HL7 v3 Message Extraction using Semantic Web Techniques

Priya Jayaratna and Kamran Sartipi
Department of Computing and Software, McMaster University
Hamilton, ON, L8S 4K1, Canada
{jayarasp, sartipi}@mcmaster.ca

Abstract

Integration of legacy healthcare systems with new HL7 standard systems is an area of utmost importance in the overall eHealth strategy of the provincial and federal governments of Canada and several other countries. A large body of researchers from across Canada are engaged in delivering solutions such as shared Electronic Health Record, Ontario Laboratory Information System and Ontario Pharmacy System based on HL7 v3 standard. However, due to inherent complexity of HL7 v3 standard and the healthcare domain at large, successful completion of such projects is a difficult task. In this paper, we propose a well-defined framework for designing message workflow during HL7 v3 based system integration. The framework provides a behavior based classification technique for HL7 v3 Interactions that conform well with real-life healthcare transactions. The framework is assisted by a semantic web tool for extracting HL7 v3 messages from healthcare scenarios. Consequently, the proposed approach drastically reduces domain-dependant and expert-dependent complexities for software professionals who perform healthcare system integration using HL7 v3 standard.

1. Introduction

A key objective of Healthcare Informatics is to facilitate seamless integration amongst heterogeneous applications to provide a unified view of information to health practitioners and other stakeholders. Achieving such flawless integration requires interoperability amongst applications serving data sources. There are two distinct levels of interoperability: one is network-layer interoperability which refers to ability of systems built on heterogeneous platforms to communicate with each other. This is easily achieved through existing distributed computing technologies such as web services, DCOM and CORBA. The second level is semantic interoperability, which refers to the ability of systems to correctly interpret concepts and terms used by another system. This can only be achieved through standardization of information exchange and representation. Health Level 7 (HL7) [4] is an internationally accepted standard for healthcare information.

Over two decades since its inception, HL7 has undergone an evolutionary process starting from version 2.1 to its current version 3 (v3). HL7 version 2 (v2) was a text based messaging standard not driven by a consistent data model. Its scope was limited to a few healthcare domains and information exchanged was limited to basic fields. Due to its simplicity HL7 v2 could be adopted with minimum effort and required less intervention by healthcare domain experts.

HL7 v3 was a complete overhaul of its predecessor and was designed with consistency and comprehensive coverage in mind. While it has been hailed over v2 for being a ”true” standard offering precision and unambiguity, the worldwide healthcare community has so far been reluctant to adopt it due to its overwhelming complexity. HL7 v3 is sufficiently comprehensive to cover the breadth and depth of healthcare domain information. It supports a wide range of areas such as patient care, patient administration, laboratory, pharmacy, diagnostic imaging, surgical procedures, insurance, accounting and clinical decision support systems.

While all these topics are related, each of them has unique features and information requirements that need to be addressed by the standard. Furthermore, HL7 v3 uses several standard clinical terminology systems such as SNOMED and LOINC to represent information content.1 Thus HL7 v3 based integration of systems require a herculean effort on the part of IT professionals to gain sufficient knowledge of the standard itself.

This tedious process can be improved tremendously by developing guidelines, processes and tools to support legacy healthcare system integration and migration to new platforms. However, to the best of our knowledge, open-source tools supporting design and implementation of HL7 v3 based integration, are unavailable as of today. In this paper we propose a framework to support HL7 v3 message extraction for standard compliant integration projects. We present an open source tool to store, locate and explore HL7 v3 artifacts using leading edge SemanticWeb (SW) technologies [18].

The remaining sections of the paper are structured as follows. Section 2 describes related work in the area of healthcare integration. Section 3 briefly describes standards and technologies relevant to this paper. Section 4 is about the proposed message extraction process and our support tool. Section 5 concludes with suggestions for future research.

2. Related work

There are a number of commercial support tools available for HL7 version 2. 7Scan [3] is a specialized browser and editor that finds, displays, edits and transmits text-based HL7 version 2 messages with ease. 7Edit [2] is a productivity tool for browsing, editing, searching, validating HL7 messages and communicating with systems that support HL7 format. 7Edit supports HL7 versions 2.1 up to 2.6. NeoTool's NeoBrowse [22] offers a multi-view interface making HL7 2x messages easy to view and understand. Our research is geared towards developing an open-source tool that supports HL7 version 3 which is fundamentally different from version 2.

The HL7 v3 mapping process proposed in this paper is continuation of work carried out by Yarmand and Sartipi [24]. Their proposed model for message standardization is based on guidelines set forth by Canada Health Infoway [5]. Interaction selection and terminology mapping are offline operations unassisted by tools. In contrast, we propose a tool-assisted approach that is independent of Canadian national guidelines.

In other healthcare integration related research, Liu et.al. [20] discuss an HL7 v2 based integration project to establish interoperability between a hospital information system (HIS) and a Picture Achieving and Communication System (PACS) based on DICOM. They propose an information exchange gateway between DICOM and HL7 v2 based on a series of parsers, transaction processors and send/receive modules capable of processing, translating and transmitting data between the two systems. Mirth [21] is a far more advanced, full-fledged, open source healthcare messaging integration engine developed by WebReach, Inc., a healthcare IT consulting company based out of Irvine, California. Mirth is based on a unique client-server and Enterprise Service Bus (ESB) architecture and consists of connector, filter and transformer modules to send/receive, parse, and transform messages from HL7 v2 to legacy formats. Mirth has been adopted by several healthcare organizations to facilitate middleware services in their standard-based integration efforts. The latest and the greatest research and development efforts in electronic health is in the arena of Integrated Electronic Health Record (iEHR). Currently Canada Health Infoway is spearheading projects to realize a SOA-based, shared Electronic Health Record system in Canada leveraging HL7 v3. EHR Infostructure (EHRi) [6], an elaborate framework supporting architectural requirements, tools and environment necessary to build a pan-Canadian EHR has been developed by Infoway to drive the initiative.

