Gathering Experience with Ontology and SPARQL Based Decision Support

Jan Pettersen Nytun[a], Tian Zhao[a], Angelique Mukasine[a], Rune Fensli[a]

[a]*Department of ICT, University of Agder, Norway*

Abstract

Use of knowledge-based decision support systems in medicine may improve the quality of medical care. We have executed several projects at UIA involving semantic web technologies like OWL, SPARQL and Jena; the application area has often been E-health; among these projects are two recent master student projects investigating the use of OWL and SPARQL to realize two different decision support systems. One project concerned monitoring vital signs of a patient estimating the state of the patient using a score system; recommendations based on score and additional rules were given. The other project defined an ontology storing medical patient information, with focus on diabetes and recommendations for diabetes patients; recommendations could be retrieved by querying the ontology. The projects used the Protégé framework when doing the development; this implied limitations that made the development cumbersome. The paper proposes another approach based on the Jena framework. The projects are analyzed in regard to technology; extensions and alternative solutions are discussed and proposed. The paper describes and recommends technology that may be used to build an advance medical knowledge-based decision support system.

Keywords: eHealth, knowledge-based decision support system, Ontology, OWL, SPARQL, Jena

Introduction

Decision support should be available to health care professionals (e.g. stated in report to Stortinget: St. meld.nr.9, 2012-2013, Én innbygger – én journal) and it may also play a role in supporting patients at home. Spring 2014 two master thesis projects [1,2] have targeted this research field. The focus of this paper has been on technological aspects and we place the students work in a wider context together with recommendations for further work.

The work of Tian Zhao was associated to a project at Sørlandet Sykehus Kristiansand, a Norwegian hospital. The project is called Tidlig Identifisering av Livstruende Tilstander (TILT) (some related work: [3]). In English this would be something like: Early Identification of Life-Threatening Conditions. By continuasly monitoring vital signs of a patient, together with application of rules, a state is "calculated" and interventions proposed. The task was then to propose a software solution (or parts of a solution) to this.

The work of Angelique Mukasine was the making of ontology for supporting patients at home, with focus on diabetes.

The remainder of the paper is structured as follows: The Method section briefly outlines description logic and the use of ontologies to represent knowledge; a general overview of a knowledge-based decision support system is delineated; some ontology technologies are also presented. The Result section presents the two master projects. The section called Discussion and Some Conclusions sees the projects in light of the greater picture presented in the Method section; limitations and problems are discussed; some conclusions are drawn. Section Recommendations for Further Work advocates some technologies and solutions.

Methods

Creation of fully envisioned systems was not within reach for the two master projects. Simplifications, focusing on only some technology and "a proof of concept approach" have been necessary. Figure 1 gives a general picture of our wider vision where the reported projects considered only a few components.

We will first describe the parts of the general picture. We apply description logic [4] where the knowledge base is composed of a terminology part defining concepts (also called classes) like *vital sign*, *blood pressure*, and *patient*. Descriptions of individual patients, etc., are found in the assertion part. Correspondingly, properties (also called roles) are described in the terminology part and applied in the assertion part, e.g., property *has blood pressure* relates *patient* with *blood pressure*; the assertion part may contain information about the blood pressure of some specific patient. This is the classical model (terminology) and model instance (assertions) separation. However, not all entities in the assertion part may have been classified and this differs from ordinary object-oriented program development.

The concepts can be categorized as *atomic* or *complex*; complex concepts are (unlike atomic concepts) built with the help of *description logic constructors* and atomic concepts. E.g., the following is the definition of a complex concept: *Potential diabetic ketoacidosis is equivalent to a patient with excessive thirst and frequent urination and confusion and,* etc. Here *Potential diabetic ketoacidosis* is the complex concept being defined; *is equivalent* and also *and* illustrate description logic constructs.

The complex concept may be used by a *reasoner* to classify individuals by inference. I.e., individuals satisfying the following belong to this complex concept: Must be of type patient, have excessive thirst, etc. Reasoners can also be used to check

