PROCEEDINGS OF THE LATVIAN ACADEMY OF SCIENCES. Section B, Vol. 63 (2009), No. 4/5 (663/664), pp. 156–158.
DOI: 10.2478/v10046-009-0033-6

Short Communication

ONTOLOGICAL RE-ENGINEERING OF MEDICAL DATABASES

Guntis Bārzdiņš, Sergejs Rikačovs, Marta Veilande, and Mārtiņš Zviedris

Institute of Mathematics and Computer Science, University of Latvia, Raiņa bulv. 29, Rīga, LV-1459, LATVIA

Communicated by Andrejs Ērglis

This paper describes data export from multiple medical databases (relational databases) into a single shared Medical Data Warehouse (RDF database structured according to an integrated OWL ontology). The exported data is conveniently accessible via SPARQL or via graphical query language ViziQuer based on UML profile for OWL. The approach is illustrated on one of Latvian Medical databases – Injury Register.

Key words: *ontologies, information system design, medical statistics.*

There are several medical statistics databases in Latvia covering the main pathologies endangering the quality of life by spread of risk factors to residents of Latvia. Our task has been to merge these legacy relational databases into a single unified Medical Data Warehouse accessible for medical researchers. Additionally, medical researchers require easy-to-understand ontologies describing the structure of available data and a flexible query mechanism for data retrieval.

Previously, all medical statistics data was stored in relational databases, which caused serious problems with data retrieval. Firstly, medical researchers were not familiar with the relational database schemas employed in each of the databases — these schemas included a lot of technical encodings only a database administrator could understand. Secondly, medical researchers were not familiar with the data retrieval language SQL, so they had to involve a programmer to program their request.

Our solution (Barzdins *et al*., 2008) to this problem is to use semantic web technology: relational database schema is transformed to OWL ontology, and visualised in UML, which is a format already readable by a medical researcher.

The transformed OWL data can be queried through standard SPARQL query language by a programmer. For medical researchers we have developed a graphical front-end ViziQuer for composing SPARQL queries directly from UML-like visualisations.

In order to collect all data in medical databases, inquiry forms are implemented for each medical register. Figure 1 illustrates a fragment of an inquiry form for the Trauma/Injury Register that is included in the Unified Information System PREDA stored in *VSMTVA* (Health Statistics and Medical Technologies State Agency). Therefore, we decided that it would be easier to build ontologies using these

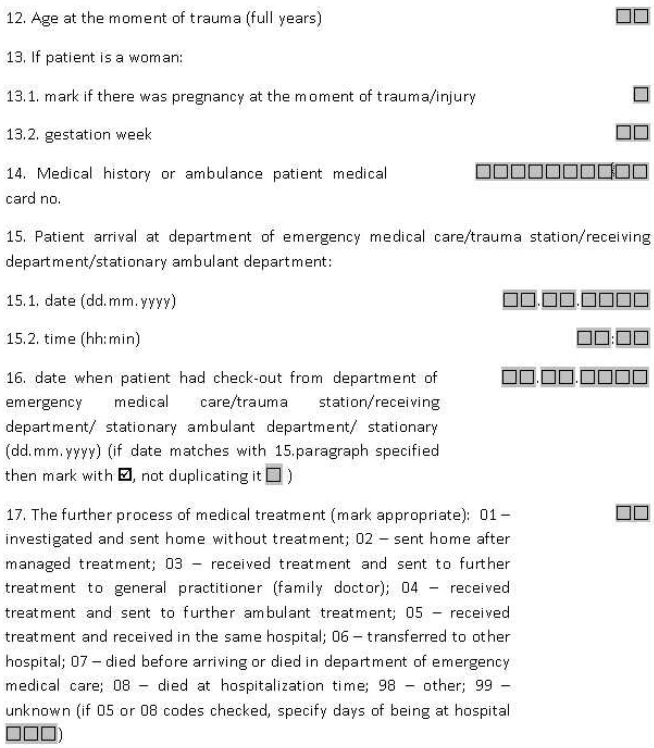
12. Age at the moment of trauma (full years)
13. If patient is a woman:
13.1. mark if there was pregnancy at the moment of trauma/injury
13.2. gestation week
14. Medical history or ambulance patient medical card no.
15. Patient arrival at department of emergency medical care/trauma station/receiving department/stationary ambulant department:
15.1. date (dd.mm.yyyy)
15.2. time (hh:min)
16. date when patient had check-out from department of emergency medical care/trauma station/receiving department/ stationary ambulant department/ stationary (dd.mm.yyyy) (if date matches with 15.paragraph specified then mark with ☒, not duplicating it ☐)
17. The further process of medical treatment (mark appropriate): 01 – investigated and sent home without treatment; 02 – sent home after managed treatment; 03 – received treatment and sent to further treatment to general practitioner (family doctor); 04 – received treatment and sent to further ambulant treatment; 05 – received treatment and received in the same hospital; 06 – transferred to other hospital; 07 – died before arriving or died in department of emergency medical care; 08 – died at hospitalization time; 98 – other; 99 – unknown (if 05 or 08 codes checked, specify days of being at hospital ☐☐☐)