We have studied an approach proposed by Dezhkam and Sartipi [17] for structuring business scenarios to extract UML design diagrams. In this paper we have extended their schema to healthcare transactions by incorporating their characteristics and composition in a messaging context. This schema is used to formally express healthcare transactions for mapping on to HL7 messages by our tool.
Overall, there's an increasing trend towards standard based integration of legacy systems leveraging emerging technologies such as Web Services, SOA and ESB [20, 23, 15]. Our mission is to contribute towards legacy system interoprability by providing guidelines, well-defined processes and tool-support to reduce complexity and improve return on investment (ROI) and turnaround time of HL7 v3 standard-based integration projects.

3. Healthcare standards and technologies

This section introduces healthcare standards and various technologies that are applicable to this paper. Health Level 7 (HL7): HL7 is a non-profit organization comprised of healthcare subject matter experts and IT professionals collaborating to develop international standards for exchange, management and integration of healthcare information in electronic format.

SNOMED CT: It is a comprehensive multilingual, clinical terminology offering a consistent way of indexing, storing, retrieving and aggregating clinical data across specialties and sites of care. SNOMED CT is organized into a hierarchical ontology with each term attached to a concept code, descriptions and relationships with other concepts.

LOINC: It is a database of codes representing terms used primarily in the Laboratory and Observation areas of healthcare.

Resource Description Framework (RDF): RDF [18, 9] consists of entities and binary relationships or statements between those entities represented as subject-predicate-object triples. In graphical notation of RDF, the source of the relationship is called the subject, the labeled arc is the predicate (also called property), and the relationships destination is called the object. The RDF data model distinguishes between resources, which are Uniform Resource Identifiers (URIs) representing a unique concept, property or object, and literals which are just strings. The subject and the predicate of a statement are always resources, while the object can be a resource or a literal.

Our tool uses RDF to represent and store metadata information about HL7 artifacts. Semantic Web technologies such as RDF offer a rich platform to implement efficient and accurate semantic search capabilities. In recent years a number of Semantic Web (SW) languages such as Web Ontology Language (OWL) [11], Ontology Inference Layer (OIL) [18] and DARPA Agent Markup Language (DAML) [14] have been developed upon RDF. Even though they offer improved descriptiveness, RDF remains the lowest common denominator among all and offers sufficient expressivity and precision for our tool.

Sesame framework: Sesame [12] is an open source Java framework for storing, querying and reasoning with RDF and RDF Schema. It can be used as a database for RDF and RDF Schema, or as a Java library for applications that need to work with RDF internally. Sesame consists of a Sesame library, Sesame server and Sesame repositories.

4. Proposed message extraction process

HL7 v3 Information Models are hierarchical in organization. At the highest level, Reference Information Model (RIM) provides the core classes from which all information structures are derived by consistent application of a standard set of constraints. RIM derives Domain Message Information Models (D-MIM) at the next level. Each D-MIM represents information pertaining to a core sub-domain of healthcare. Information of a topic is represented by a Refined Message Information Model (R-MIM) which is a refinement of the D-MIM. Hierarchical Message Description (HMD) which is a tabular structure defining message specification and Message Types are derived from R-MIMs.

HL7 organization has formed a number of technical committees to develop its information models and specifications. Each such committee is responsible for standardization of a single domain. A D-MIM may further be refined into"topics" whose names and numbers are decided by the technical committee in charge of the domain. While HL7 has dictated the manner and rules with which RIM is refined to derive subsequent data structures, no hard and fast rules have been laid out to guide how various topics and sub-domains are abstracted out within a domain. As a direct result, there is a level of inconsistency amongst peer information models of different domains. For example, Activate, Revise and Nullify message types of the Patient Billing topic of Account and Billing domain derive from the same topic and hence same R-MIM (Patient Billing Account Event), whereas the same message types for Person topic of Patient Administration domain derive from three different R-MIMs. Furthermore, even though some topics have been grouped together into one domain, possibly since their information models closely resemble each other, they seem to lack a close relationship at conceptual level. For example, topics Allergies, Care Plan and Clinical Document all fall under domain Care Provision. There are also instances where closely related concepts such as Allergies and Adverse Reactions are provided as separate topics.

The complexities associated with organization of HL7 artifacts pose difficulties for IT professionals in identifying appropriate message structures for use during system integration. As a result message workflow design with HL7 v3 typically involves top-down analysis of the entire information model hierarchy.

This section describes a process that simplifies HL7 v3 message extraction with the use of a new tool. The tool searches for HL7 Interactions (Section 4.1) suitable to represent real-world healthcare transactions. Following sections detail the proposed process and underlying concepts.

