A SEMANTIC WEB TECHNOLOGIES-BASED SYSTEM FOR CONTROLLING THE CORRECTNESS OF MEDICAL PROCEDURES IN POLISH NATIONAL HEALTH FUND

Jan Andreasik
Zamość University of Management and Administration, Zamość, Poland
Institute of Biomedical Informatics, University of Information Technology and Management in Rzeszów, Rzeszów, Poland

Andrzej Ciebiera, Sławomir Umpirowicz
Institute of Biomedical Informatics, University of Information Technology and Management in Rzeszów, Rzeszów, Poland

Keywords: Decision support system, Semantic web, Ontology, RDF, RDFS, OWL, SPARQL, SPIN.

Abstract: The correctness verification of the medical procedures described by enormous number of data stored in large, relational databases is very difficult. The database queries executed on redundant, incoherent, and contextless data may cause erroneous results. The Semantic Web technology has been considered to solve the problem. In this paper we propose the Semantic Web technologies-based system architecture for controlling medical procedures. We performed the set of inference tasks that confirmed the usefulness of the Semantic Web technology. The experience with Semantic Web modelling can be used for implementation of solutions to the contextual data integration and analysis.

1 INTRODUCTION

In the recent years, there has been intensive development of the semantic Web technologies towards the semantic analysis of data stored in large repositories. In many institutions and enterprises, large distributed databases based on relational database technologies such as ORACLE, Microsoft SQL Server and others, operate and are still extended. The data exploration process based on this model causes a series of issues that are related to preparation of reports. Moreover, the query technologies of distributed databases does not meet the user's requirements. The databases' users are not familiarized with the database query languages (SQL) because they usually are members of the administrative staff, marketing specialists in finance departments, and inspectors in various controlling offices. These specialists want to formulate queries in the natural language of the specific domain. They also want to obtain reports with data or information more concise than reports created on the basis of relationships between data. They want to extract the knowledge concerning certain issues creating the knowledge based system (KBS). Semantic Web (Hebeler et al., 2009) gives them such possibilities.

In this paper we present the architecture of system developed for the branch of the Polish National Health Fund (NHF) in Podkarpackie voivodship. It defines a list of issues including search processes of the NHF distributed databases. The main issue is related to the control of medical procedures. The NHF inspectors want to obtain the reports that respond to the specific and complex queries.

This kind of systems is under the development. Several systems, based on hospital repositories, and used for patients monitoring can be mentioned here. The HIWO (Hospital Ward Intelligent Ontology) system was presented by P. Katari, R. Juric, S. Paurobally, K. Madani (Kataria et al., 2008). HIWO ontology is shown in the tool TopBraid whereas conversion of data from the Oracle Express Edition 10g relational database to the RDF format is made with using the D2RQ tool, which uses the Jena API libraries. Another approach was proposed by P. LePendu, D. Dou, G.A. Frishkoff, J. Rong (LePendu et al., 2008), who formed the ontological base for the analysis of electroencephalographs. The data were collected in the MySQL RDBMS, rules were written in the SWRL

Andreasik J., Ciebiera A. and Umpirowicz S..
A SEMANTIC WEB TECHNOLOGIES-BASED SYSTEM FOR CONTROLLING THE CORRECTNESS OF MEDICAL PROCEDURES IN POLISH NATIONAL HEALTH FUND.
DOI: 10.5220/0003642103310336
In *Proceedings of the International Conference on Knowledge Management and Information Sharing* (KMIS-2011), pages 331-336
ISBN: 978-989-8425-81-2

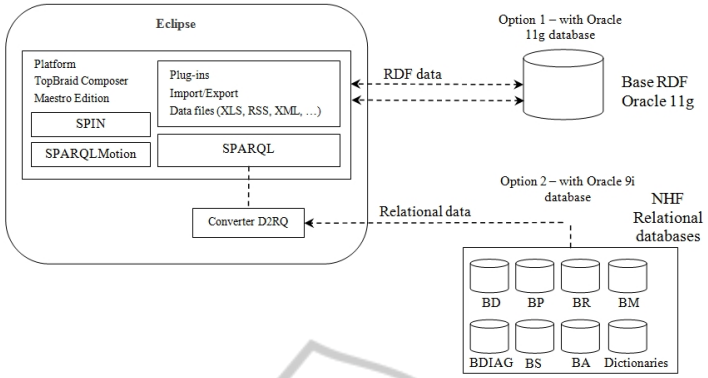

Figure 1: System architecture for controlling the correctness of medical procedures.

