Development of an EHR System for Sharing – A Semantic Perspective

Haifeng LIU 1, Xue Qiao HOU, Gang HU, Jing LI, Yu Qi DING
Information and Knowledge Department, IBM China Research Lab, Beijing, China

Abstract. Building up Electronic Health Record (EHR) systems is considered a promising approach to improve health care quality in China. In order to meet the challenges of EHR semantic interoperability, we present a HL7 Clinical Document Architecture (CDA) R2 based EHR system prototype developed for Guangdong Hospital of Traditional Chinese Medicine (GDHTCM). In particular, we demonstrate how semantic technologies are leveraged to benefit EHR users from enabling semantic search against an EHR repository. Pilot users of the system deemed it clinically useful, which encourages us to develop more capabilities based on EHR in future, like cross-institution sharing and integration with a clinical decision support system.

Keywords. electronic health record (EHR), clinical document architecture (CDA), semantic interoperability, semantic search, HL7 standards

1. Introduction

China is taking great effort to promote the reform and development of health care systems. Similar to the e-Health initiative proposed and being developed in European countries, e.g., Electronic Health Records (EHR), telemedicine etc, an urgent need has arisen for implementing HL7 Clinical Document Architecture (CDA) R2 based EHR systems [1], which has been considered an effective approach to realize health information exchange and related services [2] in China.

Motivated by meeting the challenges for implementing semantic interoperability of EHR systems [3], in 2007, we started a joint-research initiative on EHR called mCHAS with the Guangdong Hospital of Traditional Chinese Medicine (GDHTCM), one of the largest hospital enterprises in China. The major goals of mCHAS include:

- Implement a patient-centric data model for EHR based on HL7 CDA standard which provides a flexible and extensible document architecture that supports both syntactical and semantic interoperability;
- Provide profession-friendly semantic search against patient data for clinical workers and researchers;
- Enable cross-institution health care information sharing through EHR system.

Among them, being able to perform semantic search is particularly crucial for clinical workers and researchers, as they often need to search for the descriptions of a specific clinical finding from a large volume of clinical documents. An example is that a clinician needs to find out all patients who have edema in left hand after a reduction...
of fracture. While traditional keywords-based search often returns incomplete or unrelated search results, and depresses the users, a semantic search engine could be employed to complete the task by relying on the semantics encapsulated within CDA level 3 documents.

We have developed an EHR system prototype for mCHAS to fulfill the basic requirements mentioned above. The goal of this paper is to describe the mCHAS system prototype and demonstrate the power of leveraging semantic technology to enable semantic search with an EHR system. To the best of our knowledge, mCHAS is the first attempt targeting on delivering an online production CDA-based EHR system in China. We believe that our approaches lay a foundation which could also be shared by other efforts to address EHR semantic interoperability.

2. System Architecture

![Figure 1. System architecture for mCHAS EHR prototype](image)

Figure 1 depicts the overall architecture of mCHAS EHR prototype system. The system conforms to HL7 CDA Release 2 standard and IHE XDS profile [4], and is comprised of the following main components. (Due to the space limitation, we will not introduce the methods of generating CDA level 3 documents).

- **EHR repository.** It stores all CDA level 3 documents which encapsulate semantic information as coded entries occurred in their section and content parts by inserting codes from HL7 recognized coding systems, or locally defined codes.
- **Clinical terminology repository.** It formulates and stores clinical terminologies referenced by CDA documents including SNOMED CT, LOINC, ICD-10, and a terminology for Traditional Chinese Medicine as Resource Description Framework (RDF) triples [5]. RDF is a language for processing metadata, and RDF triples are popularly used to express the formal semantics defined by terminologies. An RDF triple is conventionally written in the form of subject-predicate-object, like “Cancer is A Disease” or “Pain has Finding Site head”.
- **EHR registry.** In compliance with IHE XDS profile, this component is responsible for registering CDA documents in EHR repositories with a purpose of sharing and exchanging documents. It maintains metadata about
each registered document, including a link to the EHR repository from which it can be retrieved.

- Publisher. In the need of semantic search which is normally performed against RDF triples, this component extracts, parses, and publishes CDA coded entries, default context values associated with particular entries (such as Observation and Procedure), and additional document-centric facts into RDF triples (please refer to the next section for the details).

- Reasoner. The component aims to provide inference capability to elucidate implicit semantic information which is crucial for semantic search. It generates expanded RDF triples by reasoning about the published RDF triples from CDA documents and the referenced coding terminologies.

- RDF indexer. The component aims to index a repository of RDF triples, and assists a semantic search engine in performing efficient search.

- Semantic search engine. This engine is responsible for taking user input, performing efficient and scalable semantic search against the RDF repository (holding by the indexer), and retrieving the resulted documents.

- Web user interface. We provide a user friendly web interface for users to input their semantic search requests, initiate search, and browse returned results (list of information about clinical documents).