4.1. HL7 Interactions

HL7 Interactions provides critical contextual information required to interpret the semantics of a message correctly. HL7 v3 Interaction is a single, one way information flow. An Interaction explicitly answers the questions:

1. What the particular message type is (Message Type)
2. What caused the message to be sent (Trigger Event)
3. How a receiving system knows the type of response message to send if any (Receiver Responsibilities)

The Trigger Event that caused a particular message to be sent is encoded in the Control Act Wrapper associated with a message. While the Message Type contains the content of the message, Control Act tells the recipient how to act on that content. It’s critical for healthcare messages to be self contained especially when they are sent asynchronously or in a batch. Also, Receiver Responsibilities attached to an Interaction specifies the subsequent exchanges of information required to complete a transaction. Thus, in order to claim compliance with HL7 v3, a

Context	HL7 Domain and D-MIM	Description
Accounts and Billing	Accounts and Billing (FIAB DM000000UV)	accounts and billing, financial transactions, payment
Blood, Tissue and Organ Donation	Blood, Tissue and Organ Donation (POBB DM100000UV)	donation event, eligibility for donation, blood transfusions, blood bank
Care Provision	Care Provision (REPC DM000000UV)	patient care episodes
Care Record	Care Provision (REPC DM000000UV)	record of care
Allergies	Care Provision (REPC DM000000UV)	allergies, intolerance, adverse reactions
Care Transfer	Care Provision (REPC DM000000UV)	transfer of care provider
Patient Health Condition	Care Provision (REPC DM000000UV)	patient medical conditions
Family/Surgical History	Care Provision (REPC DM000000UV)	family history, surgical history
Discharge Report	Care Provision (REPC DM000000UV)	discharge report
Specialized Care and Professional Services	Care Provision (REPC DM000000UV)	specialists, physiotherapy, psychology, counseling
Referral Report	Care Provision (REPC DM000000UV)	referral report
Claims and Reimbursements – Special Authorization	Claims and Reimbursements (FICR DM000001UV)	insurance special authorization
Claims and Reimbursements - Eligibility	Claims and Reimbursements (FICR DM000001UV)	insurance eligibility
Claims and Reimbursements – Preapproval	Claims and Reimbursements (FICR DM000001UV)	insurance pre-approval
Claims and Reimbursements – Predetermination	Claims and Reimbursements (FICR DM000001UV)	insurance pre-determination
Claims and Reimbursements – Coverage	Claims and Reimbursements (FICR DM000001UV)	insurance coverage extension

Table 1. A portion of our HL7 v3 interaction categorization into Contexts and their affiliated D-MIMs.

healthcare transaction must be mapped to the correct set of Interactions and not a Message Type.

4.2. Extracting metadata

For the proposed search tool, HL7 v3 Interactions must be stored in a repository. Also, specific relationships between real-world healthcare transactions and Interactions must be established and built into the tool's mapping logic. However, this is a difficult task to achieve due to a number of reasons. Firstly, relationship between transactions and Interactions are not explicit or obvious in the HL7 v3 specification. Further, real-world healthcare transactions are not a bounded set and the same transaction could be expressed in many different terms using natural language. Thus, creating a one-to-one mapping between transactions and Interactions was not a solution. Instead we needed to extract metadata to describe Interactions and healthcare transactions using common concepts.

The following 4 steps describe the development and usage of the tool in a nutshell: *1) through careful analysis of HL7 Interactions and domain model, identify searchable metadata 2) express healthcare transactions in a structured, machine-readable language using the above identified* *metadata 3) store Interactions along with associated metadata 4) use structured transactions to search for matching Interactions.*

As described in the introduction to this section, we observed that existing HL7 domain model does not facilitate efficient discovery of Interactions due to overlaps and disconnects among the domains. Therefore, using a holistic view of the HL7 information model, we developed a more intuitive categorization for HL7 domains. We call these ”Contexts” to avoid confusion with original HL7 domains. We have come up with 50 Contexts to represent areas of healthcare that superimpose well with actual healthcare transactions. We have experimented with a large number of real world healthcare scenarios to test the effectiveness of these new categories. Each Interaction was then associated with a single Context. Context acts as a key piece of metadata in the search tool. A list of 16 (out of 50) contexts along with the affiliated D-MIMs are presented in Table 1.

Next, we classified Interactions into a hierarchy of classes based on the purpose of the message they convey. This classification has been designed to be intuitive and general enough so that it could also be used to formally express a transaction as described in Section 4.3. We call this classification the Interaction Classification Model and discuss it in detail in the next section. The class of an Interaction is the next key piece of metadata that would drive our tool.

4.3. Healthcare transaction schema

We define a ”Transaction” as a set of messages exchanged between two or more distinct systems in order to complete a particular task. Our approach to expressing healthcare transactions in a structured language was based on a technique proposed by Dezhkam and Sartipi for structuring business scenarios for design automation [17]. The proposed transaction schema is illustrated in Figure 1.

A message conveys some information required to complete the overall goal of the transaction it belongs to. In the transaction schema, each message can be viewed as a composition of constituents Actor, Operation and Data. All messages have one sender and one or more receivers. Combined, we refer to these components as Actors participating in a message exchange.

The remainder of the message can be further decomposed into Operational and Informational components. Operational component, referred in our schema as ”Operation” is a generalization of the action contained in the message description. For example, consider message EMR requests EHR for patient allergy details. The action contained in message description which is, ”request for allergy details” can be generalized as ”request for information”. This would form the Operation component of the message. The remaining information in the message acts as qualifiers for the Operation and is collectively known as Data. Data derives Content and Context components. Content refers to fields of data that need to be communicated to the receiver. We associate each message with the Context that best describes its domain affiliation. It is any one of the Contexts described in Table 1. The most appropriate context to represent the above message would be ”Allergy”.

