User Centred Networked Health Care
A. Moen et al. (Eds.)
IOS Press, 2011

doi:10.3233/978-1-60750-806-9-699

Ontology-based Knowledge Management for Personalized Adverse Drug Events Detection

Feng CAO [a, 1], Xingzhi SUN [a], Xiaoyuan WANG [a], Bo LI [a], Jing LI [a], and Yue PAN [a]
[a] *IBM Research – China*

Abstract. Since Adverse Drug Event (ADE) has become a leading cause of death around the world, there arises high demand for helping clinicians or patients to identify possible hazards from drug effects. Motivated by this, we present a personalized ADE detection system, with the focus on applying ontology-based knowledge management techniques to enhance ADE detection services. The development of electronic health records makes it possible to automate the personalized ADE detection, i.e., to take patient clinical conditions into account during ADE detection. Specifically, we define the ADE ontology to uniformly manage the ADE knowledge from multiple sources. We take advantage of the rich semantics from the terminology SNOMED-CT and apply it to ADE detection via the semantic query and reasoning.

Keywords. Adverse Drug Events (ADE) detection, knowledge management, ontology, semantic reasoning

1. Introduction

Adverse Drug Event (ADE), including adverse reaction from a single medication, and drug interaction from two or more medications, is increasingly becoming a serious problem for public health. According to the Journal of American Medical Association [1], ADE has been the 4th leading cause of death. The yearly cost of ADE is $136 Billion in US 2000, greater than total cost of cardiovascular or diabetic care. In this way, there arises high demand for advanced systems to help doctors and patients to detect the potential ADE in clinical practice.

There exist the following challenges for ADE detection: First, knowledge sources for ADE detection are heterogeneous, including relational data, semi-structured data, and RDF data. How to build a general system to uniformly manage ADE knowledge is a non-trivial task. Second, personalized ADE detection requires the integration of patients' record system with ADE knowledge base. There lacks a general framework for the interoperation between these two systems. Third, clinical terms are used in both patients' records and ADE knowledge. During ADE detection, it is not straightforward to apply rich relationships among clinical terms so that ADE knowledge and patients' records can be matched based on their semantics instead of literal characters.

[1] Corresponding Author: E-mail: {caofeng, sunxingz, wangxyxy, libocrl, jingli, panyue}@cn.ibm.com; IBM Research – 6F, Building 10, 399 Keyuan Road, Zhangjiang Innovation Park, Pudong New District, Shanghai, China.

Semantic Web technology can offer great help to address the above mentioned challenges. On the one hand, RDF and ontology have the intrinsic advantage for data integration and interoperation, which can help to resolve the challenges on heterogeneous data sources and the interoperation between patient record system and ADE knowledge base. On the other hand, ontology reasoning can derive the implicit information, which can bridge the gap between patient clinical data and ADE knowledge and enable the semantic matching. This paper presents a personalized ADE detection system, with the focus on applying ontology-based knowledge management techniques to enhance ADE detection services.

2. System Architecture

The system takes ADE knowledge and patients' data as input, uses the standard vocabulary (such as SNOMED-CT) to manipulate these data, and provides the service of personalized ADE detection by semantic query and reasoning. Figure 1 shows the system architecture, in which the basic components are organized in three layers.

Storage layer has two repositories, ***Ontology Repository*** and ***ADE Repository***, both in the relational model. The ontology repository is used for storing the medical terminology (i.e., SNOMED) and the pre-defined ADE model as ontology. The ADE repository is designed for storing ADE knowledge.

Knowledge layer is responsible for pre-processing the ontologies and ADE knowledge, and loading them into the corresponding repositories. Before loading the SNOMED, ***T-Box Reasoner*** preprocesses the SNOMED terminology by reasoning out the implicit relationships among the clinical terms. ***ADE Knowledge Loader and Converter*** is used to load heterogeneous external knowledge sources to the standardized ADE repository.

Access layer enables the personalized ADE detection by semantic query and reasoning. Our ***Semantic Data Access (SeDA) engine*** [4, 8] is the key component to publish ADE repository as virtual RDF store, and to enable the semantic query over the relational database. Given a set of drugs, ***ADE Query Coordinator*** aims to generate SPARQL [6] queries for retrieving all the related ADEs (together with their conditions and consequences). Further, ***Clinical Condition Matcher*** checks the patient's clinical records against the identified ADE conditions so that personalized ADE can be selected.

