HL7 v3 message extraction using Semantic Web techniques

Priya Jayaratna

Department of Computing and Software,
McMaster University,
1280 Main Street West, Hamilton, ON, L8S 4K1, Canada
E-mail: priya_jayaratna@hotmail.com

Kamran Sartipi*

Faculty of Engineering and Applied Science,
University of Ontario Institute of Technology (UOIT),
2000 Simcoe Street North, Oshawa, ON, L1H 7K4, Canada
E-mail: kamran.sartipi@uoit.ca
*Corresponding author

Abstract: Healthcare system integration is an area of utmost importance in the overall eHealth strategy of countries. The overall goal of these efforts is to provide a large scale and unified view of clinical information to healthcare practitioners, thereby enabling them to deliver accurate and timely services to the general public in a cost-efficient manner. In this paper, we present a novel framework for identifying HL7 v3 messages to represent healthcare transactions that take place in an integration scenario. The proposed technique provides a new categorisation of HL7 v3 message functionality according to a set of message contexts extracted by extensive study of HL7 v3 information hierarchies and messaging infrastructure. These contexts allow us to map the key terms in a healthcare scenario to the corresponding HL7 v3 messages using Semantic Web technology. We have developed a prototype tool and will present two healthcare case studies to demonstrate our solution.

Keywords: health informatics; healthcare system; Semantic Web; SW; HL7 v3; scenario; system interoperability; information model; interaction; transaction; context; knowledge management.

Reference to this paper should be made as follows: Jayaratna, P. and Sartipi, K. (xxxx) ‘HL7 v3 message extraction using Semantic Web techniques’, *Int. J. Knowledge Engineering and Data Mining*, Vol. X, No. Y, pp.000–000.

Biographical notes: Priya Jayaratna graduated from McMaster University with a Masters in Software Engineering in 2009. She is also a certified HL7 v3 Specialist and has over ten years of experience designing and developing enterprise applications for improving healthcare delivery across North America. She is currently working at eHealth Ontario as a Technical Consultant specialising in healthcare application integration.

Kamran Sartipi is an Associate Professor at the Faculty of Engineering and Applied Science of the University of Ontario Institute of Technology (UOIT). He received his BSc and MSc in Electrical Engineering from Tehran University and MMath and PhD in Computer Science (Software Engineering) from the University of Waterloo. He is a licensed Professional Engineer. He is the author or co-author of more than 50 published peer-reviewed papers in prestigious journals and conferences, as well as book and book chapters in different scientific areas. He has been the co-developer of a new graduate programme eHealth (Health Informatics) at McMaster University among three faculties business, health sciences, and engineering. His research is inter-disciplinary, including: software engineering and revere engineering, knowledge engineering, data mining, service and semantic interoperability, work-flow analysis, data privacy, decision support systems, and embedded systems.

1 Introduction

Proper management of healthcare information is of paramount importance in providing safe and efficient patient care. Recent studies show that adoption of information technology in healthcare can result in improved quality of care, prevention of medical errors, reduction in healthcare costs and increase in administrative efficiencies (Department of Health and Human Services, http://healthit.hhs.gov/; Southern California Evidence-based Practice Center, 2006; Chaudhry et al., 2006). However, compared with other business domains such as banking, telecommunication and media, the IT spending and adoption rates of the healthcare industry have historically been lagging behind (HIMSS, 2008). The slow pace of IT adoption in healthcare can be attributed to a variety of factors. Doubts about benefits vs. costs of IT, high initial cost of implementation, ongoing support and maintenance concerns and reluctance to change business practices to accommodate information systems are amongst major contributing factors. Governments and healthcare authorities have also been reluctant to back wider adoption of IT in healthcare systems due to prohibitive initial costs of implementation, lack of conclusive evidence of return on investment (ROI) and information security and privacy concerns (Vishwanath and Scamurra, 2007; Massachusetts Technology Collaborative, 2003).

However, with educational institutions taking the lead in providing IT education to health professionals, there is a new generation of cross-domain experts capable of changing the attitude of the healthcare community towards information systems. New research carried out by national and international healthcare IT organisations such as the Canada Health Infoway (http://www.infoway-inforoute.ca/lang-en), Agency for Healthcare Research and Quality (Southern California Evidence-based Practice Center, 2006) and Healthcare Information and Management Systems Society (HIMSS, 2008) indicates tangible benefits of using information technology in healthcare. The US Congressional Budget Office (http://www.cbo.gov/ftpdocs/99xx/doc9968/hr1.pdf) estimates that the use of electronic medical records could save the nation $12.5 billion over ten years. As a result, harnessing IT for betterment of health services is becoming an integral part of the healthcare strategy of governments worldwide. The current US Government has pledged $20 billion towards implementing a nationwide electronic health record (EHR) by the year 2014. The federal government of Canada also invested $500 million into Canada Health Infoway's EHR project in 2008, bringing up Canada's total investment in EHR to date to $2.1 billion (eHealth Ontario, http://www.ehealthontario.on.ca).

There is a flurry of activity in the field of healthcare informatics to help translate the change in attitude and commitment to tangible business results. The public require secure and easy access to individual health records; providers require patient medical information sharing capabilities; laboratories need to exchange order and result information with providers and peer-laboratories; pharmacies need to be integrated with practitioner networks and provide ePrescribing facilities to clients (Massachusetts Technology Collaborative, 2003). In summary, there’s a growing need for unification of information across application and organisation boundaries in a secure and reliable manner. Healthcare systems currently in use are as diverse as the healthcare domain itself. Information exchange between different organisations is still mostly non-electronic, largely depending on telephone, fax and e-mail. Even where systems have been integrated, they are mostly point-to-point integrations. This is obviously a rather non-scalable and maintenance-intensive approach. Thus, widely accepted and adhered-to standards are increasingly important in order to integrate healthcare systems.

