Semantic Search Engine as Tool for Clinical Decision Support in Register for Acute Coronary Syndrome

Pavle Kostić, Dejan Novičić, Suzana Blesić, Zorana Vasiljević, Jelica Grujić Milanović, Siniša Pavlović, and Slađan Milanović

Abstract – **This paper presents the implementation and use of Semantic Search Engine (SSE) as part of knowledge management system functionalities in Register for Acute Coronary Syndrome (REACS). REACS SSE is part of a clinical decision support system and is used as an aid in decision making in clinical processes related to the care and treatment of patients with Acute Coronary Syndrome (ACS).**

Keywords — **Semantic search engine, Register, Acute Coronary Syndrome, Knowledge Management, Cardiology.**

I. Introduction

SEMANTIC search seeks to improve the accuracy of search results by understanding the intention and contextual meaning of the term that appears in the search knowledge-domain, regardless of whether on the Web or within a closed system, generating more relevant search results [1]. Starting from the definition of semantic search and bearing in mind the needs of physicians in the coronary units (CUs), to enable them to easily find similar cases from the Registry of Acute Coronary Syndrome (REACS), this paper describes a semantic search engine (SSE) as part of the comprehensive knowledge management system for support in clinical decision making in real-time and real-life scenarios and clinical processes related to the diagnostics, care and treatment of patients with ACS. Following is an overview of basic concepts and definitions related to the ACS and the REACS. We describe the relationship of guides to good clinical practice and clinical support system for decision-making, as well as connections with knowledge management systems in which the semantic search engine was implemented. The paper describes the technical platform on which a semantic search engine is implemented, semantic and functional requirements, as well as the way of realization of semantic search application.

Finally, some conclusions on the possibility of extended usage of semantic search engine in similar applications are derived.

A. Acute Coronary Syndrome

ACS is one of the quite frequent diseases that accounts for about 50% of all cardiovascular disease. ACS includes the spectrum of clinical presentations which, according to the clinical findings, electrocardiographic ST-segment image and to the presence of cardiospecific enzymes, can be regarded as Acute Myocardial Infarction (AMI) with ST elevation (STAIM), AMI without ST elevation (NSTAIM) or as Unstable Angina Pectoris (NAP) [2] - [3]. Timely diagnosis and treatment with modern methods have greatly increased the prognosis in patients with ACS [4]. Thus, enabling better and faster clinical decision making in accessing the patient status and coming to accurate diagnosis in a shorter time is the assumption for the improvement of treatment and ultimately the increase of survival rate, reducing the level of disability and increasig the working capacity of patients in real-world scenarios.

B. Good Clinical Practice Guideliness and Systems for Clinical Decision-making Support

In many medical domains are defined Good Clinical Practice Guidelines (GCPG), mainly as "systematically developed instructions to help doctors and patients in deciding about appropriate health care for specific clinical circumstances" [5]. While computerized GCP are considered critical components of evidence-based

This project was funded by the Ministry of Science and Technology of the Republic of Serbia, technology development project No. TR13010-A, entitled "System for knowledge management - applications in cardiology."

Pavle Kostić, University of Belgrade, Institute for Medical Research, Laboratory of Biomedical Engineering and Telemedicine, PO BOX 102, 11129 Belgrade, Serbia (phone: 381-11-2685788, e-mail: pkostic@imi.bg.ac.rs)

Dejan Novičić, University of Belgrade, Institute for Medical Research, Laboratory of Biomedical Engineering and Telemedicine, PO BOX 102, 11129 Belgrade, Serbia (phone: 381-11-2685788, e-mail: dnovicic@imi.bg.ac.rs)

Suzana Blesić, University of Belgrade, Institute for Medical Research, Laboratory of Biomedical Engineering and Telemedicine, PO BOX 102, 11129 Belgrade, Serbia (phone: 381-11-2685788, e-mail: suzana@imi.bg.ac.rs)