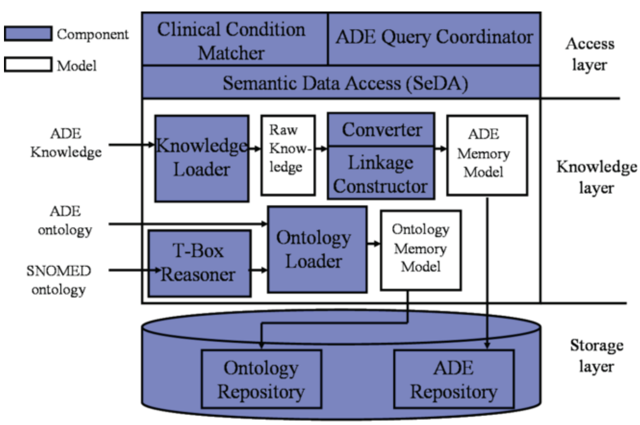

Figure 1. Architecture of Knowledge Model Management

3. Method

3.1. Semantic Query for ADE Detection

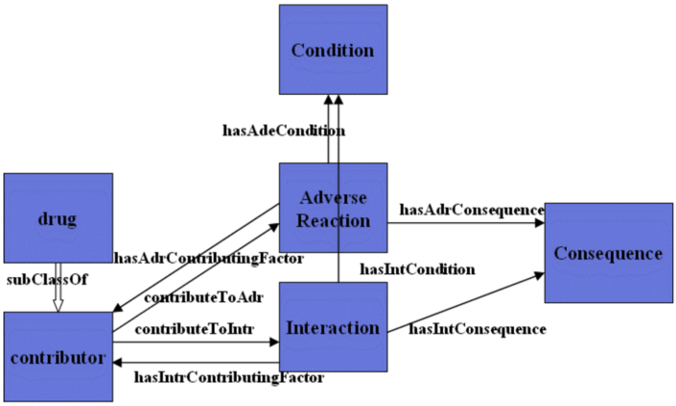

Figure 2. ADE ontology (only shows class and relationship property)

The knowledge of ADE is stored in the ADE database. The ADE ontology, shown in Figure 2, provides the logical view to the external. In addition to drug information, ADE knowledge can be classified as adverse reaction (of a single drug) and drug interaction (of multiple drugs). We use the SPARQL language to express the semantic query over ADE ontology data. Given a patient's prescription (i.e., a list of drugs), we issue semantic queries to find the adverse reactions caused by each drug d_i, and the drug interactions for each pair of drugs (d_i, d_j), together with their consequences and conditions.

Technically, we build a mapping between ADE ontology and the data stored in the ADE database, according to which, the semantic query in SPARQL retrieves the corresponding ADE information from the database. The mapping follows a broadly-used mapping language, called D2R (Database to RDF) [2]. Based on the mapping, the SeDA engine translates the SPARQL query into the SQL statement for ADE detection.

The advantages of this mapping-based approach are twofold. First, converting relational data into virtual RDF data makes it possible to integrate ontology reasoning with the semantic query. Second, from the system design point of view, the mapping based approach adds the independency between the ADE query logic and the physical storage of ADE knowledge.

3.2. Semantic Reasoning over SNOMED

The goal of semantic reasoning is to guarantee the completeness of ADE detection by applying the semantics captured in the terminology (in our system, SNOMED). The SNOMED reasoning can be classified into two types in terms of functionality.

Enrich ADE knowledge: SNOMED terms are used in the ADE knowledge. By referencing the relationships defined in SNOMED ontology, the ADE knowledge can be enriched. Table 1 shows the types of ADE knowledge enrichment we applied. Let us take type II as an example. Given the ADE knowledge "ingredient_I leads to disorder_D", if in SNOMED, we find "drug_A has active ingredient J, and J is a subclass of ingredient_I", we can infer "drug_A leads to disorder_D".

Table 1. Types for ADE knowledge enrichment

Type	ADE Knowledge	SNOMED Relationship(S)	Inferred Knowledge
I	drug_B -> disorder_D	drug_A sct:subClassOf drug_B	drug_A-> disorder_D
II	ingredient_I -> disorder_D	drug_A sct:hasActiveIngredient ingredient_J AND ingredient_J sct:subClassOf ingredient_I	drug_A-> disorder_D
III	Nil (no ADE information for a given drug A)	drug_A sct:hasActiveIngredient ingredient_J AND ingredient_J sct:subClassOf ingredient_I AND disorder_D sct:hasCausitiveAgent ingredient_I	drug_A-> disorder_D

TBox reasoning is applied for the ADE knowledge extension. Thus, in the ADE detection, we check ADEs based on both explicit knowledge and implicit knowledge.

Enable the semantic matching between ADE conditions and patient's records: After the semantic query finds the potential ADEs for a list of drugs, to enable the personalized ADE detention, we need to match the identified ADE conditions against the patient's EMR, which is given as a CDA document. Applying SNOMED reasoning can make this matching done semantically rather than literally.