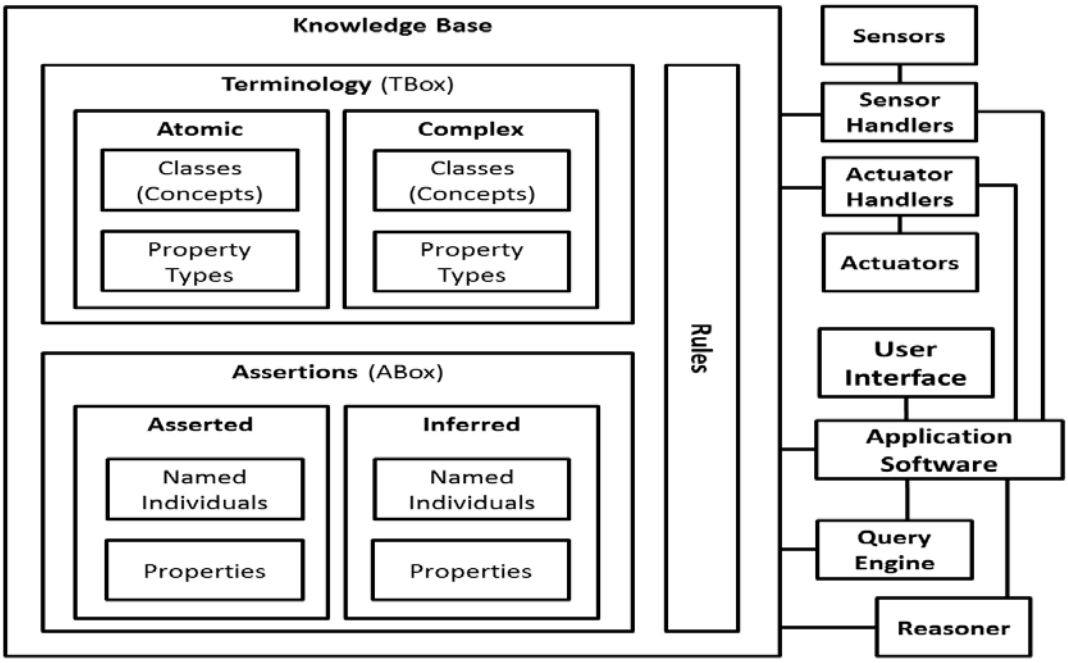

Figure 1 - Envisioned knowledge-based decision support system

the soundness of the classification hierarchies, e.g., decided if it is impossible for a concept to have individuals.

The knowledge base, often called ontology, can be described in many ontology languages. We have chosen a Semantic Web [5] approach and selected The Web Ontology Language (OWL) [6]. One argument for this is the already existing medical ontologies in this language, e.g., a version of Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) [7] and International Classification for Nursing Practice (ICNP) [8]. The most used ontology editor and knowledge acquisition system [9], Protégé [10], also supports OWL. There is also a Java Semantic Web framework, Jena [11,12,13, 14], which supports an API for OWL (and other Semantic Web technologies).

Complex concepts (see Figure 1) can be seen as rules, but there may be additional rules specified in some other language. SPARQL Protocol and RDF Query Language (SPARQL) [15] is a part of the Semantic Web technology. The use of SPARQL has, out of interest, been a requirement for the student projects. SPARQL can be used to query and gather new information given some OWL ontology (see Rules and Query Engine in Figure 1). There are examples of OWL/SPARQL being used in medical informatics [16,17].

Protégé and Jena both support several reasoners like Pellet [18] and FaCT [19].

Sensors for measuring vital signs, etc., and actuators for medications, etc., could be applicable in a practical situation. Furthermore, application software tying the components together and a user interface is needed.

From before we had some experience with both Protégé and Jena; they could both be used separately or together as described later. Experience indicated that making a Java application utilizing Jena could be very time consuming if lacking previous experience; Protégé was consequently proposed as a tool for the projects. A shift of technology from Protégé to Jena would still allow the results to be utilized; the ontology language used in Protégé would be OWL which is also understood by Jena software; the same goes for SPARQL statements. However, it turned out that the Protégé environment had some unforeseen limitations, more on this later.

The Master Theses

To sum up the previous given information, both student projects were developed in Protégé, which supplied a user interface, application software, reasoners, OWL, SPARQL and visualization. Actuators were not involved. Sensors/sensor handlers were simulated by manually inserting sensor data in the form of OWL statements. Also input about users and their states were manually inserted with OWL statements in Protégé; a fully developed application would have a tailored user interface to handle the human machine communication instead of using the generic Protégé user interface. This gave a set up that was manageable in the limited time frame and it gave the possibility to focus on defining the OWL ontology and the SPARQL rules, i.e., the kernel of an ontology based decision system as we envisioned it.