Zorana Vasiljević, University of Belgrade, Faculty of Medicine, Department of Internal Medicine, Dr Subotića 8, 11000 Belgrade, Serbia (phone: 381-11-3615552, e-mail: zoranav@eunet.rs)

Jelica Grujić Milanović, University of Belgrade, Institute for Medical Research, Laboratory of Cardiovascular Physiology, PO BOX 102, 11129 Belgrade, Serbia (telefon: 381-11-2685788, e-mail: jeca@imi.bg.ac.rs)

Siniša Pavlović, University of Belgrade, Faculty of Medicine, Department of Internal Medicine, Dr Subotića 8, 11000 Belgrade, Serbia (telefon: 381-11-3615617, e-mail: pavlosini@yahoo.com)

Slađan Milanović, University of Belgrade, Institute for Medical Research, Laboratory of Biomedical Engineering and Telemedicine, PO BOX 102, 11129 Belgrade, Serbia (telefon: 381-11-2685788 e-mail: sladjan.milanovic@imi.bg.ac.rs)

medicine [6] - [8] their true potential has yet to be realized [9].

The main challenge is the integration of the GCPG models in real-world clinical business processes to generate patient-centric personalized recommendations and actions instead of generalized theoretical assumptions.

Systems to support clinical decision making based on GCPG appear as a paradigm for help in reducing errors, reducing costs and improving quality in evidence based medicine.

The functions of the system to support clinical decision making can be categorized in the areas of: context, knowledge and data sources, decision support systems, information delivery and work processes [10].

For successful implementation in a complex environment of health information systems and particular eHealth systems, computerized GCPG and other related technologies to support clinical decision making are considered to be the foundation platform for a knowledge management system that includes also well-defined medical informatics standards (i.e. for terminology (*Systematic Nomenclature of Medicine - Clinical Terms* (**SNOMED-CT**); *Logical Observations, Identifiers, Names, and Codes* (**LOINC**); *Unified Medical Language System* (**UMLS**); *General Architecture for Languages, Encyclopedias and Nomenclatures in Medicine* (**GALEN**)), exchange of information (*Health Level Seven* (**HL7**) for general health information; *Digital Imaging and Communications in Medicine* (DICOM) for medical images, ISO / IEEE 11073 Medical / health device communication standards etc.) and the Electronic Health Record platforms for data collection and dissemination.

C. *ACS Knowledge Management System*

ACS Knowledge Management System is developed and implemented comprising of several technologically and functionally integrated modules:

1. Register (with an accompanying database and end-user application) for Acute Coronary Syndrome (REACS) for patients and episodes of care data collection,
2. Business Intelligence and Analytics module for statistical data analysis and research and
3. Semantic search module for real-time and real-life clinical decision support.

All modules are exposed through the KardioNet internet portal (http://www.kardionet.org/) accessible to both patients and doctors [11].

In this paper we focus on a Semantic Search module which is implemented and functioning over the REACS patients and episodes of care database with imported data period from 2002 to 2008 and on-line data entry from 2009 to 2011, spread over the territory of Serbia. This data volume is significant and can be considered as highly relevant for patient-centric case comparison and knowledge extraction.

II. Research Aims And Application Goals

The main research aim in the development of Semantic Search module in the context of REACS was to enable doctors in CUs to achieve easier and more efficient search on similar (comparable) clinical cases, and standardization of the quality and effectiveness of treatment in CUs. Thus, by enabling faster but more robust case comparison and filtering in the presentation of most relevant results to doctors, the usability of this system can be aligned with real-time needs in clinical processes typical for CUs.

The main functional requirements and application goals were defined as:

- Developing a semantic meta-model for representation of medical/clinical knowledge contained in the registry
- Developing a suitable and efficient User Interface (UI, portlets) for the management of meta-model of knowledge and semantic search.