The high level schema of a transaction can be expressed in regular expression syntax as follows. Here ”+” stands for composition and ”1..N” represents multiplicity:

$$\textit{Transaction: } \{\textit{Message}\}^{1..N}$$
$$\textit{Message: } \{\textit{Actor}\}^{2..N} + \textit{Operation} + \{\textit{Data}\}^{1..N}$$

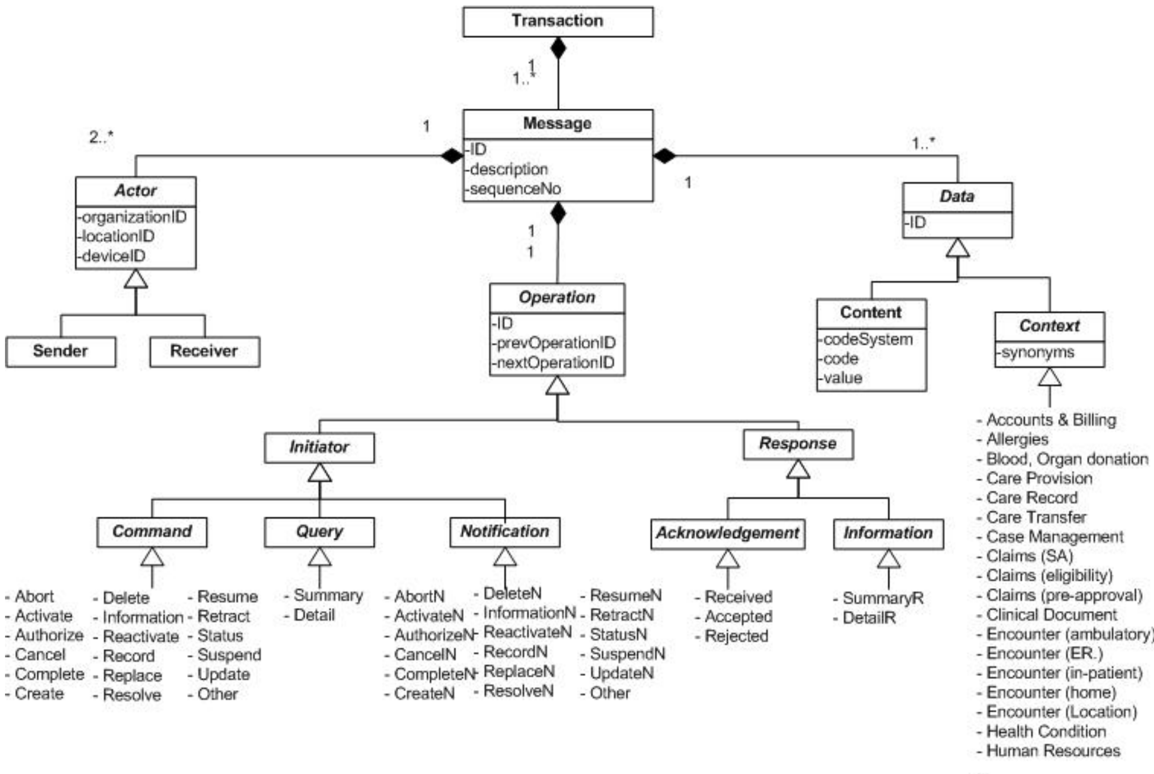

Figure 1. Healthcare transaction schema. The schema describes a healthcare transaction as a composition of Actors, Operations and Data.

Interaction classification model

Now that we have decomposed healthcare Transactions and Messages into their components, the next step is to map them to HL7 v3 Interactions. As mentioned in the previous section, we have classified HL7 v3 Interactions into a hierarchy of classes based on the core operation they perform. Each Interaction is also assigned a context from the set of Contexts we have identified.
Thus an Interaction is described by a Context and an Operation class. For mapping Interactions to Healthcare messages, we match:

1. Context of healthcare message with that of HL7 v3 Interaction.
2. Operation component of healthcare message with Operation class of HL7 v3 Interaction.

The Interaction classification model classifies HL7 v3 Interactions at three levels of abstraction. Abstraction Level I: in this level an Interaction is divided into Initiator and Response. Initiator class represents Interactions that initiate an information exchange. For example, “a request for a report” or “reporting an event such as a medical appointment” fall under Initiator category. Response class represents Interactions that are non-initiators and are sent by a receiver in response to a previous message.

Abstraction Level II: in this level Initiator Interaction is further divided into Command, Query and Notification. Command refers to an Interaction that orders the receiver to perform a task. Query represents requests for information. Notification refers to Interactions that notify a third party of

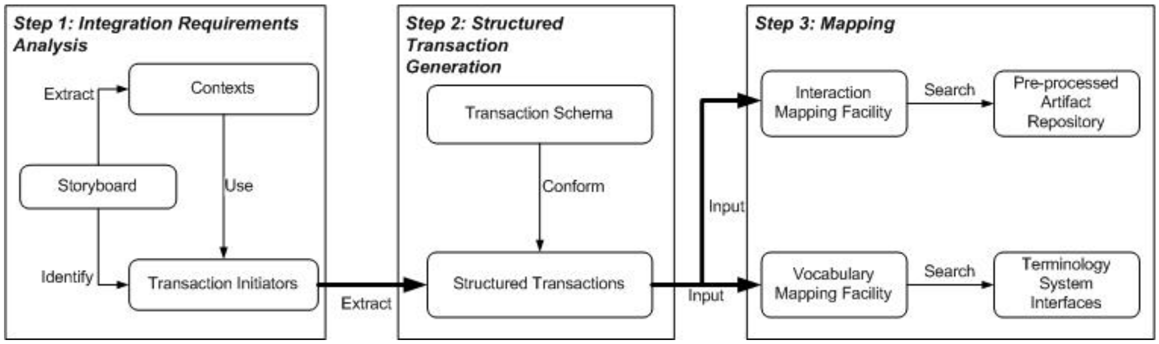

Figure 2. Message extraction process that maps healthcare transactions to HL7 messages.

occurrence of an event. Response class is divided into Acknowledgement and Information. Acknowledgement is an Interaction that is sent to acknowledge its receipt in general and at times, the accept/reject status by the receiver. Information represents Query results or information sent as response to a Command.