An Ontology-Based Decision Support System for Interventions based on Monitoring Medical Conditions on Patients in Hospital Wards

This project is a first start making the TILT system as a software system. The main intention with TILT is to give an early identification of life-threatening conditions based on parameters like measurements of respiration frequency, pulse oximetry, blood pressure, etc.. The parameters are measured regularly, how regularly should be given by the decision support system. The different parameters are given score, e.g., respiration frequency between 9 and 14 (i.e., normal frequency) is given 0, below 9 give 2, between 15 and 20 gives a score of 1. In essence 0 indicates a normal state, while a higher score indicates something that may be abnormal or a risk. The scores at one point in time are inserted into a table where they are summed. A high sum indicates a high risk and consequently: The higher sum the more frequent a new table should be assembled based on new measurements.

Suggestions when it comes to symptomes were decided with help of additional rules, e.g., resparation frequency score 3 and pulse oximetry score less than 2 and blood pressure score less than 2 gives suggesition: Hyperventilation.

An ontology modelling the following was sucessfully made:

- Classes *Patient*, *VitalSign*, *TILTTable*, *Suggestion*, etc.
- Relation types *hasVitalSign*, *hasSuggestion*, etc.
- Test data in the form of individuals of type *Patient*, *VitalSign*, etc.
- Test data in the form of application of the relation types (i.e., properties) *hasVitalSign*, *hasSuggestion*, etc.

SPARQL was used to define rules that gave the sum of the scores of a filled TILT table. Also suggestions were produced by applying SPAQL queries.

Ontology-Based Personalized System to Support Patients at Home

The proposed ontology model concepts like *Patient*, *Disease*, *Examination*, *Medication*, *GlucoseLevel*; and, properties like *hasMedication*, *hasDiet*, *hasInsulinIntake*. Individuals of the concepts and application of the property types were used to model specific patients with special medical conditions. The ontology focused on diabetes patient data and the modelling of examination results concerning diabetes. Predefined recommendations were inserted into the ontology, e.g., recommendations concerning insulin intake were connected to specific glucose levels. In this way a lot of information was modelled; SPARQL queries were given to show how the stored information could be retrieved.

An existing ontology for modelling questionnaire was integrated and exploited (this ontology had been made by both the mentioned students in an earlier project).

Discussion and Some Conclusions

OWL is a powerful language and the ontologies were expressed by utilizing only a small subset of the language. Protégé is well suited for manually defining the terminology and the assertional part of the ontology.

SPARQL can be used as a rule language as demonstrated in [17] and discussed in [20]. Protégé seems to give full support for SPARQL 1.1 which includes the *construct query*. The construct query allows a graph to be constructed based on a given query; the graph is composed of statements that can easily be inserted back into the ontology. However, the insertion is not done automatically by the construct query, implying that an insertion has to be done manually in Protégé.

SPARQL 1.1 Update [21] provides update, create, and remove operations on ontologies. In SPARQL 1.1 Update insertions can be done automatically with help of its *insert operation*. SPARQL 1.1 Update is not supported by a Protégé plugin.

And also, the SPARQL plugin found in Protégé did not allow a sequence of SPARQL statements to be run automatically (which after all is not that meaningful when insertions are not done automatically).

Using SPARQL as a rule languages were demonstrated in the project concerning TILT tables. However, using Protégé together with the SPARQL plugin, to simulate a system having rules changing the ontology, seems extremely unpractical. Such a system may run in cycles, where several rules are executed in a sequence, allowing the changes to the ontology by one rule to be used by the next. The amounts of manual work (i.e., run a rule and copy the changes to the ontology, run next rule copy changes to ontology, etc.) would be somewhat overwhelming. In other words, Protégé does not seem like well suited for this kind of experiments.

One may consider several different types of TILT tables; the proposed solution defined only one and it did not define concepts for defining TILT tables in general. To achieve this a metamodel approach seems natural; this approach foresees 3 levels (these levels are conceptual and just meant to help understand this way of thinking); one level, defined in advance, defines concepts to be used when defining one type of TILT table; the second level "uses" the concepts of the first level defining a special type of TILT table; filled in tables reside on the last level. To implement something like this demands programming (or some tools specifically designed for doing metamodelling) and it seems impractical to use Protégé for something like this. This approach defines two types of user roles, one defining tables and one filling tables. The same type of thinking may apply to rules, since rules also could be supplied by some user, i.e., users that were not IT personnel knowing SPARQL. However, making such a system is not trivial.

The project concerning personalized system to support patients at home supplied a rather big ontology, where different types of tables were represented, e.g., information relating glucose levels with recommendation of insulin intake. No functions or rules were supplied; all information was coded into the ontology. This kind of approach works well for small and uncomplicated cases, but will break down when the complexity grows – it will not be practical to (manually) insert all possibilities into the ontology.