III. REACS Semantic Search Engine

For the development and implementation of REACS SSE we selected D2RQ platform with accompanying tools and components available for Java platform. Details on development and implementation are described in the following sections.

A. *Jena framework*

Jena is an open source Java framework for building Semantic Web applications. It provides a programmatic environment for RDF, RDFS and OWL, SPARQL (RDF query language) and includes a rule-based inference engine.

The Jena Framework consists of Java libraries and packages, ready made code snippets and tools for parsing, creating and querying RDF graphs and knowledge models, including:

- A RDF API,
- Reading and writing RDF in RDF/XML, N3 and N-Triples
- An OWL API
- In-memory and persistent storage
- SPARQL query engine.

The actual version 2.8.8, which we used, is freely available for download on Internet [12].

B. *SPARQL*

RDF is a directed, labeled graph data format for representing information in the Web. SPQRQL [13] is the query language and protocol used for working with RDF which was designed by the W3C RDF Data Access Working Group. As a SPARQL query language is oriented to the data (data-oriented), it can only execute queries against data stored in the model. In the query language, there is no reasoning mechanism (inference). Of course, the Jena model can be "smart" by creating RDF triples on demand mechanism using OWL reasoning. SPARQL is thus utilized to execute what is defined in the application in the form of a query and returns a set of information in the form of a RDF graph.

C. D2RQ platform

The D2RQ platform [14] consists of:

- the D2RQ Mapping Language, a declarative mapping language for describing the relation between an ontology and a relational data model.
- the D2RQ Engine, a plug-in for the Jena and Sesame Semantic Web toolkits, which uses the mappings to rewrite Jena and Sesame API calls to SQL queries against the database and passes query results up to the higher layers of the frameworks.
- D2R Server, an HTTP server that can be used to provide a Linked Data view, a HTML view for debugging and a SPARQL Protocol endpoint over the database.

Fig. 1 depicts the architecture of the D2RQ Platform.

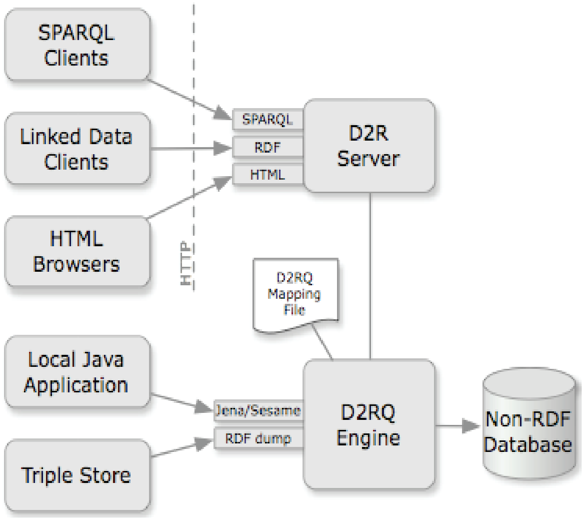

Fig.1. D2RQ platform architecture.

The D2RQ Engine is implemented as a Jena graph, the basic information representation object within the Jena framework. A D2RQ graph wraps a local relational database into a virtual, read-only RDF graph. It rewrites Jena or Sesame API calls, find() and SPARQL queries to application-data-model specific SQL queries. The result sets of these SQL queries are transformed into RDF triples or SPARQL result sets that are passed up to the higher layers of the framework. The D2RQ Sesame interface wraps the D2RQ Jena graph implementation behind a Sesame RDF source interface. It provides a read-only Sesame repository interface for querying and reasoning with RDF and RDF Schema.

D2R Server is a tool for publishing relational databases on the Semantic Web. It enables RDF and HTML browsers to navigate the content of the database, and allows applications to query the database using the SPARQL query language. D2R Server builds on the D2RQ Engine. For detailed information on how to set up D2R Server please refer to the separate D2R Server website.