Abstraction Level III: in this level Command and Notification each divided into 18 sub-categories based on the nature of the requested task. Abort, Activate, Update, Retract and Record are some examples. Level II type Query is divided into Summary and Detail based on level of detail in information requested. Acknowledgement is divided into Received, Accepted and Rejected, representing the status of the message. Information is further sub-divided into Summary and Detail based on the level of detail. The definition of each class in the Interaction Classification Model is provided in Table 2.

The search tool facilitates input of the structured transactions as search criteria. It then matches different components of the input transaction with metadata stored against Interactions to find the best match. Semantic Web techniques have been used to implement the mapping operation. The technical concepts behind the tool are described in detail in Section 4.5.

4.4. Proposed process

In this section, we explain details of each step of the message extraction process using a running example.

Step 1: Integration Requirements Analysis.

This step involves examining information exchange requirements of the systems being integrated. Typically, this step starts by laying down system requirements as an informal narrative called the Storyboard. Such information is often incomplete, unstructured and therefore of little use for automation. We propose the following four sub-steps to identify and structure messaging requirements.

1.1 Storyboarding. As explained above, prospective system users are asked to write business scenarios using their own terms. Several storyboards may be required to lay down all requirements for a particular system. Each storyboard is then entered into the tool. We take real-life scenarios in the Storyboard ”Visit to Physician to Refill Prescription” given below as our running example.

Class	Definition
Initiator	Interaction initiating a conversation with a receiving system.
Query	Query receiver for information.
Detail	Find all possible candidates matching search criteria.
Summary	Retrieve a particular record by ID.
Command	Order the receiving system to perform an operation.
Abort	Order receiving system to abort a previously activated operation.
Activate	Order receiving system to activate an account
Authorize	Order receiving system to authorize an operation/document.
Cancel	Order receiving system to cancel a previously activated operation.
Complete	Order receiving system to complete a previously activated operation.
…	
Notification	Notify receiver(s) of occurrence of an event or operation
Abort	Notify receiving systems of an abort operation.
Activate	Notify receiving systems of an activate operation
Authorize	Notify receiving systems of an authorize operation.
…	
Response	Respond to a command, query or notification.
Acknowledgement	Acknowledge the receipt of a message indicate if the requested command/notification is accepted for processing.
Received	Acknowledge that a particular message was received
Accepted	Inform that the receiver accepts to process a Command/ Query /Notification.
Rejected	Inform that the receiver rejects to process a Command/ Query /Notification.
Information	Response to a command to send information/query.
Summary	Summary information response. Detailed information response
Detail	

Table 2. Definition of a portion of the classes in the Interaction Classification Model.

"Mr. X needs to get a repeat of his usual medications -Glyburide 5 mg tid, Metformin 500 mg tid once daily (od) and Celebrex 100 mg od. He visits his FP, Dr. P. Dr. P pulls up Mr. X's chart in her EMR, which automatically queries the EHR for current medication, allergy history and medical conditions and downloads the information to her EMR. Dr. P updates her EMR with Mr. X's new allergy. She also notes that Mr. X's last HbA1c (a measure of long-term glucose control) was high and recommends that Mr. X start a new medication, Roziglitazone 4 mg od. She then re-prescribes for Mr. X all his usual medications using her EMR. Once Dr. P is satisfied that there are no drug-drug interactions, she initiates a transfer of the prescription to the EHR and tells Mr. X that she has prescribed the medications for him with 3 repeats and that he can pick them up from the pharmacy of his choice. When Dr. P closes Mr. X's chart on her EMR, it automatically updates the EHR with the updated information he has agreed to send; in this case just the allergies."

1.2 Extract Contexts. The tool searches storyboard text entered in Step 1.1 to create possible semantic maps between Contexts and words and phrases in the text. Within our tool, each Context has been annotated with Cognitive Synonyms describing it. We have used WordNet [1] and SNOMED vocabularies to incorporate as many cognitive synonyms and phrases to describe each Context.

In the future we intend to enhance this feature by using Natural Language Processing (NLP) concepts. This exercise is useful to execute successfully Step 1.3, where users identify transactions that are conceptually linked to existing HL7 domains. However, the automatic mapping is not a definitive map and can be refined or replaced by the user later.

Transaction initiator	Interaction
EMR sends request for patient medication history	Medication Profile Detail Generic Query (PORX IN060350UV)
EMR sends request for patient allergies.	Patient adverse reactions query (REPC IN000058UV)
EMR updates EHR with medication.	Medication Order Record Request (PORX IN010380UV)
EMR updates EHR with allergies.	Record adverse reaction request (REPC IN000004UV)
EMR sends prescription request to pharmacy.	Medication Order Fulfillment Request (PORX IN011070UV)

Table 3. Interactions mapped for Storyboard Medication Refill.

For the storyboard in the running example, some possible context maps include:
1. Medication: Pharmacy
2. Allergies: Allergies
3. Medical conditions: Clinical document
4. Prescriptions: Pharmacy

1.3 Identify Transaction Initiators. We define Transaction Initiator as the starting message in a sequence of messages completing a transaction. Transaction Initiators can be easily identified manually from storyboard text. Contexts identified in step 1.2 must be kept in mind to keep these transactions relevant to HL7 domains.

For our running example, possible Transaction Initiators are:
1. EMR sends request for patient medication history.
2. EMR sends request for patient allergies.
3. EMR updates EHR with medication.
4. EMR updates EHR with allergies.
5. EMR sends prescription request to pharmacy.

Step 2: Structured Transaction Generation.

Each transaction initiator is then structured according to the proposed transaction schema so that they are in machine readable format.