We consider both the projects helpful and successful taking the limited time into account.

Recommendations for Further Work

The high number of medical ontologies and the rather mature semantic web technologies fuel our interest in the *knowledge-based decision support system approach*. Central to this approach is an explicitly represented knowledge base, defined in some ontology language. The Ontology Users' Survey [22] indicates that OWL is by far the most used ontology language in 2013. Adding our good experience from other projects [23] makes OWL a natural choice.

One of the purposes of using ontologies is to establish common languages and facilitate unambiguous communication in single system and among different systems; at a practical level it allows interoperability between existing applications; the ontologies are believed to be long lasting valuable assets.

We already have some experience [23] with The International Classification for Nursing Practice (ICNP®) [24]. ICNP aims at improving communication among nurses, and between nurses and other health care professionals. ICNP already contains concepts like *Diabetes*, *Blood Glucose* and *Effective Blood Glucose Level*; reusing these concepts in the aforementioned ontology would be an improvement. Another often mentioned medical ontology is SNOMED CT.

Ontologies like SNOMED CT function as standards defining medical concepts; before defining your own "new" concepts the most common medical ontologies should be checked to see if they already contain a representation of the concept. There has are also been done some work on defining sensor ontologies [25].

The two projects reported in this paper did not utilize the possibility to define complex concepts. This option is very powerful; to give an example: Alarms are common in the surveillance of health conditions; the following setup may be implemented:

- The conditions identifiying the triggering of alarms may be defined in advance as a complex concept, i.e., the definition of different types of alarms.
- One execution activity may continuously input sensor data into the knowledge base (in accordance with some defined sensor ontology). At some point in time the sensor data may satisfy the conditions for triggering an alarm (this could be the combination of data from multiple sensors); however, the duty of this execution activity is only to insert the sensor data.
- There are several reasoners available. One reasoner may be set to run continuously to infer individuals of complex concepts. In our case such individuals would correspond to alarms.
- One execution activity may continuously query the knowledge base for individuals of alarms, and if found then give a corresponding alarm.

Analyzing the proposed handling of alarms makes it clear that handling of time is essential when dealing with the sensory data. In general, treating of patients typically involves patient history; consequently the handling of time is important when it comes to medical informatics, and high priority should be given to doing this in an intelligent way.

There are ways to adapt Protégé and make some suitable user interface and also make plugins; however, as already mentioned the SPARQL 1.1 Update is not supported.

We have good experience with the Jena framework which also supports SPARQL 1.1 Update, reasoners, etc.; it seems to cover what we need. Using Jena means to develop a Java application and utilize the rich set of semantic web support given by existing Java libraries.

OWL ontology can be seen as a graph; the graph is broken up into triples when stored; a special database technology called triple store [26] has been developed for efficient storage and retrieval of triples. Jena gives close to transparent support for this kind of storage. The knowledge base may be distributed through different SPARQL endpoints; an SPARQL endpoint is conformant to the SPARQL protocol service (i.e., supporting queries and updates); a SPARQL endpoint may be offered at a server deploying a triple store.

References

[1] Zhao T. An Ontology-Based Decision Support System for Interventions based on Monitoring Medical Conditions on Patients in Hospital Wards. Master Thesis in Information and Communication Technology, UIA, 2014.

[2] Mukasine A. Ontology-Based Personalized System to Support Patients at Home. Master Thesis in Information and Communication Technology, UIA, 2014.

[3] Mapp ID, Davis LL, Krowchuk H. Prevention of Unplanned Intensive Care Unit Admissions and Hospital Mortality by Early Warning Systems. Dimensions of critical care nursing : DCCN, 32. [Internet]. [cited 2014 May] Available from: http://www.biomedsearch.com/nih/Prevention-Unplanned-Intensive-Care-Unit/24100432.html

[4] Baader F, Calvanese D, McGuinness DL and van Harmelen F. The Description Logic Handbook: Theory, Implementation and Applications. New York, NY, USA: Cambridge University Press, 2003.

[5] W3C. Semantic Web. [Internet] Available: http://www.w3.org/standards/semanticweb/, Accessed May 2014.

[6] W3C. OWL 2 Web Ontology Language. December 2012. [Internet]. [cited 2014 May] Available from: http://www.w3.org/TR/2012/REC-owl2-primer-20121211/.

[7] International Health Terminology Standards Development Organisation (IHTSDO). SNOMED CT. [Internet]. [cited 2014 May]. Available from: http://www.ihtsdo.org/snomed-ct/.