D. Details on SSE implementation

For the implementation of semantic search engine (SSE), we used the previously described technology. We used the Jena framework as a basis for knowledge representation located in REACS database. To be able to use data from an existing database and not to have to make a new database that would be suitable for use in semantic applications, we used D2RQ platform. It has enabled us to create a new layer above the existing relational database that allows us to search the database as it is a read-only graph (important to preserve original data and to enable real-time data collection and management). In order to be able to use that layer, we had to first create a mapping of source database model (entity-relational) to a semantic graph model. Within D2RQ platform, there is a tool that facilitates the very extensive work mapping tables from the database and their connection to classes in the semantic model (using D2RQ Mapping Language).

As a result of the mapping, a mapping file was obtained (asc-mapping.n3). As an illustration, in Fig. 2 is shown a part of the mapping that relates the table EPISODEOFCARE with semantic concepts.

Mapping file is an essential element for D2RQ platform concept. Based on this file, the application is able to monitor and search a relational database as if it comes to a read-only graph.

After mapping the relational database, we got a virtual graph that we are able to crawl using the SPARQL query language. Fig. 3 shows the graph segment referring to the episode of care (ACS cases).

```
# Table ACS.EPISODEOFCARE
<jdbc:oracle:thin:@imiserver:1521:imidb/vocab#ACS.EPISODEOFCARE> a d2rq:ClassMap;
  d2rq:dataStorage map:database;
  d2rq:uriPattern
"jdbc:oracle:thin:@imiserver:1521:imidb/vocab#ACS.EPISODEOFCARE/@@ACS.EPISODEOFCARE.ID@@";
  d2rq:class <jdbc:oracle:thin:@imiserver:1521:imidb/vocab#ACS.EPISODEOFCARE>;
  d2rq:classDefinitionLabel "ACS.EPISODEOFCARE";
  .
<jdbc:oracle:thin:@imiserver:1521:imidb/vocab#ACS.EPISODEOFCARE__label> a d2rq:PropertyBridge;
  d2rq:belongsToClassMap <jdbc:oracle:thin:@imiserver:1521:imidb/vocab#ACS.EPISODEOFCARE>;
  d2rq:property rdfs:label;
  d2rq:pattern "ACS.EPISODEOFCARE #@@ACS.EPISODEOFCARE.ID@@";
  .
map:EPISODEOFCARE_ID a d2rq:PropertyBridge;
  d2rq:belongsToClassMap <jdbc:oracle:thin:@imiserver:1521:imidb/vocab#ACS.EPISODEOFCARE>;
  d2rq:property vocab:EPISODEOFCARE_ID;
  d2rq:propertyDefinitionLabel "EPISODEOFCARE ID";
  d2rq:column "ACS.EPISODEOFCARE.ID";
  d2rq:datatype xsd:decimal;
```

Fig. 2. Sample D2RQ mapping of EPISODEOFCARE.