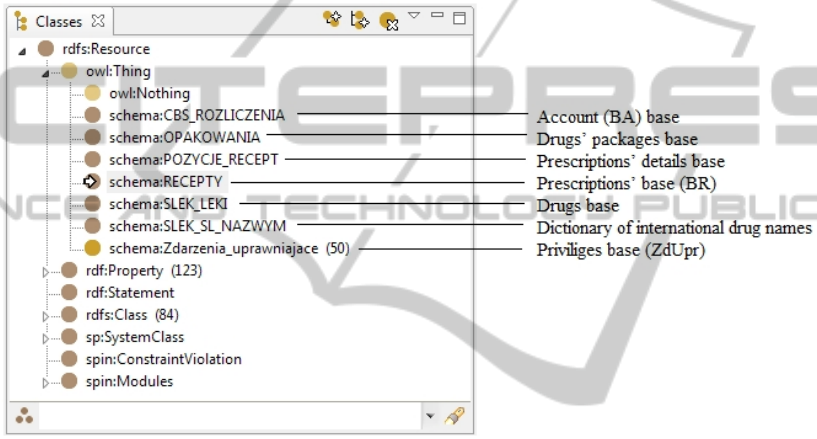

Figure 2: Controlling system class diagram in the TBC editor.

language and queries were constructed in the following languages: SPARQL, OWL-QL, and SQL. H. Chen (Chen et al., 2006) and co-workers developed a tool called Dartgrid for processing data from the relational database using the SPARQL query language. An application for the examination of Chinese medicine procedures based on China Academy of Traditional Chinese Medicine has been developed. The other approach is to create systems, designed in accordance with the principles of the SW technologies. Such systems are based on the medical ontologies such as UMLS (NLM, 2011), SNOMED (IHTSDO, 2011), and Galen (OpenGALEN, 2011). The system for medical procedure registering with the use of OWL was presented by A.L. Rector, R. Qamar, T. Marley (Rector et al., 2009). The medical information management system in a hospital, based on the following languages: OWL-S, OWL, SWRL, was presented by M.A. Casteleiro, J.J. Des Diz (Casteleiro and Des Diz, 2008). They have developed an interface between the UMLS thesaurus and the OWL language editor called Protege-OWL. The model of public health information network with use of the semantic modeling was developed by Mirhaji P, Casscells SW, Allemang D, Coyne R (Mirhaji et al., 2007).

2 SYSTEM ARCHITECTURE

At present, the usage of Semantic Web requires connection to databases that supports the RDF model storage or imports data into the RDF structures. Source data for the system for controlling the correctness of medical procedures is stored by the NHF in their relational Oracle databases. Therefore, to make connection to the data sources, the D2RQ (FUB, 2011) converter was used. We used the TopBraid Composer (TBC) Free Edition (Quadrant, 2011) as the SW modeling tool, which is the extended language editor for RDF, RDFS and OWL, and moreover, explorer of instances and a tool for performing the SPARQL queries with a graphical user interface. An additional advantage of TBC is the built-in rule engine SPIN (SPARQL Inferencing Notation), that

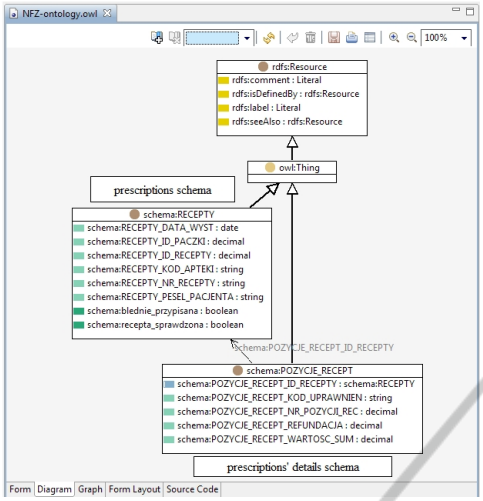

Figure 3: Organization of the class POZYCJE_RECEPT (prescriptions' details) related to the class RECEPTY (prescriptions).

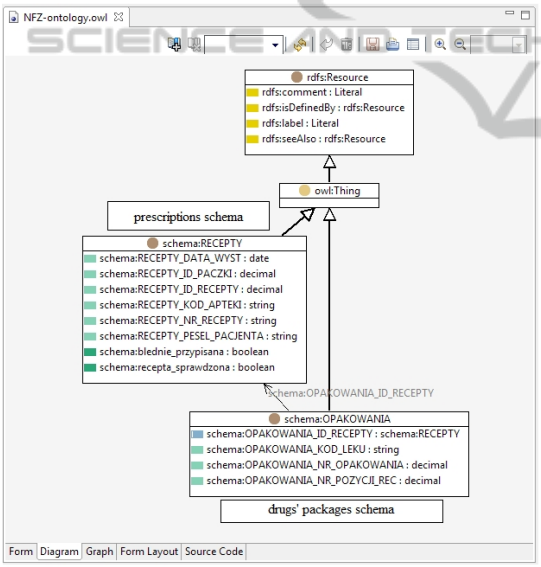

Figure 4: Organization of the class OPAKOWANIA (drugs' packages) related to the class RECEPTY (prescriptions).

makes the process of definition of additional conditions and rules in the constructed queries of the system easier.

