

An HL7 Rim-Driven, Ontology-based Approach to Healthcare Knowledge Base Development

Y.F. Zhang, Y. Wang, P.F. Li, J.S. Li

EMR and Intelligent Expert System Engineering Research Centre, Ministry of Education, Innovation Joint Research Center for Cyber-Physical-Society System, College of Biomedical Engineering and Instrument Science
Zhejiang University
Hangzhou, China

Abstract—The development of the core component of clinical decision support systems (CDSS), the knowledge base, has become a significant bottleneck in its successful application due to lack of effective knowledge representation models. To solve this problem, this paper presents a novel approach by developing a healthcare knowledge base using semantic technologies, of which the semantic healthcare knowledge ontology (SHKO) was constructed, based on HL7 reference information model (RIM), and the semantic healthcare expression repository (SHER) was built in Jena rule format and simple protocol and resource description framework query language (SPARQL). The proposed solution is validated in a case study of the management of patients with type 2 diabetes mellitus to demonstrate its technical feasibility.

Keywords—HL7 RIM; ontology; knowledge base; clinical decision support

I. INTRODUCTION

A clinical decision support system (CDSS) is defined to be a computer system that can assist healthcare providers and administrators with such decision tasks as information retrieval, data analysis, diagnosis and test, procedure and case management recommendation[1]. The use of CDSS to facilitate the practice of evidence-based medicine promises to substantially improve healthcare quality[2]. However, despite its great potential and a history of successful trials to this day, clinical decision support systems have not found wide use outside of a handful of mostly academic medical centres[3 4]. One of the major impediments to its widespread adoption is the lack of standard knowledge representation models to guide the development of the knowledge base, which is the core component of a CDSS.

The ontology, originally defined as “a formal, explicit specification of a shared conceptualization”[5], has been considered as a key technology in knowledge engineering. A variety of ontologies have been constructed in the biomedical domain, while the distributed nature of ontology development has led to a proliferation of medical ontologies, often with significant heterogeneity and overlapping domains, which itself creates obstacles to knowledge integration.

Health Level Seven International (HL7) is an ANSI-accredited standards developing organization dedicated to providing a comprehensive framework and related standards for the exchange, integration, sharing, and retrieval of electronic health information. In version 3 of HL7 standards,

all message specifications are based on the reference information model (RIM), a shared information model for integrating patient data and medical domain knowledge[6], which can serve as the basis of unified domain ontology development.

In this paper, we present a novel approach to solve the above mentioned problem by developing a semantic healthcare knowledge base (SHKB), of which the core component is a semantic healthcare knowledge ontology (SHKO) constructed based on HL7 RIM to represent healthcare domain knowledge; a semantic healthcare expression repository (SHER) of rules encoded in Jena rule format and queries encoded in simple protocol and resource description framework query (SPARQL) format is built to extend the SHKO with dynamic decision logic. The proposed method is validated in a case study of the management of patients with type 2 diabetes mellitus to demonstrate its technical feasibility.

II. MATERIALS AND METHODS

A. Knowledge Acquisition

The knowledge acquisition phase includes three processes: knowledge identification, knowledge extraction, and knowledge authentication. Three types of knowledge sources are identified in the knowledge identification process:

1. Published medical literature on care plans (CP) or clinical practice guidelines, standard nomenclatures and clinical information models;
2. Patient data from the electronic medical record (EMR);
3. Experiences of healthcare practitioners and domain experts.

In this process, we gained a general view of the basic healthcare workflow. We used the UML activity diagram to describe the healthcare workflow, as shown in Figure 1.

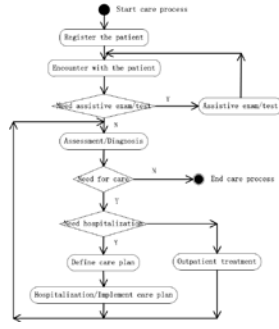


FIGURE I. MODEL HEALTHCARE WORKFLOW AS UML ACTIVITY DIAGRAM.

We then developed a process-oriented description of the recognized healthcare workflow with the HL7 RIM class diagram, containing all the concepts, attributes and relationships that constitute the information model of each stage in the healthcare workflow, as shown in Figure II;

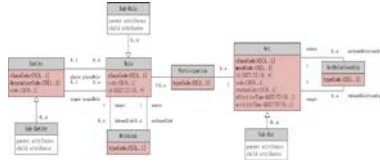


FIGURE II. HL7 RIM CLASS DIAGRAM (ADAPTED FOR SIMPLIFICATION).

Finally, we drew information semantics from these class diagrams using the meta-ontology approach[7]: concepts are expressed as classes; attributes and relationships are expressed as properties; decision algorithms are expressed as rules, as shown in Figure III.