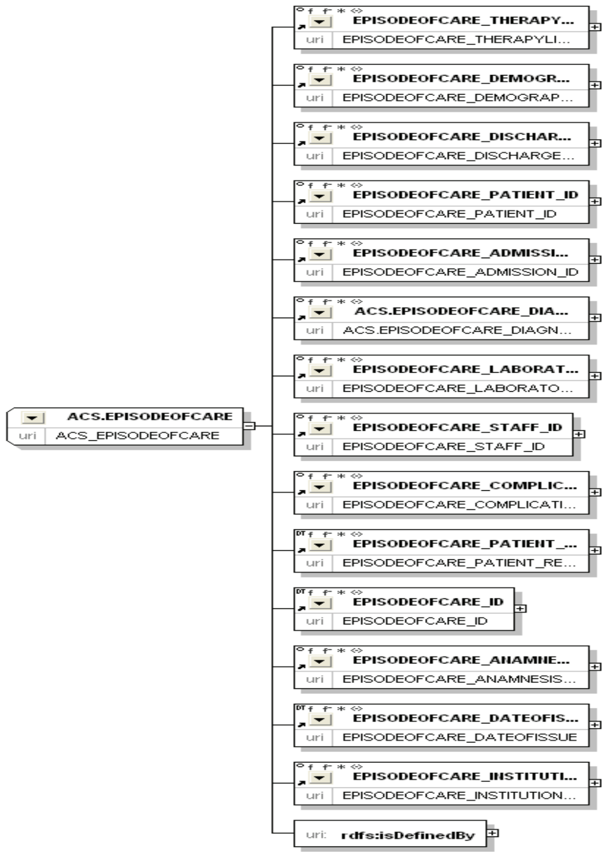

Fig 3. Graph segment referring to episode of care.

On the basis of SPARQL query language specification, we made a set of predefined queries for the most common usage scenario - retrieval of patients with given similar clinical conditions in their cases e.g. episodes of care. This scenario enables physicians to compare patient status with previous cases (patients), learn from their similarity and compare possible outcomes of different diagnostic / treatment possibilities (decision making support).

IV. Results

Time to the treatment for symptoms of ACS can be a matter of life and death. Semantic search enables a physician to quickly review recorded outcomes for similar patients (similarity by a number of parameters – search criterion, Fig. 5) when the two most common treatment therapies (Thrombolytic or PCI) are applied. A decision on which therapy is to be applied is very complex, and is mostly driven by guidelines. Still, guidelines are designed around statistically significantly distinguishable cases (patients selected for inclusion by a well defined criterion). All guidelines, also, provide the option of a different expert's opinion, in cases not entirely standard. Thus, patients with uncommon symptoms or physical and demographic statuses, or when such statuses are not known (which is not so rare a situation in emergency units) are not easy to triage by non-specialized caregivers. In such situations, when a physician has doubts in decision making using guidelines, semantic search as part of a knowledge dissemination system can help in the reevaluation of initial assumptions by allowing the physician to compare possible outcomes for similar cases with his/her assumptions. So, semantic search in REACS is not intended to be a "blind guess" decision making tool but to help in the reevaluation of assumptions in atypical situations.

