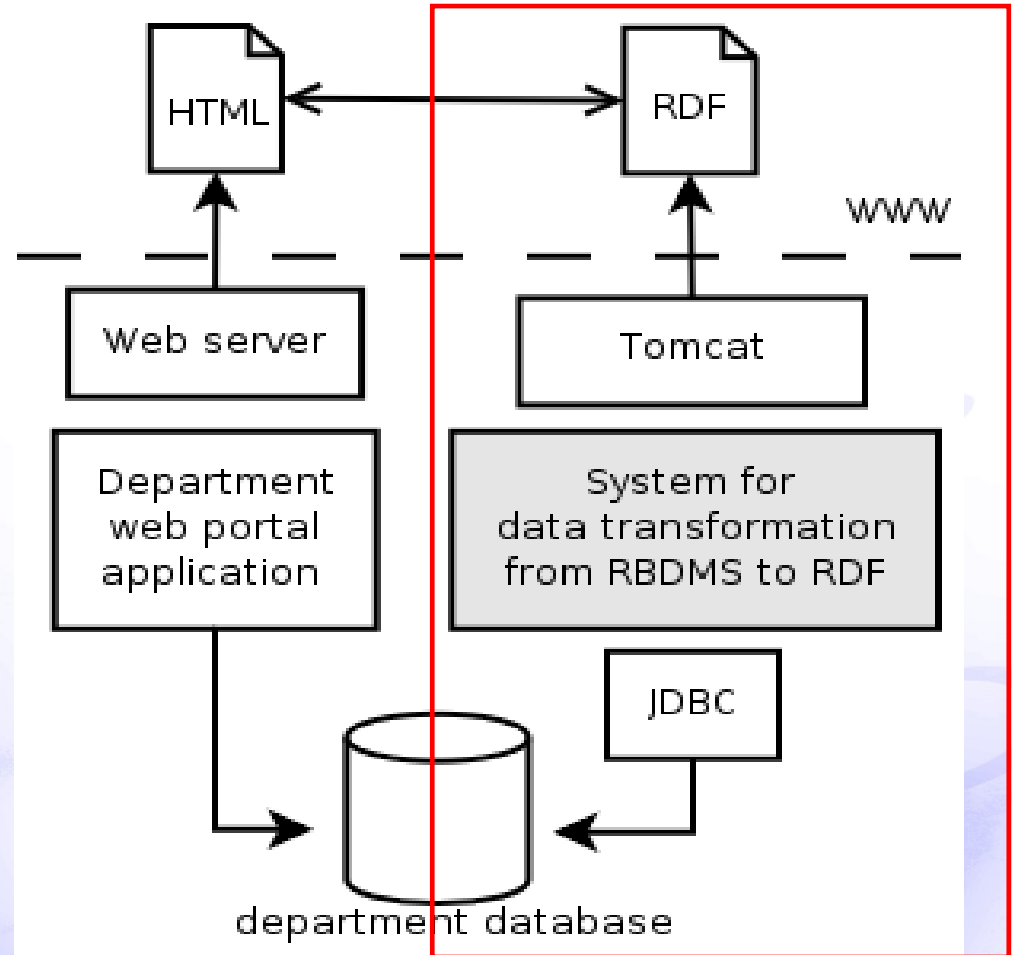
2. Načtení šablony
person
<Person>
<name/>
<surname/>
</Person>