For example, a fragment of CDA document (EMR) states that there is an observation of "Neoplasm on hilus of lung" for a patient:

```
<Observation>
    <code code="ASSERTION" codeSystem="2.16.840.1.113883.5.4"/>
    <value xsi:type="CD" code="126707007" codeSystem="2.16.840.1.113883.6.96"
    codeSystemName="SNOMED CT" displayName=" Neoplasm of hilus of lung "/>
</Observation>
```

Suppose the relevant ADE knowledge is "for patient who has disorder on lung, Drug A leads to adverse reaction R". The query for clinical condition matching is to check if the patient has the observation of disorder with finding site at lung.

Q(x) :- emr :patient(x), emr: hasObservation(x, y), sct :Disorder(y), sct :findingSite(y, z), sct :Lung(z) .

From SNOMED, we have:

1) sct: Neoplasm of hilus of lung $\subseteq$ sct:Neoplasm of lung $\cap$ sct:findingSite. sct:hilus of lung
2) sct:Neoplasm of lung $\subseteq$ sct:Disorder
3) sct:hilus of lung $\subseteq$ sct:Lung

By referencing SNOMED, we know the above patient meets the ADE condition specified by the query. The semantic matching is realized by ABox reasoning over SNOMED and patient's EMR. More technical details can be found in [5].

4. Conclusion and Discussion

We have developed the system for personalized ADE detection, with the following features highlighted:

1. Two kinds of ontologies are introduced into the system, the external standardized terminology system and the specially designed ADE domain ontology. With our Semantic Data Access engine, the system exposes the underlying relational data into virtual RDF view.

2. The system supports semantic query on RDF view and semantic reasoning over clinical ontology. With embedded semantic reasoning capability, the system can infer more explicit relationship from both ADE knowledge data and patient's clinical records, which realizes the personalized ADE detection semantically.
3. We set up a general ADE knowledge management system that can load, store, and query ADE data from heterogeneous sources with different formats. Based on the ADE model, the system supports multiple knowledge sources incorporation.

In practice, we have built a comprehensive system that provides four kinds of services to end users, that is, Terminology Service, Drug Service, ADE Loading Service, and ADE Detection Service. We implemented all of these services in J2EE environments with DB2 V9 (as ontology repository and ADE repository), and published them by WebSphere V7 as Web-based applications. The system is deployed in Korean Gil Hospital and allows clinicians to perform ADE search and report together with patients' health records.

Since this kind of intelligent clinical system tends to be knowledge-intensive in commercial use, ADE knowledge enrichment and enhancement is essential for the system. We divide it into four phases: seek for more ADE-relevant information from various data sources, extract ADE knowledge from multiple sources, encode ADE knowledge using standardized codes and terms, and incorporate different knowledge and keep semantic consistency. Our system accommodates three kinds of data sources currently: customer data from Korean Gil Hospital, Structured Product Labeling (SPL) [7] documents from FDA, and Linked Open Drug Data (LODD) [3]. The ADE knowledge extracted from different sources needs to be normalized with the same coding system so that the semantic reasoning methodology can be applied to enrich knowledge and complete semantics. We are developing the SNOMED encoding toolkit to support standard terminology encoding on different sources. In addition, we consider exploring more open data and verifying its effectiveness from the ADE perspective.

Acknowledgements. We thank to IBM Korean Ubiquitous Computing Lab, Haifa Research Lab for harmony collaboration in the BlueMedics project, and Gachon University Gil Hospital (Incheon, Korea) for their strong support.

References

[1] Lazarou J, BH. Pomeranz, PN. Corey. Incidence of Adverse Drug Reactions in Hospitalized Patients. In *Journal of the American Medical Association*. Vol. 279 No. 15, 1998.
[2] Chris Bizer JG, Cyganiak R, Maresch O. The d2rq platform v0.5.1 - treating non-rdf relational databases as virtual rdf graphs, in *Proceedings of ISWC*, 2004.
[3] Linking Open Drug Data (LODD). Semantic Web for Health Care and Life Sciences Interest Group. http://esw.w3.org/HCL-SIG/LODD.
[4] Ma L, Sun X, Cao F, et al. Semantic Enhancement for Enterprise Data Management. In *Proceedings of International Semantic Web Conference* 2009.
[5] Liu S, Ni Y, Mei J, et al. iSMART:Ontology-based Semantic Query of CDA Documents, in *Proceedings of AMIA*, 2009.
[6] SPARQL Query Lanaguge for RDF. http://www.w3.org/TR/rdf-sparql-query/
[7] Structured Product Labeling Resources. http://www.fda.gov/ForIndustry/DataStandards/StructuredProductLabeling/default.htm
[8] Wang X, Sun X, Cao F, et al. SMDM: Enhancing Enterprise-Wide Master Data Management Using Semantic Web Technologies. In *Proceedings of VLDB*, 2009.