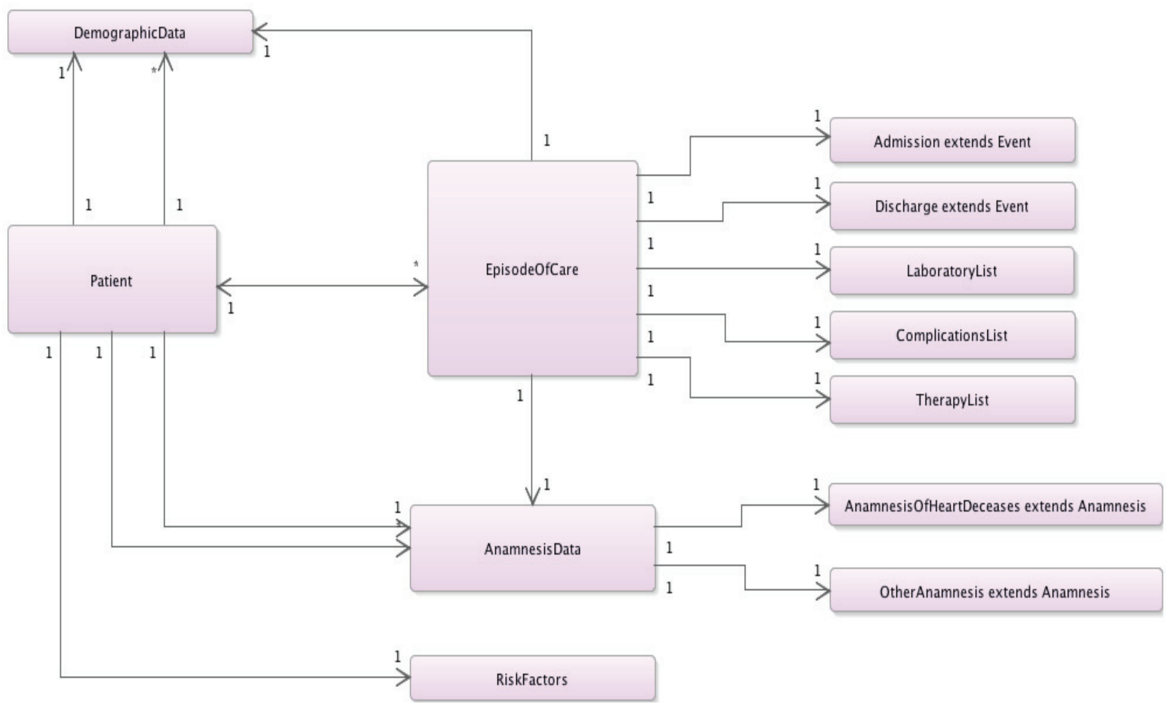

Fig. 4. REACS physical data model.

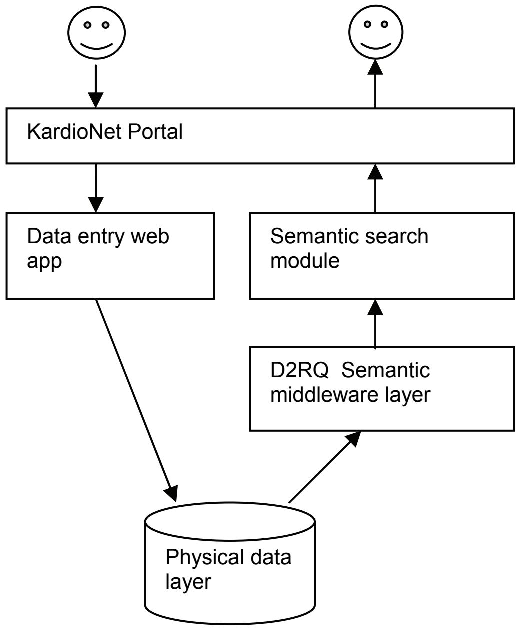

Fig. 5. REACS semantic mapping model.

This approach is different from other decision making support tools by not relying on statistical interpretations or quantifications but allowing a physician to improve decision making by a direct insight into personalized medical records of similar cases. This approach emphasizes personalized care derived from evidence based medicine.

A. REACS physical and semantic model

The basis of REACS physical data model, which was made on the basis of Coronary Questionnaire (CQ) [15], is *Patient-EpisodeOfCare* relationship which represents the domain ontology for ACS. The model is made respecting the relevant medical informatics standards (ATC / DDD, CDISC, HL7 ICD 10, CARDS ACS / CCU (**C**ardiology **A**udit and **R**egistration **D**ata **S**tandards-**A**cute **C**oronary **S**yndrome / **C**oronary **C**are **U**nit)) [16] and is shown in Fig. 4. This model deals with domain ontologies and taxonomies recommended by expert bodies to semantically model concepts and related relationships applied on the domain of ACS and related CU clinical processes. This model includes several information dimensions, of both clinical and demographic data, interrelated to both chronological / time-variant (episode of care domain) and non-chronological / time-invariant (risk factors, patient biological profile etc.) information.

REACS semantic model (Fig. 5.) is derived from a physical data model by semantic middleware (through D2RQ platform, as explained in the previous section), which provides mapping of underlying domain ontologies into upper semantic model based on graph representation of ACS cases (episodes of care). Upper ontologies represent not only the episode of care but also a demographic and personal health profile of people (patients) related to prescribed therapy and applied methods of treatment. Such a semantic model is closer to the perception of physicians which are neither used nor willing to cope with a complex entity-relational domain

.