For our running example, Transaction Initiators identified in 1.3 can be expressed as follows:

EMR sends request for patient medication history.
Actor: EMR
Operation: QueryDetail
Context: Pharmacy
Content: medication history

EMR sends request for patient allergies.
Actor: EMR
Operation: QueryDetail
Context: Allergy
Content: patient allergies

EMR updates EHR with patient medication.
Actor: EMR
Operation: CommandUpdate
Context: Pharmacy

Transaction initiator	Interaction
HIS sends EHR a request for Health Record	Patient health condition details query (REPC IN000025UV)
HIS sends a request to EHR to update demographic information	Patient Registry Revise Request (PRPA IN201314UV02)
HIS sends ER record to X's Primary care physician	Emergency Encounter Ended (PRPA IN403003UV02)
HIS sends ER record to EHR	Emergency Encounter Ended (PRPA IN403003UV02)

Content: patient medication

Table 4. Emergency encounter - mapping of transaction initiators and interactions.

Step 3: Mapping.

3.1 Interaction Mapping. Structured transactions extracted in Step 2 are entered into the tool using its web interface. The tool's advanced semantic search feature searches a history archive to locate if similar search criteria have been used successfully before. If not, the main artifact repository is searched. The user can confirm or reject the results. If confirmed, user can choose to save search criteria and results in the history archive. For the running example, Table 3 gives Interactions returned in the mapping step.

3.2 Vocabulary Mapping. While the previous steps ensure HL7 compliance for message schema, this step ensures that data fields communicated are interpreted accurately by the receiver. This is achieved by converting local terms to standard terminology codes for transmission. The tool integrates with terminology systems SNOMED and LOINC to search for the most appropriate code for a particular legacy clinical term. Data fields extracted during Step 1 are used as search criteria.

Case Study - Emergency Encounter

Storyboard: *"Mr. X arrived at hospital emergency room via ambulance. Mr. X was in respiratory distress and had an accelerated heart beat. The physician on duty, Dr. E, decided Mr. X should be treated at this time. Mr. X was checked-in for an ER visit. The emergency room clerk pulled up Mr. X's health record in the HIS4 which automatically quizzes the EHR. Clerk created the emergency check-in. The ER clerk reviewed the contact information in Mr.X's patient record with him. Mr. X stated that he needed to change his emergency contact information. Mr. X's daughter was out of town so Mr. X informed that he wanted to put his son, Mr.S, down as the emergency contact. He provided S' phone number and address. System was updated and notification sent to EHR. ER specialist, Dr. E decided that after a nebulizer treatment Mr. X was stable and was ready to be checked-out. Dr. Emergency noted that Mr. Everyman needed to schedule a follow-up visit with Dr. P, pulmonologist. The ER clerk completed the check-out information for Mr. Everyman and checked him out of the Emergency Room. HIS sends EHR the Mr. X's emergency record. His primary care physician, Dr. P was also sent the emergency record."*

Context maps:
1. Emergency - Encounter (Emergency)
2. Health record - Health Condition
3. Patient registry - Patient Administration
4. Pulmonologist - Specialized Care

Transaction initiators:
1. HIS requests EHR for Health Record.
2. HIS requests EHR to update demographic information.
3. HIS sends emergency record to X's Primary care physician.
4. HIS sends emergency record to EHR.

Structured transactions:
HIS requests EHR for Health Record.
Actor: HIS

Operation: QueryDetail

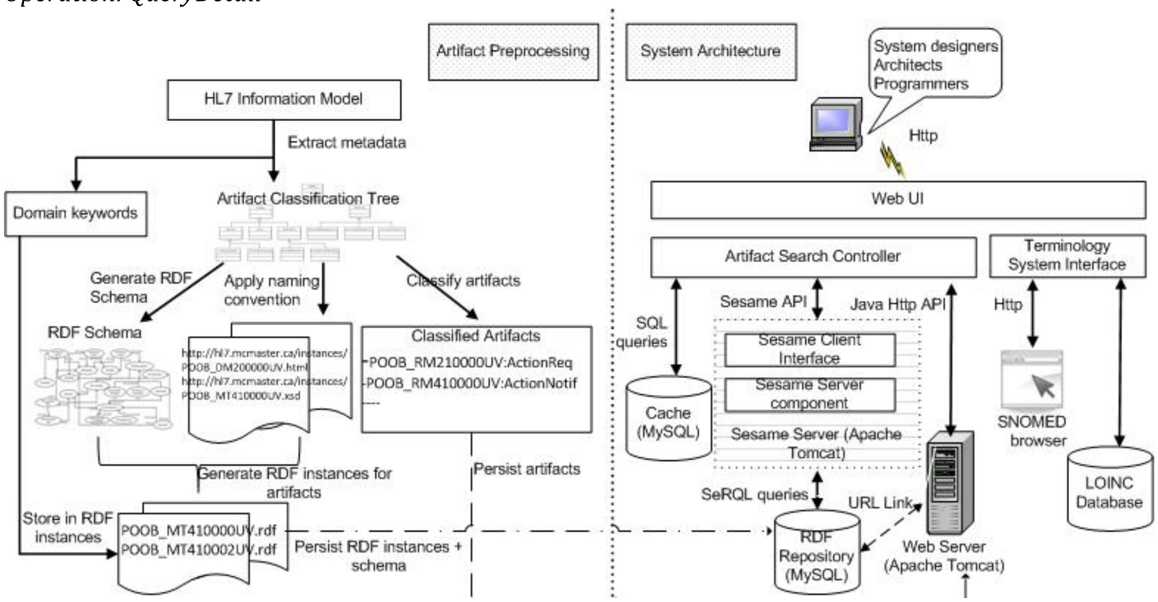

Figure 3. Tool architecture. Left part illustrates steps involved in pre-processing HL7 v3 artifacts persisting in artifact repositories. Right part illustrates the system architecture.