7. DATA PERSON.
name=Jan
surname=Novák

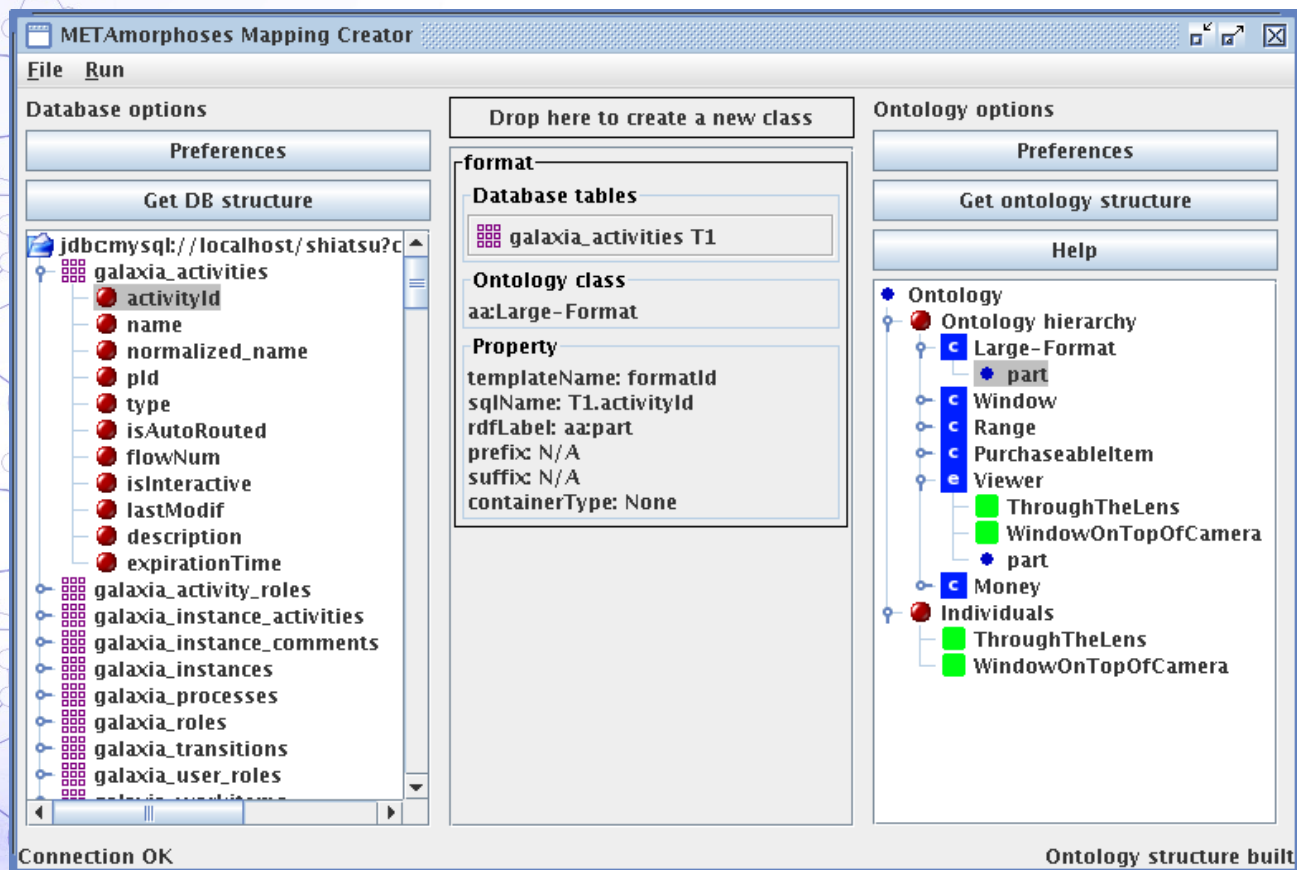
Mapping
document

Publikovanie RDF na webe

- Vytvoriť ontológiu
- Vytvoriť mapovanie a RDF šablóny
- Publikovať relačné dáta ako RDF



Vytváranie mapovania: Mapping editor



- drag&drop tool, vytvára mapovací dokument
- obmedzenie: len jedna ontológia

Implementačné detaily

- Transformačný procesor – Java 1.4
- Mapovací a template dokument – XML
- Pripojenie na DB – JDBC
- Publikačný server – Java Servlet
- Mapping editor – Java 5



KEG 7.12.2006

(9)



Nasadenie

- Sémantická verzia webového portálu katedry počítačov FEL ČVUT
- Každá stránka portálu (o človeku, publikácii...) má svoj RDF obraz
- Case study:
 - zber dát
 - ukladanie v Sesame
 - publikovanie v rôznych formátoch na webe
 - semantic web search



KEG 7.12.2006

(10)



Testy

- veľmi predbežné benchmarks, bez metodiky
- dotaz vypíše všetky publikácie z databázy s ich titulom()
 - D2RQ s Jena Graph API ~ 300ms (tuple result)
 - D2RQ s Jena SPARQL query API ~ 600ms (tuple result)
 - Sesame2 SeRQL query ~ 300ms (RDF graph result)
 - **METAmorphoses ~ 70ms (RDF graph result)**

Koniec úvodu – priestor na otázky

- **Formalizácia**

- Formálne ciele a definície
- Formálny model transformácie
- Úplnosť – relačná, RDF, RDFS
- Rozšírenie na OWL
- Diskusia



KEG 7.12.2006

(12)



Úvodné definície

Definition (Schema) The schema S is a structure consisting of concepts, their relationships and restrictions:

$$S = (\text{CON}, \text{REL}, \text{RES})$$

Definition (Schema formalism) Schema formalism is a triple

$$F = (Cf, Rf, Sf)$$

Definition (Instance) The instance I is data representing a concept from a schema S and conforming to the schema.

Definition (Database) The database D is a set of instances that conform to the schema S and satisfy all constraints in S .

Formálna definícia cieľa

Definition (Data transformation) Data transformation is an operation ts, t , which is able to create a target database Dt conforming to a schema St from data stored in a source database Ds conforming a schema Ss .

$ts, t : (Ds \rightarrow Dt) \mid Ds \text{ conforms } Ss, Dt \text{ conforms } St$

Schema mapping

Definition (Schema mapping) Given a source schema $S_s = (\text{CONs}, \text{RELs}, \text{RESs})$ and a target schema $S_t = (\text{CONt}, \text{RELt}, \text{RESt})$ describing one knowledge domain. The schema mapping M is a set of mapping elements each of which indicates that certain concepts or relationships of schema S_s are mapped to certain concepts or relationships in S_t .

$$M = (E_c (cs \rightarrow ct), E_r (rs \rightarrow rt) \mid cs \text{ in CONs}, rs \text{ in RELs}, ct \text{ in CONt}, rt \text{ in RELt})$$

where

- E_c is a set of mapping elements between concepts,
 - E_r is a set of mapping elements between relationships
- and mapping elements are created in accordance with restrictions from RESs and RESt .