Epizoda posete	Otpust	Dijagnoza	Fibrinoliza	Urgentna PTCA	Pacijent	Datum rodjenja	Pol	Hipertenzija	Diabetes melitus
61666	Oporavljen	Infarkt miokarda sa ST-elevacijom	Ne	Da	61666	08.08.1954.	Muški	Da	Ne
71885	Oporavljen	Infarkt miokarda sa ST-elevacijom	Ne	Da	71885	13.01.1956.	Muški	Da	Ne
65723	Oporavljen	Infarkt miokarda sa ST-elevacijom	Ne	Da	65723	12.04.1947.	Muški	Da	Ne
75412	Oporavljen	Infarkt miokarda sa ST-elevacijom	Ne	Da	75412	12.11.1944.	Muški	Da	Ne
66576	Oporavljen	Infarkt miokarda sa ST-elevacijom	Ne	Da	66576	04.05.1956.	Muški	Da	Ne
72503	Oporavljen	Infarkt miokarda sa ST-elevacijom	Ne	Da	72503	18.07.1942.	Muški	Da	Ne
67179	Oporavljen	Infarkt miokarda sa ST-elevacijom	Ne	Da	67179	30.06.1938.	Muški	Da	Ne
67768	Oporavljen	Infarkt miokarda sa ST-elevacijom	Ne	Da	67768	20.06.1937.	Muški	Da	Ne

Fig. 6. User interface of semantic search module (episodes of care with accompanying clinical and demographic information relevant to given criterion).

model in everyday emergency decision making. Thus, an upper semantic model (RDF-graph based) enables flexibility in meeting the needs of simplification of decision making queries (through SPARQL) while data collection is maintained in a standards-based model (entity-relationship based) which enables interoperability.

B. REACS Semantic Search application

The application (REACS module) for semantic search is composed of:

- Input parameters web form
- Display of search results.

Search parameters are comperhensive, covering several domains of health / medical information related to:

1. demographic data (date of birth / age, sex,...)
2. risk factors (family heart and blood vessels deseasses history, smoking status, hyperlipoproteinemia, hypertension, diabetes mellitus etc.)
3. medical history (diagnosis of acute coronary syndrome, therapy during hospitalization)
4. the outcome.

An important property of developed semantic search engine is the ability to sort search results in a manner that promotes results most similar to a specified criterion. In practice, this means that the most relevant (similar) cases are presented on the top of the result list, allowing the physician to quickly scan and review the most relevant information for further decision making processing, which is of high importance for timely providing an optimal ACS treatment [17]. We can also agree that the interest of the patient and the experts is not limited only to the immediate outcome of treatment then, in future projects, this kind of system is possible to upgrade with long-term follow-up data on treatment outcomes.

V. Conclusion

Recent research and applications of semantic technologies are showing a trend of rising applicability in a similar manner. As shown in [18], improved search and data integration are emerging areas of semantic technology applications in healthcare and medical informatics. By combining both of these functionalities in our REACS semantic search module, we are expecting to achieve most beneficiary effectiveness.

This is one of the first attempts in our country to build a clinical knowledge management system despite unavailable Electronic Health Record system. The potential of this approach will be explored in the future, especially by comparing it with existing systems that were developed around the world [10], [19] - [21].

Acknowledgement

Authors owe special gratitude to members of the Expert Team for Acute Coronary Syndrome, Ministry of Health of the Republic of Serbia and Cardiac Care Units Association of Serbia. We acknowledge the support of the Ministry of Education and Science of the Republic of Serbia project TR13010A.

References

[1] Semantic search http://en.wikipedia.org/wiki/Semantic_search/ (accessed July 2011.)

[2] F. Van de Werf, D. Ardissino, A. Betriu, et al. Management of acute myocardial infarction in patients presenting with ST-segment elevation. Eur Heart J 2003; 24:28-66.

[3] J. Bassand, C. Hamm, D. Ardissino et al. Guidelines for the diagnosis and treatment of non-ST-segment elevation acute coronary syndrom. Task Force for the Diagnosis and Treatment of non-ST-Segment Elevation Acute Coronary Syndromes of the European Society of Cardiology. Eur Heart J 2007; 14:1-63.