In the future, when using Oracle 11g, the D2RQ tools can be removed. The set of the relational databases in NHF consists of over ten bases. For the experiment purposes, three databases were chosen: base of drugs, base of services and base of prescriptions. Each base is characterized by a large number of records (approximately 50 million records).

During the analysis of data based on the SQL queries, carried out by NHF, we identified inconsis-

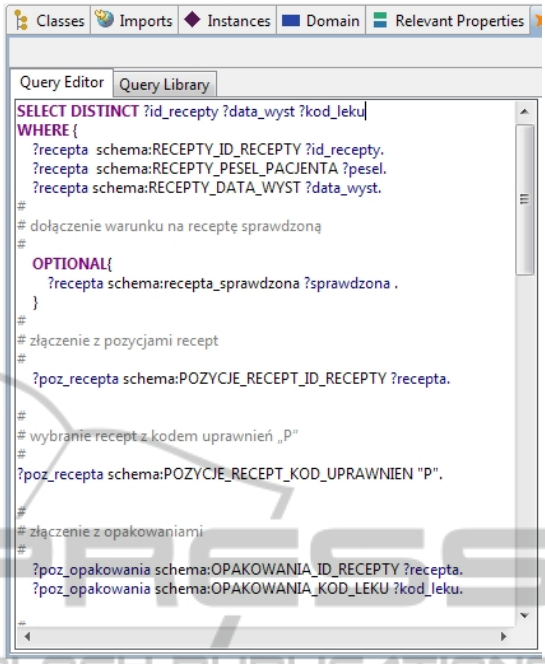

Figure 5: SPARQL query.

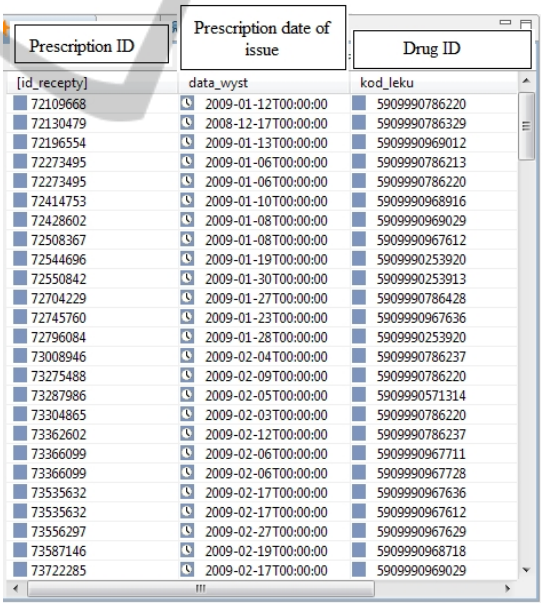

Figure 6: Recipes identifiers - the result of SPARQL query.

tencies in the source data. One of the objectives of the project was also to confirm the suitability of the Semantic Web data model for database integration. Source data for the controlling system were identified as:

- Doctors' base (BD),
- Patients' base (BP),

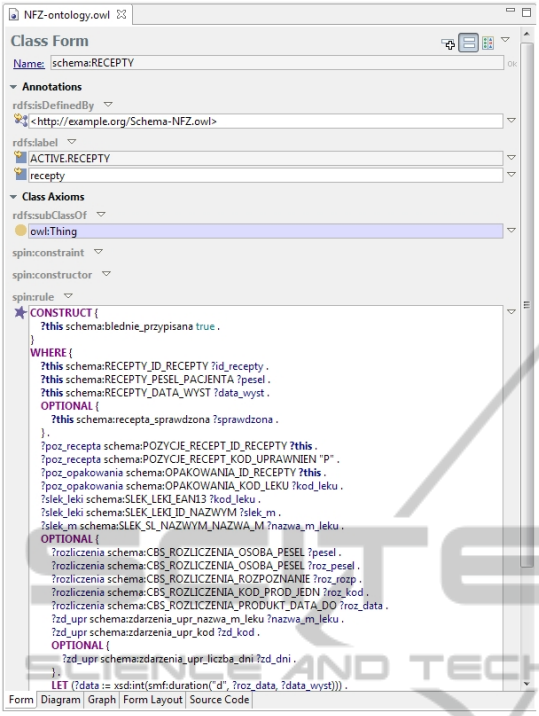

Figure 7: SPIN rule defined in the TBC.