Context: Health Condition
Content: health record

HIS requests EHR to update demographic information.
Actor: HIS
Operation: CommandUpdate
Context: Patient Administration
Content: demographic information

HIS sends emergency record to X's Primary care physician.
Actor: HIS
Operation: NotificationInformation
Context: Emergency Encounter
Content: emergency record

HIS sends emergency record to EHR.
Actor: HIS
Operation: NotificationInformation
Context: Emergency Encounter
Content: emergency record

Table 4 presents the transaction initiators and corresponding interactions mapped to them.

4.5 The tool

We developed an open source, web-based tool to supports message workflow design activities associated with HL7 v3 based integration. The architecture of the tool is illustrated in Figure 3.

Architecture

The tool consists of Web UI, Web Server, Sesame Server, Artifact Search Controller and Terminology System Interface components. The Java Servlet based GUI is a user friendly environment for searching, browsing, navigating and exploring artifacts. Web Server is an Apache Tomcat server where HL7 artifacts such as xml schema, documentation, xml sample instances, information models and other representations are stored. Sesame Server consists of an RDF repository, Server component and Sesame API. RDF repository is a MySQL database of RDF instances with metadata pertaining to each artifact. Sesame server component handles connections and communications with the RDF repository to execute search and retrieve RDF instances. Artifact Search Controller accesses Sesame infrastructure with Sesame API to leverage its services. It also parses retrieved RDF instances to obtain URLs of artifacts referred to, and requests the web server for them. The Terminology System Interface supports searching for SNOMED and LOINC codes for local terms by integrating into existing SNOMED browser by BT [16] and a MySQL database of LOINC codes. The right portion of Figure 3 illustrates the high level architecture.

RDF-based search and retrieval

Our approach to implementing semantic search is to create an RDF instance with metadata for each HL7 Interaction. The RDF instance will carry information such as other HL7 artifacts related to a particular Interaction, including: the ”Operation” class that best represents it, ”Context” that it belongs to, and “keywords” that are cognitive synonyms for this context. Each D-MIM has also been associated with appropriate keywords and phrases that describe information it represents. Left part of Figure 3 indicates the activities involved in offline artifact pre-processing stage. This metadata model was converted to an RDF schema by applying rules of RDF syntax and semantics specified by W3C. Since RDF requires all resources to be uniquely identifiable, we adopted an artifact naming convention based on their HL7 artifact ID which is unique. For example, Observation Request message schema will be named POOB MT210000UV.xsd based on its HL7 artifact ID POOB MT210000UV.

Finally, RDF instances describing the metadata and relationships of each artifact are generated in conformance with the schema and by analyzing the HL7 information models. Artifacts are persisted in the Web Server and RDF instances are stored in the RDF repository for access by the application.

At runtime, the user accesses the search function and inputs domain keywords or phrases such as ”patient”, ”vital signs” or ”blood donation” that appear in use case descriptions derived in the Integration Requirements Analysis step. User may also select the artifact classification tree node that best describes the use case. The Search Controller then generates SeRQL [13] queries based on the search criteria and accesses the RDF repository via Sesame API. Depending on the strength of search criteria, more than one match per use case may be returned. Information in resulting RDF instances will be parsed and a hierarchy of related URLs pertaining to each RDF instance will be retrieved from the web server which is displayed in a user friendly, browse-able format. Message schema retrieved can also be downloaded via the tool. Figure 4 illustrates RDF Graph of HL7 v3 Interaction metadata model.
A section of the RDF instance for Interaction ”Re- quest to record subject observation” (Artifact ID POOB IN000001UV) that is persisted in the RDF store is as follows:

```
1. <?xml version="1.0"?>
2. <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-
3.syntax-ns#"xmlns:hl7="http://hl7.mcmaster.ca/hl7/schema#">
4. <hl7:Interaction rdf:about="http://hl7.mcmaster.ca/
5. instances/POOB IN000001UV.html">
```