Fig. 1. Fragment of an inquiry form for Trauma/Injury Register.

inquiry forms than understanding physical data schemas of relation databases.

As a result we obtained 11 ontologies for all medical registers that are included in this project. An example of ontology of Trauma/Injury Register is illustrated in Figure 2. These constructed ontologies are like data structures representing what kind of data will be stored in the Medical Data Warehouse and how classes are linked to each other. There can be seen classes, like Person, their attributes, like date of

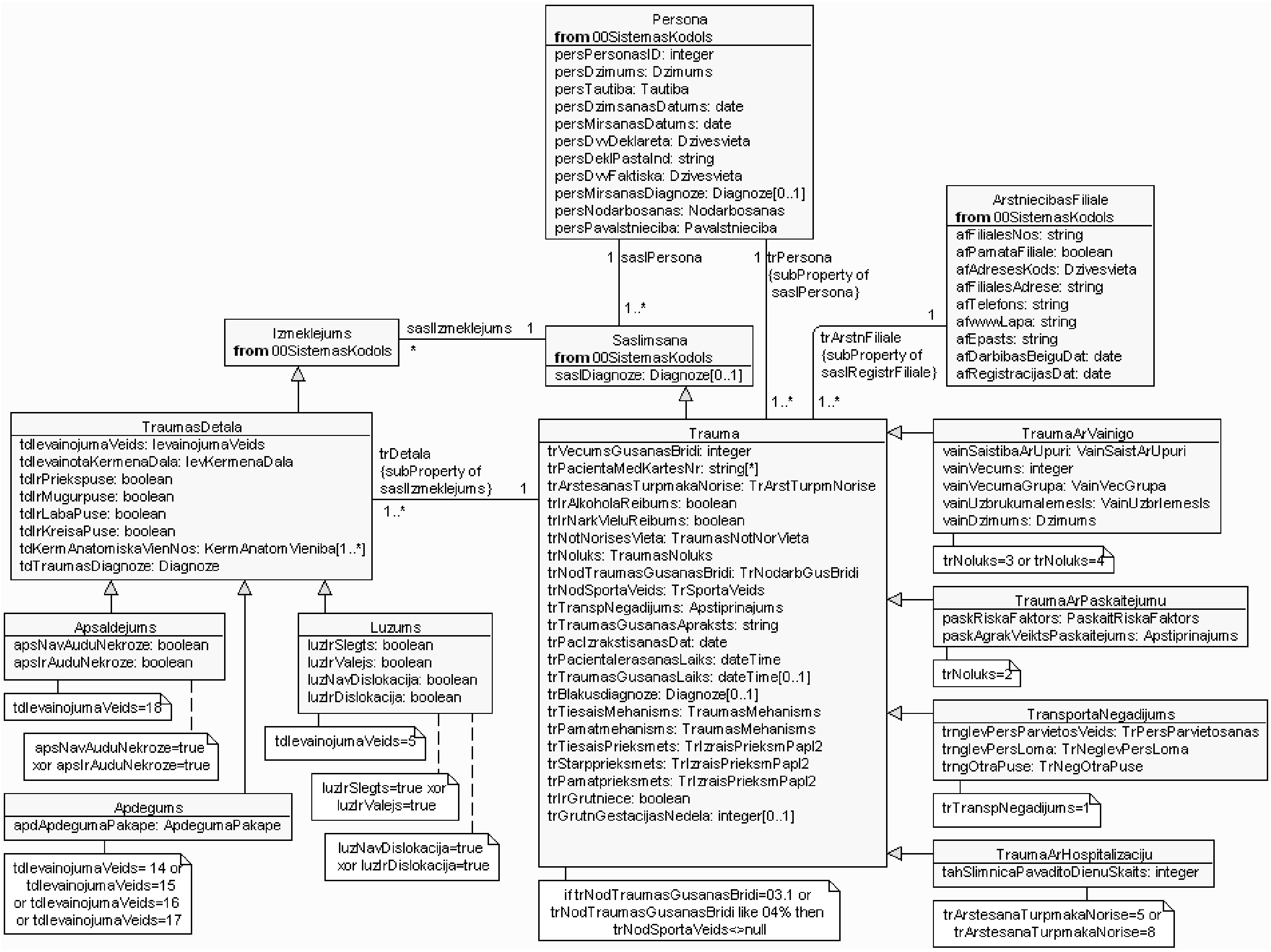

Fig. 2. Ontology of Trauma/Injury Register.

birth, data types of attributes (integer, string, date, etc.), and relationships (associations) between classes, as between Person and Injury. Some of associations have a triangle symbol, which indicates that one of the classes is a subclass of another, for example, Traffic Accident is a subclass of Injury. This is a so called generalisation. All associations have cardinality (except generalisation), for example, cardinality between classes Person and Injury is "1...*", which means that a person can have one or more injuries. We chose to use the enumeration data type for class attributes like nationality, profession, etc., as we know all the possible values for the corresponding attribute. All these types are depicted as separate classes in ontology only with a stereotype "<enum>". For some of <enum> classes potential values are shown, for others there are too many values to show in ontology.

After an ontology is built, the next task is to export data from original databases (legacy relation databases) to the Medical Data Warehouse (RDF database). In this step we explore data given in original databases and verify if ontologies were developed correctly. Then data from relation databases are translated to RDF triples using specifically crafted SQL queries. Consequently, these RDF data triples and corresponding OWL ontology are loaded into RDF a database — Medical data warehouse. This RDF database now can be accessed by medical researchers using our graphic ViziQuer Tool.

We have developed a graphical query language ViziQuer, which supports three main elements — class (concept), association (relation) and attribute, and a tool that supports the developed language. Using these elements the researcher constructs a graphical query. The query is then translated into SPARQL, executed, and results presented to the researcher as a table. A similar approach was used in Chen *et al*. (2006), but the main difference of our approach is support for graphical query construction.