This study focuses on addressing issues related to modelling and designing information communication amongst different systems. The following sections describe in detail the research problem addressed by this paper and the proposed solution.

2 Problem definition and solution

Achieving seamless integration amongst heterogeneous healthcare systems is not an easy task. One of the major barriers in implementing nationwide integrated solutions such as the EHR is the problem of interoperability (Massachusetts Technology Collaborative, 2003). Semantic interoperability refers to the ability of systems to correctly interpret concepts and terms used by another system. This can only be achieved through standardisation of information exchange and representation. Health Level 7 (HL7) (http://www.hl7.org) is the internationally accepted standard for healthcare information.

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 HL7 v2 for being a ‘true’ standard offering precision and un-ambiguity, the worldwide healthcare community has so far been reluctant to adopt it due to its overwhelming complexity. HL7 v3 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 (http://www.ihtsdo.org/snomed-ct) and LOINC (http://www.loinc.org) to represent information content.

Thus HL7 v3-based integration of systems requires a Herculean effort on the part of IT professionals to gain sufficient knowledge of the standard itself in order to perform message design tasks independently. Employing healthcare professionals to provide necessary domain knowledge would be costly and inefficient since typically they have little IT knowledge. Further, since HL7 is an evolving standard, integrators would require constant upgrading of their knowledge in order to be productive.

HL7 v3 is organised into a hierarchy of information models from which messages are progressively derived. These information models are described in detail in Section 4 on standards and technologies. HL7 organisation has formed a number of technical committees to develop its information models and specifications. Each such committee is responsible for standardisation of a single domain of healthcare represented by a domain message information model (D-MIM). A D-MIM may further be refined into ‘topics’ as refined message information domain (R-MIM). Topic 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.

The complexities associated with organisation of HL7 artefacts pose difficulties for non-domain-expert 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.

The tedious process of HL7 v3-based integration of systems can be improved tremendously by developing guidelines, processes and tools to support system integrators. However, to the best of our knowledge, well-defined frameworks and open-source tools supporting design and implementation of HL7 v3-based integration, are unavailable as of today. As such, message workflow design typically involves wading through pages of HL7 documentation with the help of a primitive text search alone. Thus, we define the problem of this study as:

> “Devising novel frameworks, techniques and tools to support HL7 v3 standard compliant integration of healthcare systems.”

We propose a process to guide users through the communication design phase of healthcare integration projects. The proposed process streamlines translation of healthcare scenarios into HL7 v3 messages in a seamless manner by using the concept of structured healthcare transactions. The process consists of three stages: *integration requirements analysis*; *structured transaction generation*; and *mapping*. The proposed process improves efficiency and accuracy of HL7 v3-based integration projects by aiding system developers with little knowledge of the standards to extract appropriate messages to meet communication requirements. Section 5 describes the proposed process in detail.

The proposed approach simplifies the process of identifying HL7 v3 messages required to represent real world healthcare scenarios. The approach relies on a search tool developed based on metadata extracted by extensive study of HL7 v3 information hierarchies and messaging infrastructure. The proposed approach takes advantage of scenario decomposition and structured scenario representation techniques proposed by Dezhkam and Sartipi (2008). An open-source, Semantic Web (SW)-based prototype search tool is built upon Sesame RDF framework. It provides advanced semantic search facilities for identifying and browsing HL7 artefacts suitable for representing a structured healthcare transaction.

2.1 Contributions

The contributions of this paper to the healthcare informatics field are as follows:

- Devised a novel, well-defined process to guide translation of healthcare transactions to HL7 v3 interactions.

- Re-categorised HL7 v3 interactions based on their behavioural traits in a messaging context. These categories provided valuable metadata to be used by the proposed search and mapping tool.
- Developed a prototype tool based on SW technologies to automate the process of identifying HL7 interactions appropriate to represent healthcare transactions.
- Extended an approach by Dezhkam and Sartipi (2008) for formal representation of business scenarios and adapted it to represent healthcare transactions.
- Demonstrated the use of the framework and the tool with two real world healthcare case studies.

3 Related work

3.1 HL7 V2 tools

Due to relative simplicity of HL7 v2 data model and message format, the process of building tool-support is straightforward and less complex. There are a number of widely used commercial support tools available for HL7 version 2. *7Scan* (http://www.7scan.com/) is a specialised browser and editor that finds, displays, edits and transmits text-based HL7 v2 messages with ease. 7Scan is an ideal tool to develop, test, and maintain HL7 interfaces. 7Scan can also be used as an endpoint simulator to send and receive messages with any HL7 interface being developed. 7Scan assists users to understand HL7 v2 messages by converting the coded, flat structured messages into hierarchical structures with user friendly field definitions. *7Edit* (http://www.7edit.com/home/index.php) is a productivity tool for browsing, editing, searching, validating HL7 messages and communicating with systems that support HL7 format. With 7Edit, HL7 v2 can be extended by creating Z-segments2 and message structures can be customised to meet integration needs unsupported by HL7 v2. 7Edit supports HL7 versions 2.1 up to 2.6. *NeoTool* is a company that provides healthcare systems integration and offers software solutions, consulting, and training for healthcare application interfacing. NeoTool’s HL7 Analyser (http://www.corepointhealth.com/products/hl7-analyzer/hl7-analyzer) (formerly NeoBrowses) offers a multi-view interface simplifying how a programmer can view, edit, test, validate, and repair any HL7 v2 message, increasing productivity.