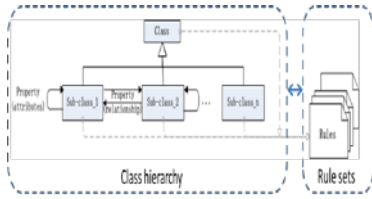


FIGURE III. SEMANTIC EXTRACTION USING THE META-ONTOLOGY APPROACH.

Close consultation with domain experts are involved in the concurrent knowledge authentication process to examine the necessity of extraction and the correctness of extracted information semantics.

B. Knowledge Representation

Knowledge representation focused on the construction of the semantic healthcare knowledge base (SHKB), including the ontology of healthcare domain knowledge and an expression repository of rules and queries. We developed SHKB on Protégé, an open-source ontology editor and framework from Stanford University[8].

1) *Semantic healthcare knowledge ontology(SHKO)*: SHKO is the backbone ontology of SHKB to represent healthcare domain knowledge. We developed SHKO following the ontology developing guide on[9], main steps include: defining classes, defining the properties of classes and

creating instances. Following a top-down design pattern, we defined three top-level classes are defined in SHKO, i.e., Data, Expressions and Model. Below is a brief description of each class:

1. **Data**. **Data** models the standard data types defined in HL7 Version 3 Data Types Abstract Specification[10].
2. **Expression**. **Expression** is the linkage class between SHKO and SHER. We defined the following properties to represent linkage semantics: **hasExprId** is the expression identifier in SHKO, **hasExprFile** is the pointer to the external expression file, **hasTriggerSource** marks the trigger source of the expression, either time-based, event-based or user defined, **hasExpression** is a textual description of the expression body.
3. **Model**. **Model** is compliant to the model definition of the HL7 RIM Version 2.41 [11], involving three main classes, **Act**, **Role**, and **Entity**, linked together with three association classes: **ActRelationship**, **Participation** and **RoleLink**. **Act** represents all the acts, performed and recorded by responsible actors in a healthcare workflow; **ActRelationship** is a directed association between a source act and a target act; **Entity** represents physical thing, group of physical things or an organization capable of participating in an act while in a role; **Role** is a competency of an that plays the role as identified, defined, guaranteed, or acknowledged by the entity that scopes the role; **Participation** is an association between an act and a role; **RoleLink** is a connection between two roles expressing a dependency between those roles.

In addition, we imported the SWRL temporal ontology [12]to enhance the temporal expressivity of SHKO with temporal models and built-ins.

2) *Semantic healthcare expression repository (SHER)*: SHER is designed to extend SHKO expressivity with expressions so that it can be queried and reasoned over for knowledge-based decision making. We considered two kinds of expressions: queries and rules. Queries are written in simple protocol and RDF query language (SPARQL) [13]. Rules are encoded in Jena rule format[14].Query-1 is an example query to select all the patients with their participating acts.

Query-1

```
SELECT ?p ?act
WHERE{
    ?p shko:hasRolePtn ?ptn.
    ?ptn shko:typeCode hl7:SBJ.
    ?ptn shko:hasPtnAct ?act.
}
```

Rule-1 is an example rule to specify the eligibility criteria for a care plan, which can be interpreted as: if the patient has a diagnosis that matches the precondition of a care plan, the patient can be set as the target of the care plan.

Rule1

```
(?p shko:hasRolePtn ?ptn)(?ptn shko:typeCodehl7:SBJ)
(?ptn shko:hasPtnAct ?diag)(?diag shko:code ?diag_cd)
(?diag shko:moodCode hl7:EVN)(?cp shko:moodCode hl7:DEF)
(?cp shko:hasOutRel ?rel)(?rel shko:typeCode hl7:PRCN)
(?rel shko:hasActRelTgt ?tgt)(?tgt shko:code ?tgt_cd)equal(?diag_cd,?tgt_cd)
->(cp shko:hasActPtn ?cp_ptn)(?cp_ptn shko:hasPtnRole ?p)
(?cp_ptn shko: typeCode shko:SBJ)
```

C. Knowledge Validation

We developed a prototype semantic-based clinical decision support system (S-CDSS) in Eclipse IDE[15] to enable SHKB-based clinical decision making. The proposed method is validated in a case study of the management of patients with type 2 diabetes mellitus. We instantiated the SHKB with the domain knowledge from the Chinese medical guideline to prevention and treatment of type 2 diabetes mellitus[16] and the standards of medical care in diabetes from American Diabetes Association (ADA)[17]; the CDA admission summaries containing patient data are then imported into SHKO as input fact data; given patient data and rules from SHER, we can use the Jena inference engine to derive new facts about the patient (diagnosis or care plan recommendations).

III. RESULTS

A total of 130 classes (6 top-level classes and 124 subclasses), 196 properties (158 object properties and 38 data type properties) and 126 rules are created in SHKB. Figure IV illustrates the ontology model of the type 2 diabetes mellitus care plan in definition mode. The precondition, outcome, responsible author, target patient and constituent care plan activities are expressed as instances of SHKO classes and properties. When implemented, this model is replicated with different mood and status designations to represent different application scenarios. Notice that we use namespace prefixes to differentiate ontologies from different sources: shko for the local SHKO, and snomedct, loinc, hl7 for the imported SNOMED CT, LOINC, HL7 vocabulary ontologies, respectively.