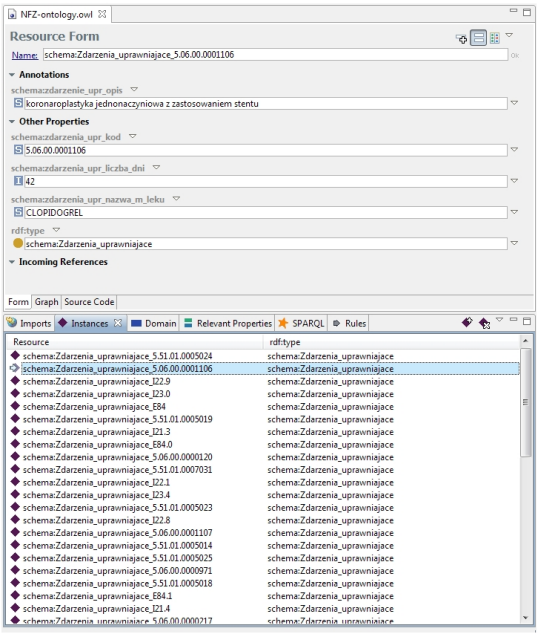

Figure 8: Instances of the class Zdarzenia_uprawniajace (ZdUpr) defined in the TBC.

- Prescriptions' base (BR),
- Medicines' base (BM),
- Diagnoses' base (BDIAG),
- Services' base (BS),
- Accounts' base (BA),
- a dictionary of international drug names (SLEK_SL_NAZWYM),
- the dictionary ICD 9 (International Classification of Medical Procedures),
- the dictionary ICD 10 (International Statistical Classification of Diseases and Related Health Problems).

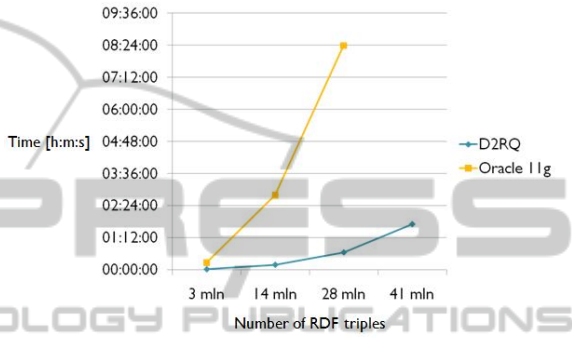

Figure 9: Processing times depending of the database access mode.

3 CONTROLLING SYSTEM ONTOLOGY

Because the controlling system was designed to process the existing source data, we applied organization of classes within the scope of structures and properties based on the structure and Polish labels of source data stored in the Oracle NHF database. This solution makes the implementation of classes and their properties in OWL complicated, but facilitates the work of NHF doctors (familiarity of existing structures) and ensures immutability of organization of existing data sources which is needed for other applications. To confirm the usefulness of the SW technology in the analysis of the medical procedures we selected two problems groups of medical procedures where drugs are used:

(i) Tramadol and Pancreatinum,

(ii) Clopidogrel.

These medicines may be prescribed on prescriptions in certain medical procedures (Proc) only with the corresponding privilege code (KodUpr) and these medicines are under supervised distribution. The difference between (i) and (ii) relays on the fact that the prescription for (i) has no time validity. The medicine is considered as a correct when it is prescribed at any

time - before, during and after the event giving the privileges (ZdUpr) for prescription. The prescription for (ii) is the case when the medicine is prescribed during the particular time depending on ZdUpr. An ontology of the system for controlling the correctness of medical procedures has been prepared in the TBC editor and it is presented in Figure 2. The prepared structure corresponds to structure organization currently used in the NHF data sources.

4 ANALYSIS OF MEDICAL PROCEDURES USING SPARQL AND SPIN IN THE CONTROLLING SYSTEM UNITS

Defined inference problems for (i) and (ii) require different search paths. For case (i), searching for the numbers of incorrect prescriptions (there is no ZdUpr record) after the modeling OWL needs only to define the query of the type SELECT in the SPARQL language. For case (ii), a query in SPARQL requires many additional conditions, as many as different periods that correspond to ZdUpr in the dictionary of events. For this group of problems, a queries of the Construct type in the SPARQL language we found useful in the definition of a rule. All the conditions defined by the medical inspectors were taken into consideration in the rule. A key SPIN features are the following:

- possibility of calculation of property values based on other properties,
- checking constraints and data validation,
- creating rule templates made under certain conditions.