[8] International Council of Nurses (ICN). International Classificatiom for Nursing Practice (ICNP®) Information Sheet. [Internet]. [cited 2014 May]. Available from: http://www.icn.ch/images/stories/documents/pillars/Practice/icnp/ICNP_FAQs.pdf.

[9] Gašević D, Djurić D, Devedžić V. Model Driven Engineering and Ontology Development (2nd ed.). Springer. p. 194. ISBN 978-3-642-00282-3; 2009.

[10] Stanford Center for Biomedical Informatics Research. Protégé. [Internet]. [cited 2014 May] Available from: http://protege.stanford.edu/.

[11] Apache Jena project. Apache Jena. [Internet]. [cited 2014 May]. Available from: http://jena.apache.org/.

[12] McBride B. An Introduction to RDF and the Jena RDF API [Internet]. 2007 [cited 2014 May]. Available from: http://jena.sourceforge.net/tutorial/RDF_API/index.html

[13] Grobe M. RDF, Jena, SparQL and the 'Semantic Web', Proceedings of the 37th annual ACM SIGUCCS fall conference, October 11-14, 2009, St. Louis, Missouri, USA

[14] McCarthy P. Introduction to Jena. [Internet]. 2004 [cited 2014 May]. Available from: http://www.ibm.com/developerworks/xml/library/j-jena/

[15] W3C. SPARQL 1.1 Query Language. W3C Recommendation, 21 March 2013 [Internet]. [cited 2014 May]. Available from: http://www.w3.org/TR/sparql11-query/.

[16] Lasierra N, Alesanco A, García J, O'Sullivan D. Data management in home scenarios using an autonomic ontology-based approach. PerCom Workshops 2012: 94-99.

[17] Beamonte NL. An ontology-driven architecture for data integration and management in home-based telemonitoring scenarios [PhD thesis]. University of Zaragoza; 2012. [Internet]. [cited 2014 May]. Available from: http://zaguan.unizar.es/record/10003.

[18] Sirin E, Parsia B, Grau BC, Kalyanpur A, Katz Y. Pellet: A practical OWL-DL reasoner. Web Sem. 5 (2) , 51-53; 2007.

[19]Tsarkov D, Horrocks I. FaCT ++ Description Logic Reasoner : System Description. pp. 292–297, 2006.

[20]Polleres, Axel. From SPARQL to rules (and back). Paper presented at the meeting of the WWW, 2007. [Internet]. [cited 2014 July]. Available from: http://www2007.org/papers/paper435.pdf

[21] W3C. SPARQL 1.1 Update. W3C Recommendation 21 March 2013. [Internet]. [cited 2014 May]. Available from: http://www.w3.org/TR/sparql11-update/.

[22] Warren P. Ontology Users' Survey – Summary of Results. Technical Report kmi-13-1, 2013 June. [Internet]. [cited 2014 May]. Available from: http://kmi.open.ac.uk/publications/pdf/kmi-13-01.pdf

[23] Nytun JP, Fossum M. Information Model for Learning Nursing Terminology. MIE2014. To be published in Proceedings of, The 25th International Congress of the European Federation for Medical Informatics; Istanbul; August 31 - September3rd; 2014

[24] International Council of Nurses (ICN). International Classificatiom for Nursing Practice (ICNP®) Information Sheet. [Internet]. [cited 2014 May]. Available from: http://www.icn.ch/images/stories/documents/pillars/Practice/icnp/ICNP_FAQs.pdf.

[25] Compton M, Barnaghi P, Bermudez L, Garcıa-Castro R, Corcho O, Cox S, Graybeal J, Hauswirth M, Henson C, Herzog A, Huang V, Janowicz K, Kelsey W, Phuoc D, Lefort L, Leggieri M, Neuhaus H, Nikolov A, Page K, Passant A, Sheth A, Taylor K. The SSN Ontology of the W3C Semantic Sensor Network Incubator Group. Web Semantics: Science, Services And Agents On The World Wide Web, 17, 2012. [Internet]. [cited 2014 May]. Available from: http://www.websemanticsjournal.org/index.php/ps/article/view/312/312

[26] Garshol LM. RDF triple stores — an overview. 2012 September. [Internet]. [cited 2014 May]. Available from: http://www.garshol.priv.no/blog/231.html

Address for correspondence

Jan Pettersen Nytun, Department of ICT, University of Agder, Norway, jan.p.nytun@uia.no, address: Postboks 509, 4898 Grimstad, Norway