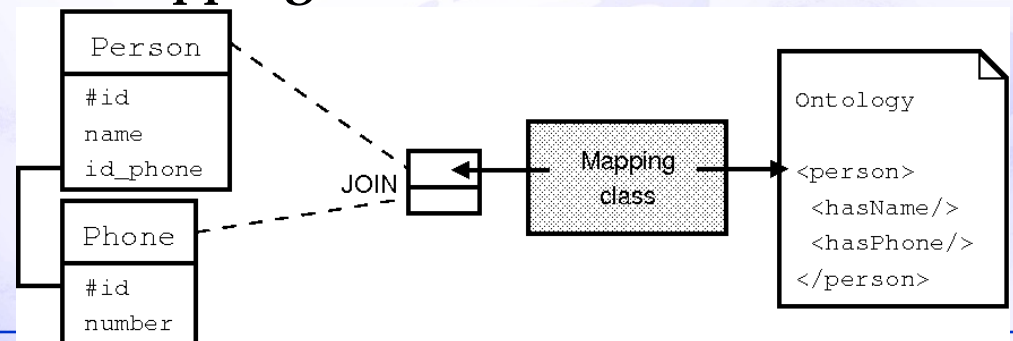


Concept mapping: class

Definition (Mapping class) A mapping class C_m is a triple
 $C_m = (rs, co, id \mid rs \text{ subset } S_s, co \text{ in } Ot)$

where

- rs is a relation schema of a composed relation referenced by an relation algebra operator,
- co is a referenced ontology class,
- S_s is a source relational database schema,
- O_t is a target ontology and
- id is a unique identificator of a mapping class.



KEG 7.12.2006

(16)

Concept mapping: property

Definition (Mapping property) There is a mapping class M_c connecting concepts from a relational database schema S_s and an ontology O_t . A mapping property P_m that belongs to M_c can be defined as:

$$P_m = (A_s, p_o, id \mid A_s \text{ rs}, p_o \text{ is property of } c_o)$$

where

- A_s is a referenced relation schema attribute,
- p_o is a referenced ontology property,
- rs is a relation from S_s referenced by M_c ,
- c_o is a class from O_t referenced by M_c and
- id is a unique identifier of a mapping property.

Relationships and restrictions in mapping

- relationship mapping:
 - vzťah medzi relačnou schémou a atribútom sa mapuje do rdfs:domain
 - foreign keys: rdfs:range? vzťahy medzi inštanciami?
- restrictions:
 - dáta v relačnej schéme zodpovedajú relačným constraints
 - RDFS nemá restriction primitives
 - OWL podpora v budúcnosti

Template layer

- template je XML dokument
- template tvorí graf z odkazov na **mapping elements**
 - uzly = odkazy na mapping classes
 - hrany = odkazy na mapping properties
- mapping property môže byť:
 - object property: nemá hodnotu a obsahuje uzol
 - datatype property: hodnota z databáze (literál)
- template je vlastne šablóna pre RDF
- pri transformácii sa template plní dátami z RDB a vzniká RDF graf

Úplnosť

- relačná: SQL dokáže všetky operácie relačnej algebry
 - dôkaz príkladom
- RDF: template je vlastne RDF graf
 - z RAL vlastnosti uzlov a hrán
 - ostatné RDF vlastnosti (kontainery atd.) vychádzajú z dátového modelu
- RDFS: podporované všetky vlastnosti
 - výpis

Ciel' splnený:

Definition (Data transformation) Data transformation is an operation ts, t , which is able to create a target database Dt conforming to a schema St from data stored in a source database Ds conforming a schema Ss .

$$ts, t : (Ds \rightarrow Dt) \mid Ds \text{ conforms } Ss, Dt \text{ conforms } St$$

- môžeme vytvoriť RDF graf (target database)
- RDF graf je vytvorený na základe RDFS ontológie (cez odkazy v mapovaní)
- podpora všetkých vlastností RDF a RDFS

Rozšírenie na OWL

- OWL postavené na RDFS
- z toho vyplýva podpora mnohých OWL elementov:
 - koncepty (class, properties)
 - pasívne podporované aj niektoré OWL vlastnosti (tak jako `rdfs:subclassOf`)
- náš model nepodporuje OWL obmedzenia
 - budúcnosť?
 - riešenie: kontrola generovaného RDF na splnenie podmienok obmedzení?

No a to je zatiaľ všetko...

...ďakujem za pozornosť a prosím o komentáre.

Martin Švihla (svihlam@vse.cz)



KEG 7.12.2006

(23)