3.2 HL7-based integration approaches

The HL7 v3 mapping process proposed in this paper is continuation of the work carried out by Yarmand and Sartipi (2008). Their proposed model for message standardisation is based on guidelines set forth by Canada Health Infoway (http://www.infoway-inforoute.ca/lang-en). 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. (2008a) discuss an HL7 v2-based integration project to establish interoperability between a hospital information system (HIS) and a picture archiving and communication system (PACS) based on DICOM. Mirth (Liu et al., 2008b) is a far more advanced, full-fledged, open source healthcare messaging integration engine. 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, transform messages from HL7 v2 to legacy formats. Mirth has been adopted by several healthcare organisations to facilitate middleware services in their standard-based integration efforts.

3.3 Electronic health record systems

HIMSS (http://www.himss.org/ASP/topicsehr.asp/) defines EHR as follows:

> "The electronic health record (EHR) is a longitudinal electronic record of patient health information generated by one or more encounters in any care delivery setting. The EHR automates and streamlines the clinician's workflow. The EHR has the ability to generate a complete record of a clinical patient encounter, as well as supporting other care-related activities directly or indirectly via interface including evidence-based decision support, quality management, and outcomes reporting."

The most extensive research and development efforts in electronic health are in the arena of integrated EHR. Currently Canada Health Infoway (Infoway) is spearheading projects to realise a service-oriented architecture (SOA)-based, shared electronic health record system (EHR-S) in Canada leveraging HL7 v3. EHR Infostructure (EHRi) (Infoway, http://www.infoway-inforoute.ca/en/WhatWeDo/Infostructure.aspx), 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. openEHR (http://www.openehr.org/home.html/) is an international not-for-profit foundation, working towards making the interoperable, life-long EHR a reality and improving healthcare in the information society through developing open specifications, open-source software and knowledge resources, engaging in clinical implementation projects, participating in international standards development and supporting health informatics education. Rong Chen (2007) describes implementation of an open source reference information model (RIM) for openEHR project. openEHR has also designed and developed a template-based EHR system called Julius that was integrated with existing EHR systems (Rong Chen and Klein, 2007).

3.4 Structured scenarios

We have studied an approach proposed by Dezhkam and Sartipi (2008) for modelling business scenarios for automation by decomposing scenarios into their constituent components. In this paper, we have extended their schema to healthcare transactions by capturing transaction actors, behaviour and data as their components. This schema is used to formally represent healthcare transactions for mapping on to HL7 messages by our tool.

Overall, there is an increasing trend towards modernising legacy healthcare IT infrastructure. Our research is concentrated on standard-based integration of legacy systems leveraging emerging technologies such as web services, SOA and ESB (Liu et al., 2008a; Yang, 2006). Our mission is to contribute towards legacy system interoperability by providing guidelines, well defined processes and tool-support to improve complexity, ROI and turnaround time of HL7 v3 standard-based integration projects.

3.5 Semantic analysis of medical data

As far as we know, there is no research performed on semantic analysis for the purpose of a semi-automatic mapping clinical narratives onto the standard HL7 v3 interactions. The current approaches mostly address the semantic analysis of the medical data for knowledge extraction. In our approach, we also apply three traditional techniques for semantic analysis: *synonym approach* is used in the WordNet's cognitive synonyms (or synsets); *syntax-oriented* approach is used through Apache Lucent's text searching engine; and *semantic parse* approach is used through SNOMED clinical terminology system, as well as through categorising the HL7 v3 interactions using our proposed transaction schema in Section 5.2.

Rasmussen and Bassøe (1993) describe a program for automatic semantic analysis of clinical narratives. In this approach the diagnoses are written in a free-text format during consultation with the patient and later they are collected into diagnostic classes. A lexical parser then automatically creates dictionaries from the clinical narrative associated with each disease. Via fuzzy set operations the correlations between diseases and corresponding signs, symptoms and treatments are identified and disease-specific knowledge is extracted. In a similar approach, Yousefi et al. (2009) use concept lattice analysis technique to detect the semantic relations among a patient's signs, symptoms and EHR records, with those of different diseases. This allows a decision support system to narrow down the set of candidate diagnoses to a short list for the clinician.

Pakhomov et al. (2010) developed a framework for practical and theoretical issues with creating reference standards for semantic relatedness. Currently, research on computerised approaches to semantic relatedness between biomedical concepts relies on reference standards created for specific purposes using a variety of methods for their analysis.

Bousquet et al. (2008) use two terminology systems in pharmacovigilance for coding of adverse drug reactions and statistical analysis. The available tools for automated signal detection and access to pharmacovigilance databases would benefit from terminological reasoning in order to provide improved groupings of terms describing the same medical condition. Such reasoning depends on formal definitions that are absent in both terminologies. They propose a draft for an ontological model consisting of 19 semantic categories and 16 relations for the representation of adverse drug reactions. From this model, they selected eight semantic categories for the categorical structure.

Nagy et al. (2010) describe how semantic interoperability among contemporary EHR-Ss with support of the HL7 v3 messages and concept mapping standards could improve patient safety. A notation-wide implementation of a semantic interoperability platform would include adopting and translating international coding systems and nomenclatures, and developing implementation guidelines to facilitate the migration from national standards to international ones. Such semantic interoperability to preserve message meaning is very challenging.

Pesquita et al. (2009) reviewed different semantic similarity measures applied to biomedical ontologies and propose their classification according to different strategies. Semantic similarity is used for validating the results of biomedical studies such as gene clustering, gene expression data analysis, prediction and validation of molecular interactions, and disease gene prioritisation.

Roth-Berghofer and Forcher (2011) describe how the results of a semantic search engine can be more understandable by adding an explanation facility for justifying and exploring the search result. This mechanism has been integrated with their search engine and is used for semantic understandability of medical documents.

Reichle et al. (2011) propose an application-independent architecture that features knowledge acquisition from a web-community, knowledge modularisation, and agent-based knowledge maintenance. They introduce a travel medicine as application domain, which applies the agent-based system architecture for extracting, analysing, sharing and providing community experiences in an individualised way.