Figure 7 shows a rule defined in SPIN. The structure of this rule allows creation of libraries of similar queries which will be important facilities for the medical inspectors.

For case (ii) in our analysis, we prepared the class Zdarzenia_uprawniajace (class for privileges to prescribe drugs) with properties: Zdarzenie_upr (event identifier), Zdarzenie_upr_kod (event code), Zdarzenie_upr_liczba_dni (the number of days just after the date of event within the drug can be applied), Zdarzenie_upr_nazwa_m_leku (international drug name), each of data type properties, which allow to refer to values of the ICD 9 and ICD 10 dictionaries. The model of the controlling system supported by TBC with embedded SPARQL and SPIN (Fuber and Hepp, 2011) allows to create the optimal structures of queries, and also allows further, flexible development of the system model with another structures, rules and queries. Furthermore, TBC allows the development of friendly interfaces for end users of the system, i.e., medical specialists.

5 PROCESSING TIME OF SEMANTIC DATA

After the system correctness confirmation the processing time tests were carried out with the relational database and the D2RQ converter and with the Oracle 11g database, which natively supports the storage of RDF data.

Configuration of the server:

- Hardware: Intel Core 2 Duo E7600, 8 GB RAM.
- Software: Windows Server 2003 R2, Oracle 11.1.0.6.0, TopBraid Composer ME 3.2.0.

Processing times depending of the database access mode shows the Table 1 and the Figure 9.

Table 1: Processing times depending of the database access mode.

Number of RDF triples	Time [h:m:s] D2RQ	Time [h:m:s] Oracle 11g
3 452 958	00:01:09	00:16:00
14 233 326	00:11:10	02:47:30
27 708 786	00:38:54	08:23:30
41 184 246	01:42:10	More than 24 hours

6 CONCLUSIONS

The result of our experiment is the achievement of a high degree of data integration in order to obtain the expected reports without any intervention to the existing distributed structure of the databases. The system for controlling the correctness of medical procedures model based on the TBC and D2RQ tools confirmed the usefulness of SW in the analysis of the medical procedures described by the records stored in the relational databases. The experience with controlling system modelling can be used to:

- extension of the developed model by new data structures, relations, inference rules, and the tools for result data visualization,
- sharing knowledge structures stored in the RDF, RDFS, OWL, and the SPIN / SPARQLMotion rules in description of the medical procedures,

- implementation of the SW tools based on a similar data model or analogical purposes of inference,
- implementation of solutions to the contextual data analysis.

REFERENCES

Casteleiro, M. and Des Diz, J. (2008). Clinical practice guidelines: A case study of combining owl-s, owl, and swrl. *Knowledge-Based Systems*, 21:247–255.

Chen, H., Wang, Y., Wang, H., Mao, Y., Tang, J., Zhou, C., Yin, A., and Wu, Z. (2006). Towards a semantic web of relational databases: a practical semantic toolkit and an in-use case from traditional chinese medicine. In *Proc. of the Fifth International Semantic Web Conference*, pages 750–763.

FUB (2011). The d2rq platform-treating non-rdf databases as virtual rdf graphs. http://www4.wiwiss.fu-berlin.de/bizer/d2rq/.

Fuber, C. and Hepp, M. (2011). Using sparql and spin for data quality. http://heppnetz.de/files/feurber-hepp-sparql-spin-dqm.pdf.

IHTSDO (2011). Snomed ct. http://www.ihtsdo.org/snomed-ct/.

Kataria, P., Juric, R., Paurobally, S., and Madani, K. (2008). Implementation of ontology for inteligent hospital wards. In *Proc. of the 41st Hawaii International Conference on System Sciences (HICSS 2008)*, pages 1–9. IEEE.

LePendu, P., Dou, D., Fishkoff, G., and Rong, J. (2008). Ontology database: A new method for semantic modeling and an application to brainwave data. In Ludacher, B. and Mamoulis, N., editors, *Proc. of the SSDBM 2008*, pages 313–330, Berlin Heidelberg. Springer-Verlag.

Mirhaji, P., Casscells, S., Allemang, D., and Coyne, R. (2007). Improving the public health information network through semantic modeling. *IEEE Intelligent Systems*.

NLM (2011). Umls. http://www.nlm.nih.gov/research/umls.

OpenGALEN (2011). http://www.opengalen.org/.

Quadrant, T. (2011). Topbraid composer. http://www.topquadrant.com/products/TB_Composer.html#free.

Rector, A., Qamar, R., and Marley, T. (2009). Binding ontologies and coding systems to electronic health records and messages. *Applied Ontology*, 4:51–69.