3. Methods

We adopted state-of-the-art semantic web technologies to develop the semantic search capability for the system described above. It consists of the following steps:

1. We first employ the publisher to internally publish the coded CDA entries into RDF triples. Figure 2 shows an example of publishing RDF triples from a CDA 3 fragment of document “D00028” which represents an observation of “weakness of limb was found in entire limb”.

   ```xml
   <observation classCode="OBS" moodCode="EVN">
     <code code="309774006" display="Weakness of limb" codeSystem="2.16.840.1.113883.6.96" codeSystemName="SNOMED CT"/>
     <qualifier>
       <name code="363698007" display="Finding site (attribute)"/>
       <value> <code code="243996003" display="Entire limb" codeSystem="2.16.840.1.113883.6.96" codeSystemName="SNOMED CT" /> </value>
     </qualifier>
   </observation>
   ```

   CDA coded entry triples:

   ```xml
   <Weakness_of_limb/309774006> <findingSite> <Entire_limb/243996003>
   ```

   Default context-value triples for SNOMED CT concept “weakness of limb”:

   ```xml
   <Weakness_of_limb/309774006><findingContext>Known_present/410515003>
   <Weakness_of_limb/309774006><temporalContext><Current/410512000>
   <Weakness_of_limb/309774006><subjectRelationshipContext><Subject_of_record/410604004>
   ```

   Additional triples to associate the observation with a document:

   ```xml
   <D00028> <hasObservation> <Weakness_of_limb/309774006>
   ```

**Figure 2.** Example of publishing coded entries of CDA level 3 document
2. The reasoner with EL++ reasoning capability [6] performs reasoning over the existing RDF triples fed by the publisher together with the triples from the referenced terminology repository, thus expands the RDF triples that could be harnessed by the RDF indexer. For example, as illustrated in Figure 3, relying on the content related to “Weakness of limb” of the referenced terminology SNOMED CT, the reasoner is able to generate four more triples associated with document “D00028” from the published triple “Weakness_of_limb/309774006> <findingSite> <Entire_limb/243996003>” in the previous step.

3. The RDF indexer adopts existing information retrieval (IR) structures and functions to index the RDF triples acquired in step 2), and supports efficient and scalable search capability [7]. Basically, the indexer translates RDF data into concepts of IR indexing (documents, field, and terms). Based on a classical IR inverted index structure, the semantic search engine can answer queries about instances of a semantic concept (e.g., weakness of limb) and a semantic relation (e.g., findingSite), and relates the results with the specific documents using document-centric triples published in step 1).

4. Users submit their search requests through the web interface where users start from typing in a keyword related to a clinical concept, and repeatedly add in multiple relation and concept constraints to form a semantic pattern which explicitly describes the question they want to ask. The pattern can later help the search engine return complete and meaningful results by automatically expanding the search with related clinical concepts that can be inferred from the question.

5. The semantic search engine performs the search against the index files according to the user input, and feeds the resulted document IDs satisfying the search criteria into the EHR registry. The registry then returns a list of corresponding documents information to the search engine which in turn forwards it to the user interface. Users can further retrieve the specific documents of interest from EHR repository through the interface.

Figure 3. Example of reasoning over terminology

```xml
<Weakness_of_limb/309774006> <isA> <Muscle_weakness/26544005>
<Weakness_of_limb/309774006> <isA> <Finding_of_limb_structure/302293008>
<Weakness_of_limb/309774006> <findingSite> <Skeletal_muscle_structure/127954009>
<Weakness_of_limb/309774006> <findingSite> <Limb_structure/66019005>
```
4. Conclusion and Discussions

We have developed an EHR system prototype for the largest health service provider (GDHTCM) in South China in compliance with HL7 CDA level 3 standard in order to support semantic interoperability. The first promising benefit received by clinicians is the semantic search capability against a large repository of EHRs. The system prototype was implemented using J2SE 5.0 with DB2 V9 (as EHR repository and registry), and currently holds around ten thousand CDA documents. Pilot users from GDHTCM have had a fresh taste of using the prototype. They achieved great satisfaction with the overall system, and especially deemed that the semantic search (implemented using Semplore [7]) is clinically useful for them to locate specific patient cases, and expect to explore the full capabilities of a production system where more documents will be included to help them fulfill further analytics tasks.

On the other hand, the current registered EHR documents all come from one site, GDHTCM, and are only internally shared by different departments. By following the IHE XDS profile, we consider to develop regional EHR systems for the four subordinate hospitals of GDHTCM in next phase of mCHAS, and implement semantic interoperability among them. In addition, we consider integrating the developed EHR system with a knowledge-centric clinical decision support system where semantic technologies also play a key role for the success [8].

Acknowledgement. We thank to GuangDong Hospital of Traditional Chinese Medicine for their harmony collaboration and great support.

References