While the above approaches deal with semantic level of biomedical data, none discuss the semantic preservation during the translation of the medical narratives (storyboards) to identify relevant context and HL7 v3 interactions. However, the approaches by Bousquet et al. and Rasmussen and Bassøe are the closest to ours.

4 Healthcare standards and technologies

This section introduces healthcare standards and various technologies that are used in our approach.

4.1 Health Level 7

HL7 is a non-profit organisation 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. The term HL7 also refers to the standards created by the HL7 organisation.

HL7 version 2.1, originally created to support hospital workflow was improved at version 2.6 to realise interoperability between electronic patient administration systems (PAS), electronic practice management (EPM) systems, laboratory information systems (LIS), dietary, pharmacy and billing systems and electronic medical record (EMR) systems. However, this standard did not adhere consistently to a data model and was text-based as opposed to XML-based. HL7 v3 was envisioned and designed to overcome these limitations.

HL7 v3 comprises a pair of base specifications – an object-oriented information model called the RIM and a set of vocabulary domains. RIM and its derivatives describe structure of data in terms of classes, attributes, constraints and relationships whereas the vocabulary domains encapsulate domain concepts and terms. HL7 message refinement process describes how message types are derived from core RIM classes.

The strategy for development of HL7 v3 messages and related information structures is based upon the consistent application of constraints on these two base specifications. Upon the extension of the specifications, the created constrained representations address a specific healthcare requirement.

4.2 HL7 v3 interactions

HL7 defines interactions as a unique association between a specific *message type* (information transfer), a particular *trigger event* (initiating or trigging the transfer) and the *receiver responsibilities* (response interactions associated with the receipt of the interaction). Thus, interactions provide critical contextual information required by a recipient to interpret the semantics of a message and to trigger an appropriate response.

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. 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 healthcare transaction must be mapped to the correct set of interactions. Therefore, interactions form the heart of the proposed process for HL7 message extraction.

4.3 Systematised Nomenclature of Medicine – Clinical Terms

Systematised Nomenclature of Medicine – Clinical Terms (SNOMED CT) 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 recommended by HL7 organisation as a terminology standard for clinical data exchange.

4.4 Logical Observation Identifiers Names and Codes

Logical Observation Identifiers Names and Codes (LOINC) is a database of codes representing terms used primarily in the laboratory and observation areas of healthcare. LOINC (http://www.loinc.org) was initiated in 1994 as a voluntary effort to meet the demand for electronic movement of clinical data from laboratories that produce the data to hospitals and physician's offices. LOINC has been identified by the HL7 Standards Development Organization as a preferred code set for laboratory test names in transactions between healthcare facilities, laboratories, laboratory testing devices, and public health authorities.

4.5 Resource description framework

Resource description framework (RDF) (Fensel, 2000; W3C RDF Primer, http://www.w3.org/TR/2004/REC-rdf-primer-20040210/) 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 labelled arc is the predicate (also called property), and the 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 artefacts. SW technologies such as RDF offer a rich platform to implement efficient and accurate semantic search capabilities. Efforts are underway to produce RDF-enable object-oriented modelling tools such as UML (Stanford Infolab, http://infolab.stanford.edu/melnik/rdf/uml/) to allow these tools to be integrated with other UML-based tools in the application design phase of the integration projects.

In recent years a number of SW languages such as Web Ontology Language (OWL) (W3C, http://www.w3.org/2004/OWL/), Ontology Inference Layer (OIL) (Fensel, 2000) and DARPA Agent Markup Language (DAML) (Arroyo et al., 2004) 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.

4.6 Sesame framework

Sesame (OpenRDF.org, http://www.openrdf.org/index.jsp) 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. The library can be deployed as a Java Servlet application on Apache Tomcat server. The repository can be in-memory or a relational database such as MySQL. Sesame supports an advanced inference and query language Sesame Query Language (SeRQL) (OpenRDF.org, http://www.openrdf.org/doc/sesame/users/ch06.html) to query and find implicit information in RDF schema and data.

5 Proposed approach

In this section, we describe our approach that simplifies translation of healthcare transactions to HL7 v3 interactions with the use of a novel tool. The following subsections provide the details of the proposed process and underlying concepts.

5.1 Extracting HL7 v3 metadata

The developed tool aids the system integrators to map healthcare transactions with HL7 v3 interactions most appropriate to communicate their content and context. For this purpose, specific relationships between real-world healthcare transactions and interactions are needed to establish and built into the tool's mapping logic. However, the relationship between transactions and interactions are not explicit or obvious in the HL7 v3 specification. Also, 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 is not possible. Therefore, our approach for construction of the search tool is to discover important metadata in HL7 v3 interactions that can also be used to describe a healthcare transaction. The default metadata associated with HL7 interactions are the D-MIMs (domains) and R-MIMs (sub domains) that they belong to. However, these pieces of information alone would not be sufficient to act as metadata for a search tool. Also, as observed in the introduction to this paper, there are inconsistencies among information hierarchies of different domains. Based on extensive study of HL7 v3 information models and the obtained knowledge, we developed the following metadata to drive the search tool.