FIGURE IV. TYPE 2 DIABETES MELLITUS CARE PLAN ONTOLOGY IN DEFINITION MODE.

Figure V shows the actual clinical actions the patient (ID: PAT101) has taken and the recommended care plan after executing relevant rules and queries.

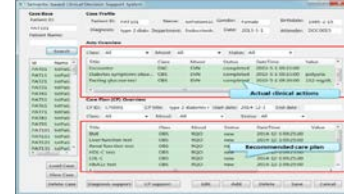


FIGURE V. S-CDSS USER INTERFACE, ACTUAL CLINICAL ACTIONS AND RECOMMENDED CARE PLAN FOR PAT101.

IV. DISCUSSION AND CONCLUSION

In this paper, we present a novel approach to enhance effective use of CDSS by developing SHKB, a healthcare knowledge base based on HL7 RIM and ontology. HL7 RIM offers a static model of health care information, which includes classes and state-machine diagrams and is accompanied by use case models, interaction models, data type models, terminology models, and other types of models based on which domain-specific information models are then derived through a series of constraining or refinement processes. Ontology provides a standard model for data and knowledge representation.

The validation results from the case study of the management of patients with type 2 diabetes mellitus demonstrated the benefits of the combination use of the HL7 RIM and ontology, which include: (1) the unified representation formalism for both patient data and domain knowledge to facilitate the semantic interoperability between CDSS and healthcare information systems; (2) semantic-based integration and reuse of computer technologies and domain knowledge for convenient knowledge base maintenance; (3) knowledge-based inference for dynamic and personalized healthcare.

Future work will focus on: (1) SHKB maintenance and extensions based on other case studies; (2) investigation of knowledge base evaluation methods; and (3) integration with machine learning techniques for automatic knowledge acquisition and close-loop knowledge management.

ACKNOWLEDGMENTS

This work was supported by the China National Natural Science Foundation (Grant No. 61173127), National High-Tech R&D Program (No. 2013AA041201), and the Fundamental Research Funds for the Central Universities.

REFERENCES

- [1] Ranjit Bose, Knowledge management-enabled health care management systems: capabilities, infrastructure, and decision-support. *Expert Systems with Applications*, 24(1), pp. 59-71, 2003.
- [2] Alejandro R. Jadad, R. Brian Haynes, Dereck Hunt, George P. Browman, The Internet and evidence-based decision-making: a needed synergy for efficient knowledge management in health care. *Canadian Medical Association Journal*, 162(3), pp. 362-365, 2000.
- [3] Jaspers M.W., Smeulers M, Vermeulen H, Peute L.W., Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. *Journal of the American Medical Informatics Association*, 18(3), pp. 327-334, 2011.
- [4] Wright A, Sittig D.F., A four-phase model of the evolution of clinical decision support architectures. *International journal of medical informatics*, 77(10), pp. 641-649, 2008.
- [5] Studer R, Benjamins V.R., Fensel D, Knowledge engineering: principles and methods. *Data & Knowledge Engineering*, 25(1), pp. 161-197, 1998.
- [6] Schadow G, Russler D.C., Mead C.N., et al, Integrating medical information and knowledge in the HL7 RIM. *Proc. of the AMIA Symposium*, American Medical Informatics Association, pp. 764, 2000.
- [7] Z Hu, J.S. Li, T.S. Zhou, H.Y. Yu, Suzuki M, Araki K, Ontology-based clinical pathways with semantic rules. *Journal of medical systems*, 36(4), pp. 2203-2212, 2012.
- [8] Protégé. <http://protege.stanford.edu>.
- [9] Noy N.F., McGuinness D.L., *Ontology development 101: A guide to creating your first ontology*: Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880, 2001.
- [10] HL7 v3.0 Data Types Specification. <http://amisha.pragmaticdata.com/v3dt/report.html>.
- [11] HL7 Reference Information Model. <http://www.hl7.org/implement/standards/rim.cfm>.
- [12] SWRLTemporalOntology. <http://protege.cim3.net/cgi-bin/wiki.pl?SWRLTemporalOntology>.
- [13] SPARQL 1.1 Query Language. <http://www.w3.org/TR/2013/REC-sparql11-query-20130321/>.
- [14] Apache Jena. <http://jena.apache.org/>.
- [15] Eclipse. <http://www.eclipse.org/>.
- [16] Chinese medical guideline to prevention and treatment of type 2 diabetes mellitus. <http://www.diab.net.cn/guideline-1.jsp>.
- [17] American Diabetes Association, Standards of medical care in diabetes—2013. *Diabetes care*, 36(Suppl 1), S11, 2013.