To better understand the use of our query tool, we present a simple example. For instance, we might be interested in the correlation between injured persons age and if they where intoxicated. Also, we want to consider only persons born before 1 January 1989. First we analyse our query for involved concepts, their relationships and attributes. Here we can identify that we are talking about *persons* and *injuries* as concepts. We know that there is some relationship between them and that person date of birth is before 01.01.1989. We want to know if a person was intoxicated and how old he/she was when the accident happened. We

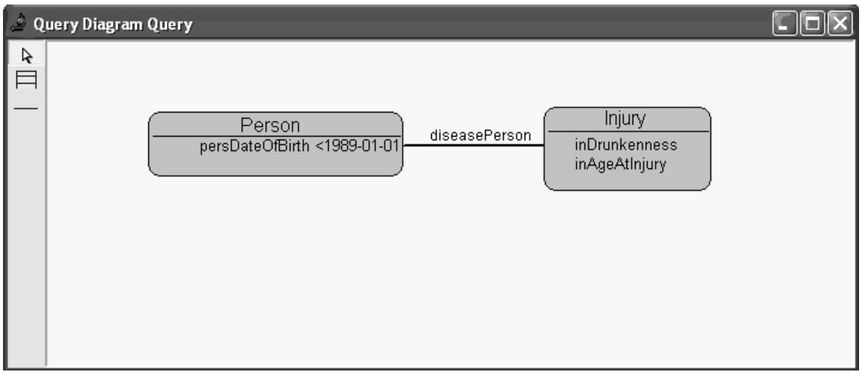

Fig. 3. Example query in ViziQuer.

can see the constructed query in Figure 3. After query construction, the researcher receives an answer table that can be exported to MS Excel for further analysis. The researcher does not even need to know that this picture is actually translated into the SPARQL query shown below:

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX med: <http://www.owl-ontologies.com/
Ontology1228562295.owl#>

PREFIX injury: <http://www.owl-ontologies.com/
Ontology1228576854.owl#>

PREFIX rdf: <http://www.w3.org/
1999/02/22-rdf-syntax-ns#>

select ?sa0 ?sa1 where

{?t0 rdf:type med:Person. ?t0 med:persDateOfBirth ?a0.

FILTER (?a0<xsd:date("1989-01-01")).

?t1 med:diseasePerson ?t0. ?t1 rdf:type injury:Injury.

OPTIONAL {?t1 injury:inDrunkenness ?sa0. }

OPTIONAL {?t1 injury:inAgeAtInjury ?sa1. } }

We have successfully implemented the first version of the described approach and it is successfully used by the medical researchers. Further improvements are possible, like adding aggregation functions in queries. For example, if we want to select only those diabetes patients who have at least five examinations, without aggregation functions, currently, we can select only those having at least one examination.

ACKNOWLEDGMENT

The work was supported by the National Research Programme in Medicine 2006?2009, project No. 14, „Creation of the unified and generally accessible data base on the main life expectancy and life quality threatening pathologies and epidemiology of their risk factors in Latvian population".

REFERENCES

Barzdins, G., Liepins, E., Veilande, M., Zviedris, M. (2008). Semantic Latvia approach in the medical domain. In: *Proceedings of the 8th International Baltic Conference on Databases and Information Systems*. Haav, H.M., Kalja, A. (eds.). Tallinn University of Technology Press, pp. 89–102.

Chen, H., Wang, Y., Wang, H., Mao, Y., Tang, J., Zhou, C., Yin, A., Wu, Z. (2006). Towards a semantic web of relational databases: a practical semantic toolkit and an in-use case from traditional chinese medicine. In: *5th International Semantic Web Conference. Lecture Notes in Computer Science*. Cruz, I.F., et al. (eds.). Springer, pp. 750–763.

Received 11 July 2009

MEDICĪNAS DATU BĀZU ONTOLOĢISKĀ PĀRBŪVĒŠANA

Aprakstīts datu eksports no vairākām medicīnas datu bāzēm (relāciju datu bāzēm) uz vienotu Medicīnas datu noliktavu (RDF datu bāzi, kas veidota atbilstoši integrētai OWL ontoloģijai). Eksportētajiem datiem var ērti piekļūt ar SPARQL vai grafisku vaicājumu valodu ViziQuer, kurai pamatā UML profils OWL valodai. Šī pieeja ilustrēta, izmantojot vienu no Latvijas Medicīnas datu bāzēm – Traumu reģistru.