Table 1 A portion of the proposed HL7 v3 contexts, descriptions and associated D-MIMs

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
Specialised care and professional services	Care provision (REPC DM000000UV)	Specialists, physiotherapy, psychology, counselling
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
Referral report	Care provision (REPC DM000000UV)	Referral report
Claims and reimbursements – special authorisation	Claims and reimbursements (FICR DM000001UV)	Insurance special authorisation
Claims and reimbursements – eligibility	Claims and reimbursements (FICR DM000001UV)	Insurance eligibility
Claims and reimbursements – pre-approval	Claims and reimbursements (FICR DM000001UV)	Insurance pre-approval
Claims and reimbursements – pre-determination	Claims and reimbursements (FICR DM000001UV)	Insurance pre-determination
Claims and reimbursements – coverage extension	Claims and reimbursements (FICR DM000001UV)	Insurance coverage extension
Invoice	Claims and reimbursements (FICR DM000001UV)	Invoice
Payment notice	Claims and reimbursements (FICR DM000001UV)	Payment notice
Statement of financial activity	Claims and reimbursements (FICR DM000001UV)	Financial statement
Immunisation	Immunisation (POIZ DM000000UV)	Vaccination, substance administration, immunisation

Table 1 A portion of the proposed HL7 v3 contexts, descriptions and associated D-MIMs (continued)

Context	*HL7 domain and D-MIM*	*Description*
Laboratory	Laboratory (POLB DM000000UV)	Laboratory, diagnostics, pathology, results, specimen, laboratory report
Drug knowledge-base	Medication (POME DM000000UV)	Drug information, drug document
Inventory management	Material management (PRMM DM000001UV)	Inventory, material management
Consent to share medical record	Medical record (RCMR DM000050UV)	Patient consent
Electronic medical record	Medical record (RCMR DM000050UV)	Electronic medical record
Non-laboratory observation	Observation (POOB DM200000UV)	Vital signs, vitals, observation
Order health services	Order (POOR DM100000UV)	Order services
Patient registry	Patient administration (PRPA DM000000UV)	Register, patient account, person account, create
Person registry	Patient administration (PRPA DM000000UV)	Register person
Location registry	Patient administration (PRPA DM000000UV)	Register location
Encounter (in patient)	Patient administration (PRPA DM000000UV)	Hospital admission, in-patient encounter
Encounter (ambulatory)	Patient administration (PRPA DM000000UV)	Ambulatory encounter, out-patient encounter
Encounter (ER)	Patient administration (PRPA DM000000UV)	ER, emergency
Encounter (home health)	Patient administration (PRPA DM000000UV)	Home health encounter
Encounter (general)	Patient administration (PRPA DM000000UV)	Encounter
Human resources	Personnel management (PRPM DM000000UV)	Healthcare workers, human resources

5.1.1 Interaction context

Using a holistic view of HL7 information model, we reclassified the domains and sub-domains of original HL7 v3 model in a more intuitive and precise manner. We grouped those domains that are conceptually related in an intuitive way and separated those domains that grouped together seemingly unrelated areas. In this study, we termed the new set of domains thus derived 'contexts' to avoid confusion with original HL7 domains. We have developed 50 such contexts to represent different areas of healthcare. To verify that the new contexts superimpose well with healthcare transactions, we conducted a large number of exercises of associating the new contexts with healthcare domains found in storyboards in HL7 literature. Once the contexts were finalised, each interaction was associated with a single context. Context acts as a key piece of metadata in the search tool. Table 1 presents a list of 36 (out of 50) contexts along with their associated D-MIMs.

5.1.2 Interaction classification model

Each HL7 interaction is designed to convey a specific set of data (payload) and some contextual information. The concept of contexts described above captures metadata about the actual data that the payload portion of the interaction conveys. The contextual information contained in the control act wrapper portion of the interaction describes the action that the message triggers or dictates at the recipient.

Therefore, we have classified interactions into a hierarchy of classes based on the action dictated by them. The class model is exhaustive and represents all possible actions dictated by interactions specified in the HL7 v3 information model. We call this classification *interaction classification model*. The classes in the model and their descriptions are given in Table 2. The class of an interaction is the next key piece of metadata that would drive our tool. The interaction classification model hierarchy consists of three levels of sub-classes, as follows:

- *Level 1:* An interaction is sub classed into *initiator* and *response*. *Initiator* class represents interactions that initiate an information exchange. *Response* class represents interactions that are non-initiators and are sent by a receiver in response to a previous message.
- *Level 2:* Initiator interaction can further be classified into *query, command* and *notification*. *Query* represents requests for information. *Command* refers to an interaction ordering the receiver to perform a task. *Notification* refers to interactions that notify a third party of occurrence of an event. *Response* class is divided into acknowledgement and InformationR. *Acknowledgements* are interactions that are sent by the receiver to inform the status of a prior interaction. *InformationR* represent query results or information sent as response to a command requesting data.
- *Level 3:* Command and notification are classified into 18 sub-categories based on the nature of the requested task. *Abort, activate, update, retract* and *record* are some examples. Level 2 type query is sub-categorised into *summary* and *detail* based on level of detail in information requested. *Acknowledgement* is sub-divided into *received, accepted* and *rejected*, representing the status of the message. InformationR is further sub-divided into *SummaryR* and *DetailR* based on the level of detail.

Healthcare transactions as defined in the next section convey data and trigger certain actions on the part of the recipient. The action conveyed by real world healthcare transactions can also be classified as per the interaction classification model we have developed. Therefore, they can be used to relate interactions to transactions as described in Section 5.2.

HL7 v3 information model has specifications for over 900 interactions in its 2009 January ballot (Health Level 7, http://www.hl7.org). In this study, we have categorised over 600 of these interactions according to the concepts described above.

Table 2 Definition of some of the classes in the interaction classification model

Class		*Definition*
Initiator	(Level 1)	Interaction initiating a conversation with a receiving system.
Query	(Level 2)	Query receiver for information.
Detail		Find all possible candidates matching search criteria.
Summary		Retrieve a particular record by ID.
Command	(Level 2)	Order the receiving system to perform an action.
Abort		Order receiving system to abort a previously activated operation.
Activate		Order receiving system to activate an account.
Authorise		Order receiving system to authorise an operation/document.
Cancel		Order receiving system to cancel a previously activated operation.
Complete		Order receiving system to complete a previously activated operation.
Create		Order receiving system to create a record.
…		
Notification	(Level 2)	Notify receiver(s) of occurrence of an event or action.
Abort		Notify receiving systems of an abort operation.
Activate		Notify receiving systems of an activate operation.
Authorise		Notify receiving systems of an authorise operation.
Cancel		Notify receiving systems of a cancel operation.
Complete		Notify receiving systems of a complete operation.
Create		Notify receiving systems of a create operation.
Delete		Notify receiving systems of a delete operation.
Information		Notify receiving systems of information asynchronously.
…		
Response	(Level 1)	Respond to a command, query or notification.
Acknowledgement	(Level 2)	Acknowledge the receipt of a message indicate if 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	(Level 2)	Response to a command to send information/query.
Summary		Summary information response.
Detail		Detailed information response.