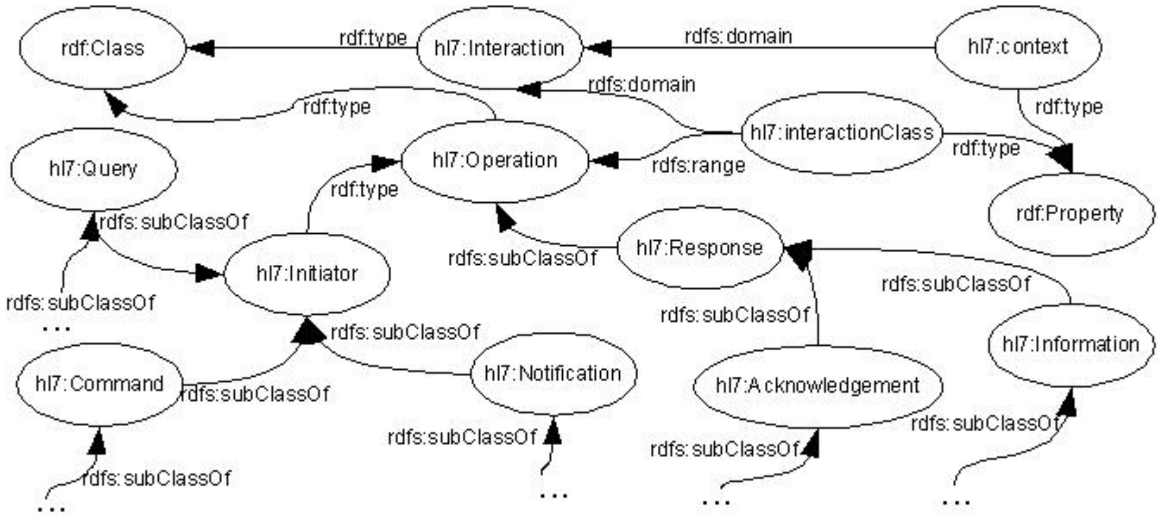

Figure 4. RDF Graph for HL7 artifact metadata model.

```
6. <hl7:interactionClass rdf:resource=
7. "http://hl7.mcmaster.ca/
8. schema/Record"/>
9. <hl7:context>observation,
10. non-laboratory observation,vital signs,
11. height,weight,blood pressure
12. </hl7:context>
13. </hl7:Interaction>
14. </rdf:RDF>
```

Line 2 begins an rdf:RDF element. On the same line, there is an XML namespace with URI address http://www.w3.org/1999/02/22-rdf-syntax-ns# which defines the meanings of the tags from the RDF vocabulary (prefixed by "rdf:"). Line 3 specifies the XML namespace with URI address http://www.mcmaster.ca/hl7/schema# that defines the tags for HL7 specific terms.
Line 4 indicates that the current RDF describes an instance of class Interaction. The URI of the resource described in the RDF is given in the "about" attribute in lines 4 and 5. Lines 6 to 12 specify some of the properties of this resource. Lines 6, 7 and 8 indicate that value of property interactionClass for the resource is http://hl7.mcmaster.ca/ instances/Command. Lines 9, 10 and 11 specify that the value of property context is "observation,clinical observation, vital signs, height, weight, blood pressure".

5. Conclusion

Increasingly, governments of many countries including Canada are recognizing the importance of the role of Information Systems in improving the quality of public health services. While IT companies and healthcare institutions engage in such collaborations, the research community has a vital role to play in conducting innovative research aimed at solving various technological issues that continue to be bottlenecks. In this paper, we presented a novel and well-defined approach for designing message workflow during HL7 v3 based system integration. We also presented a behavior-based classification for HL7 v3 Interactions that superimposes well with real-life healthcare transactions. We described a tool developed using semantic web technologies that uses this classification model to drive advanced search operations. The aim of the proposed approach and the tool is to reduce domain-dependant complexities for software professionals performing healthcare system integration using HL7 v3. This would in turn improve efficiency and ROI of projects by eliminating the necessity to involve domain experts at the design phase. Techniques used in the design and implementation of this tool can easily be adopted in other Enterprise Search and Knowledge Management applications. For future research, we will continue to improve the tool to provide extra features such as: automated message instantiation and schema editing to render Healthcare Transaction to HL7 v3 Message Translation as seamless as possible.

References

[1] WordNet. http://wordnet.princeton.edu/.
[2] 7Edit. http://www.7edit.com/home/index.php.
[3] 7Scan HL7 browser. http://www.7scan.com/.
[4] Health Level 7 website. http://www.hl7.org.
[5] Canada Health Infoway website. http://www.infowayinforoute.ca/en/ home/home.aspx.
[6] Infoway EHR Infostructure. http://www.infowayinforoute.ca/en /WhatWeDo/Infostructure.aspx.
[7] LOINC website. http://www.loinc.org.
[8] SNOMED CT website. http://www.ihtsdo.org/snomed-ct/.
[9] W3C RDF Primer. http://www.w3.org/TR/2004/ REC-rdfprimer-20040210/.
[10] Representing UML in RDF. http://infolab.stanford.edu/melnik/rdf/uml/.
[11] Web Ontology Language. http://www.w3.org/2004/OWL/.
[12] Sesame. http://www.openrdf.org/index.jsp.
[13] Sesame RDF Query Language. http://www.openrdf.org/doc/sesame/users/ch06.html.
[14] S. Arroyo, R. Lara, Y. Ding, M. Stollberg, and D. Fensel. Semantic web languages - strengths and weaknesses. http://members.deri.at/ michaels/publications/swlanguages-IADIS04.pdf.
[15] D. Berler and S. Pavlopoulos. Design of interoperability framework in a regional healthcare system. In EMBS, pages 3093–3096, 2004.
[16] BT. SNOMED browser. http://www.jdet.com/.
[17] N. Dezhkam and K. Sartipi. Knowledge Transformation from Task Scenarios to View-based Design Diagrams. In SEKE, 2008.
[18] D. Fensel. The SemanticWeb and its Languages. Intelligent Systems, pages 67–73, nov-dec 2000.
[19] R. Guha, R. McCool, and E. Miller. Semantic search. Pages 700–709. ACM, Hungary, MAY 2003.
[20] B. Liu, X. Li, Z. Liu, Q. Yuan, and X. Yin. Design and Implementation of information exchange between HIS and PACS Based on HL7 standard. In Proc. of the 5th ITAB, China, pages 552–555, may 2008.
[21] B. Liu, X. Li, Z. Liu, Q. Yuan, and X. Yin. Experiences with Mirth: An Open Source Health Care Integration Engine. In Proc. of ICSE, Germany, pages 649–652. ACM, may 2008.
[22] NeoTool. NeoBrowse HL7 browser. http://www.neotool. com/products/neobrowse.
[23] T.H.Yang. A scalable multi-tier architecture for the National Thaiwan University hospital information system based on HL7 standard. In CBMS, Germany, pages 99–104. IEEE, 2006.
[24] M. Yarmand and K. Sartipi. Standard-based Data and Service Interoperability in eHealth Systems. In ICSM, 2008. pages: 187-196.