[4] Z. Vasiljević, "Acute coronary syndrome: Pathophysiologic mechanism classification and clinical presentation," *Acta Clinica* 2006;6(1): 29-36. (in Serbian).

[5] M. J. Field, K. N. Lohr, editors. Guidelines for Clinical Practice: From Development to Use. Washington, DC: National Academy Press; 1992.

[6] M Peleg, T. SW, J. Bury, P. Ciccarese, J. Fox, R. A. Greenes, et al. "Comparing Computer-Interpretable, Guideline Models: A Case-Study Approach," *J Am, Med Inform Assoc* 2003;10(1):52–68.

[7] D. Wang, M. Peleg, S. Tu, A. Boxwala, R. Greenes, V. Patel, "Representation primitives, process models and patient data in computer-interpretable clinical practice guidelines: A literature review of guideline representation methods," *Int J Med Inform* 2002;68(1-3):59-70.

[8] P. A. de Clercq, J. A. Blomb, H. H. M. Korsten, A. Hasman, "Approaches for creating computer-interpretable guidelines that facilitate decision support," *Artif Intell* Med 2004;31:1—27.

[9] F. A. Sonnenberg, C. G. Hagerty "Computer-interpretable clinical practice guidelines: Where are we and where are we going?" In: Kulikowski C, Haux R, editors. IMIA Yearbook of Medical Informatics 2006. Methods Inf Med 2006;45 Suppl 1: 145-58

[10] A. Berlin, M. Sorani, I. Sim "A taxonomic description of computer-based clinical decision support systems," *J Biomed Inform* 2006;39(6):656-67.

[11] Kardionet portal http://www.kardionet.org/ (accessed July 2011.)

[12] Jena – A Semantic Web Framework for Java http://jena.sourceforge.net/ (accessed July 2011.)

[13] SPARQL Query Language for RDF - http://www.w3.org/TR/rdf-sparql-query/ (accessed July 2011.)

[14] The D2RQ Platform - Treating Non-RDF Databases as Virtual RDF Graphs http://www4.wiwiss.fu-berlin.de/bizer/d2rq/spec/ (accessed July 2011.)

[15] Z. Vasiljević, B. Stojanović, N. Kocev, et al. „Hospital Mortality Trend Analysis of Patients with ST Elevation Myocardial Infarction in the Belgrade Area Coronary Care Units," Srp Arh Celok Lek 2008;136(2):84-96 (in Serbian).

[16] Cardiology Audit and Registration Data Standards - Acute Coronary Syndrome / Coronary Care Unit http://www.escardio.org/Policy/Documents/CARDS-dataset-ACS-071004.pdf (accessed July 2011.)

[17] W. F. Peacock, "Time to treatment and acute coronary syndromes: bridging the gap in rapid decision making," *Rev Cardiovasc Med.* 2010;11 Suppl 2:S45-50. Review

[18] V. Janev, S. Vranes, "Applicability assessment of Semantic Web technologies, " *Information Processing & Management*, Volume 47, Issue 4, July 2011, Pages 507-517.

[19] C. C. Chen, K. Chen, C. Y. Hsu, Y. C. Li, "Developing guideline-based decision support systems using protégé and jess," *Comput Methods Programs Biomed*. 2011 Jun;102(3):288-94.

[20] J. S. Ash, D. F.Sittig, R. Dykstra A. Wright, C. McMullen, J. Richardson, B. Middleton, "Identifying best practices for clinical decision support and knowledge management in the field," *Stud Health Technol Inform*. 2010;160(Pt 2):806-10.

[21] N. C. Hsieh, C. Y. Chang, K. C. Lee, J. C. Chen, C. H. Chan, "Technological Innovations in the Development of Cardiovascular Clinical Information Systems," *J Med Syst.* 2010 Jul 23. [Epub ahead of print]