5.2 Structured healthcare transactions

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 (2008) for structuring business scenarios for automation. Each participating message in a transaction conveys some information required to complete the overall goal of the transaction. 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' represents the action information contained in the message description. For example, in message 'EMR requests EHR for patient allergies', *requests* becomes the operation component.

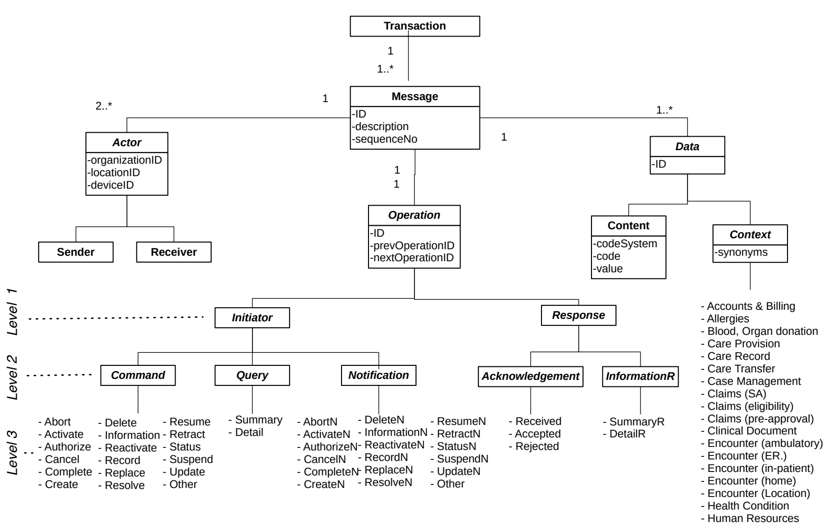
We collectively call the remaining information in the message description as 'data'. Data comprises of content and context components. Content refers to fields of data that need to be communicated to the receiver. Context describes the domain affiliation of the message itself.

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:

$$Transaction : \{Message\}^{1..N}$$

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

Figure 1 Healthcare transaction schema



Note: The schema describes a healthcare transaction as a composition of actors, operation and data.

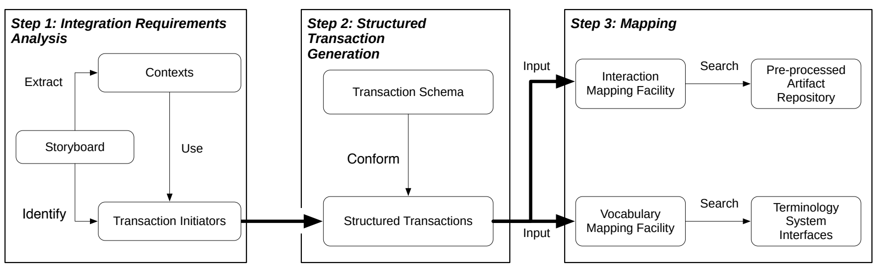
We use the concepts of contexts and interaction classes described above to represent constituent components of healthcare transactions. We derive the operation component of a transaction from the interaction class hierarchy. Also, the context component of a transaction is expressed as an item from the list of interaction contexts. The complete schema for a healthcare transaction is illustrated in Figure 1.

Now, it is possible to incorporate the above concepts into a search tool that will map healthcare transactions to interactions. The technical concepts behind the tool are described in detail in Section 6.

5.3 Proposed process

We propose a three-step process, illustrated in Figure 2, to guide users through the activities involved in identifying candidate transactions and translating them to HL7 v3 interactions. We use a running example to demonstrate our approach.

Figure 2 Proposed process for translating healthcare transactions to HL7 v3 interactions



5.3.1 Step 1: integration requirements analysis

This step involves examining information exchange requirements of the systems being integrated. Typically, a business analyst would document system requirements by conducting joint discovery sessions with the end users of the system or systems to be integrated. We streamline activities involved in integration requirement Analysis as follows.

5.3.1.1 Step 1.1: storyboarding

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 and the following scenario, 'visit to physician to refill prescription', is used as our running example.

> "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 Dr. P his Family Physician (FP). 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 allergies. 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."

As seen in the above example, information in storyboards is often incomplete, unstructured and therefore, of little use for automation. While the completeness and accuracy of the storyboards depend on human factors and hence beyond our control, we propose the following activities to impose structure on the information in storyboards.

5.3.1.2 Step 1.2: extract contexts

The proposed mapping 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 and related terms describing it.

Using an online SNOMED browser (Jdet.com, http://www.jdet.com/) we searched and found related medical terms and phrases to the contexts we identified in Section 5.1. Further we used WordNet (http://wordnet.princeton.edu/), which is a large lexical database of English. In WordNet nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets). Each synset expresses a distinct concept and synsets are interlinked by means of conceptual-semantic and lexical relations. Using the WordNet browser, we searched its network of meaningfully related words and concepts to identify cognitive synonyms to our contexts and context descriptions. These cognitive synonyms were input into the mapping tool's database.

Storyboard text entered by users would be mapped to related contexts using the mapping tool. The tool would search the database of cognitive synonyms, medical terms and phrases that we created for matches and contexts associated with matching terms and phrases are presented to the user for manual refinement. We used API provided by Apache Lucene (http://lucene.apache.org/java/docs/), which is a high performance text search engine library to implement the search.

As part of future research, we intend to enhance this feature by using natural language processing (NLP) concepts. This exercise is useful to successfully perform Step 1.3, where users identify transactions that are conceptually linked to existing HL7 domains. The automatic mapping however, is not a definitive map and can be refined or replaced by the user manually.

For the storyboard in the running example, some possible context maps are:

1. medication: *pharmacy*
2. allergies: *allergies*
3. prescriptions: *pharmacy*.

5.3.1.3 Step 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 v3 contexts.

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.

5.3.2 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 Step 1.3 can be expressed in structured format as follows:

1 EMR sends request for patient medication history
   - *actor:* EMR
   - *action:* QueryDetail
   - *context:* pharmacy – patient medication record
   - *content:* medication history

2 EMR requests for patient allergies
   - *actor:* EMR
   - *action:* QueryDetail
   - *context:* allergy
   - *content:* patient allergies

3 EMR updates EHR with patient medication
   - *actor:* EMR
   - *action:* CommandUpdate
   - *context:* pharmacy – patient medication record
   - *content:* patient medication

4 EMR updates EHR with allergies
   - *actor:* EMR
   - *action:* CommandUpdate
   - *context:* allergy
   - *content:* adverse reaction

5 EMR sends prescription request to pharmacy to dispense

- *actor:* EMR
- *action:* CommandOther
- *context:* pharmacy – dispensing
- *content:* prescription.

5.3.3 Step 3: mapping

5.3.3.1 Step 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 artefact 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 provides interactions returned in the mapping step.

Table 3 Storyboard *medication refill* – transaction initiators and corresponding interactions mapped to them

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 fulfilment request*	(PORX IN011070UV)

5.3.3.2 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.

6 Case study – emergency encounter

In this section, a case study will be presented as the second example of applying the proposed process and tool to extract HL7 v3 interactions from a healthcare storyboard.

> "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 (Emergency), 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 HIS which automatically quizzes the EHR. The 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 Mr. S' phone number and address. System was updated and notification sent to EHR. The ER specialist, Dr. E decided that after a nebulizer treatment Mr. X was stable and was ready to be checked-out. Dr. E noted that Mr. X needed to schedule a follow-up visit with Dr. P, pulmonologist. The ER clerk completed the check-out information for Mr. X and checked him out of the Emergency Room. The 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
- 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
  1. HIS requests EHR for health record
     a. *actor:* HIS
     b. *action:* QueryDetail
     c. *context:* health condition
     d. *content:* health record
  2. HIS requests EHR to update demographic information
     a. *actor:* HIS
     b. *action:* CommandUpdate
     c. *context:* patient administration
     d. *content:* demographic information
  3. HIS sends emergency record to X's primary care physician
     a. *actor:* HIS
     b. *action:* NotificationInformation
     c. *context:* emergency encounter
     d. *content:* emergency record
  4. HIS sends emergency record to EHR
     a. *actor:* HIS
     b. *action:* NotificationInformation

c *context:* emergency encounter

d *content:* emergency record.

Table 4 presents the transaction initiators and the mapping interactions.

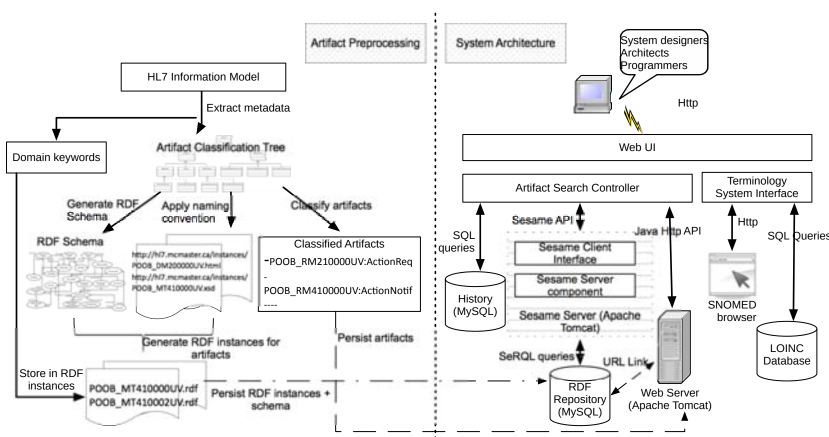
Table 4 Storyboard *emergency encounter* – transaction initiators and corresponding interactions best suited to represent them

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)

7 Developed tool

We developed an open source and web-based tool namely tool-assisted message mapping process (TAMMP). The architecture of the TAMMP tool is illustrated in Figure 3 and described in the next section.

Figure 3 Architecture of the developed tool (TAMMP), (a) steps for pre-processing HL7 artefacts and persisting in the artefact repositories (b) system architecture (see online version for colours)



(a) (b)

7.1 Architecture of TAMMP

The right portion of Figure 3 illustrates the high level architecture of the TAMMP tool. It is a client-server, multi-layer application which is composed of the following components: web user interface; artefact search controller; terminology system interface; Sesame interface; Lucene interface; and repository layer (MySQL database).

The web user interface is Java Servlet and provides a user friendly environment for searching, browsing, navigating and exploring artefacts. User interface calls upon the search controller component which interacts with various external interfaces such as Sesame API, Lucene API and the terminology system interface to leverage their services. The RDF repository access layer comprises of the Sesame server application. It interfaces with the RDF repository layer and handles connections and communications with the RDF repository to execute search and retrieval of RDF instances.

The user interface, search controller and Sesame server are hosted on an Apache Tomcat 6.0 web server. The web server also hosts a website of HL7 v3 artefacts such as XML schema, documentation, XML sample instances, information models and other representations for retrieval. The repository layer consists of MySQL databases for storing contexts and synonyms and RDF instances.

7.2 Web user interface

The Java Servlet-based web user interface guides the user through steps of TAMMP. It comprises of servlets *storyboard*, *StructuredTransactions* and *SearchResults*. The storyboard allows users to input a text-based healthcare scenario. It then leverages the services of the artefact search controller component to generate possible maps between keywords in the entered storyboard text and HL7 contexts in the tool's repository. *StructuredTransactions* servlet provides a user interface for decomposing transaction initiators into their constituent components. It then calls upon the underlying search controller to perform a semantic search to retrieve matching HL7 interactions. SearchResults servlet displays the results produced by the search operation.

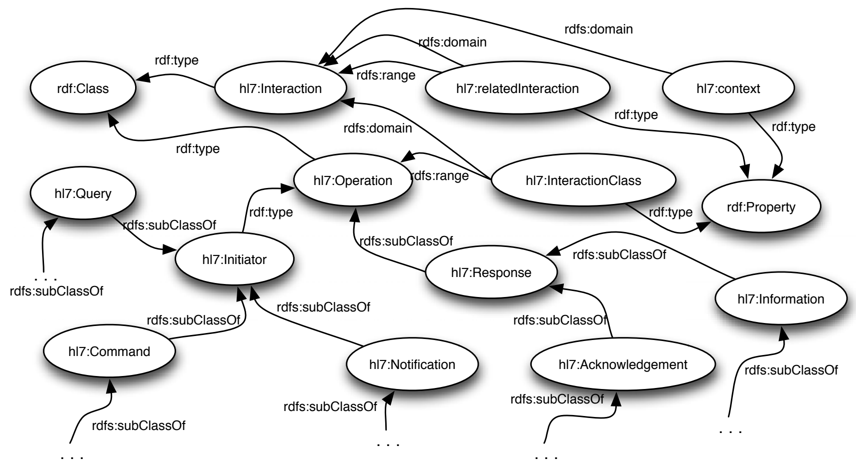
7.3 Search controller

The artefact search controller component comprises of *ContextSynonyms*, ContextMapper and *RepositoryConnector* classes. *ContextSynonyms* class retrieves the set of contexts and synonyms in the database and retains them for later use. Hence, it has been designed as a singleton pattern class. *ContextMapper* is responsible for mapping storyboard text to context synonyms leveraging indexed search features provided by Apache Lucene (http://lucene.apache.org/java/docs/), which is a popular open source full-text search engine. Hence, it interfaces with Lucene's *HttpRepository* API. The *RepositoryConnector* is responsible for connecting to the Sesame HL7 Store, constructing SeRQL (OpenRDF. org, http://www.openrdf.org/doc/sesame/users/ch06.html) queries and invoking Sesame API to search for RDF instances in Sesame repository. These RDF instances contain valuable metadata that help create a link between transactions and HL7 interactions.

7.4 Terminology system interface

The search controller also invokes terminology system interfaces to search for LOINC and SNOMED codes for terms used in the storyboards. The terminology system interface supports searching for SNOMED and LOINC codes for local terms by integrating into existing SNOMED browser by BT (Jdet.com, http://www.jdet.com/) and a MySQL database of LOINC codes.

Figure 4 RDF graph for HL7 artefact metadata model



7.5 Repository layer

The tool maintains a MySQL database with HL7 contexts and their cognitive synonyms. The synonyms are based on WordNet (http://wordnet.princeton.edu/), a lexical database for the English language maintained by the Princeton University. A separate MySQL repository has been created to store RDF instances carrying metadata on HL7 v3 interactions. HTML documentation of actual interactions are maintained as a separate website on the Apache Tomcat server.

7.6 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 artefacts related to a particular interaction, the 'operation' class that best represents it, 'context' it belongs to and cognitive synonyms describing the interaction's context (keywords). Figure 3(a) details the activities involved in offline artefact pre-processing stage.

An RDF schema has been generated to describe associations of an interaction by applying rules of RDF syntax and semantics specified by W3C. Since RDF requires all resources to be uniquely identifiable, we adopted an artefact naming convention based on their HL7 artefact ID which is unique. For example, observation request message schema will be named POOB MT210000UV.xsd based on its HL7 artefact ID POOB MT210000UV. Finally, RDF instances describing the metadata and relationships of each artefact is generated in conformance with the schema and by analysing the HL7 information models. Artefacts are persisted in the web server and RDF instances are stored in the RDF repository for access by the application.

The tool is used in the mapping step of the proposed process described in the previous step. At runtime, the user inputs the storyboard describing a healthcare scenario. The tool then executes a text search on the text of the storyboard to create maps between words and phrases in the text and HL7 v3 contexts adopted by this study. A MySQL database of HL7 v3 contexts and their cognitive synonyms is maintained to drive the search. We use Lucene search engine to execute a free-text search on the text in the storyboard. Since the text mapping is based on cognitive synonyms we have gathered from WordNet, they would only be approximate matches at this time. The purpose of this exercise is to provide new users with a starting point for selecting contexts. The user can confirm these maps or decide to select contexts manually.

In the next step, the user is prompted to enter the structured transactions identified from the storyboard. The search controller then generates SeRQL queries based on the search criteria in the structured transactions 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 displayed in a browseable format. Figure 4 illustrates RDF graph of HL7 v3 interaction metadata model.

A section of the RDF instance for interaction 'request to record subject observation' (artefact ID POOB IN000001UV) that is persisted in the RDF store is as follows:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-
syntax-ns#"xmlns:hl7="http://localhost:8080/hl7/schema#">
    <rdf:Description rdf:about="http://localhost:8080/
        hl7/interactions/POOB IN000001UV">
        <rdf:type rdf:resource="http://localhost:8080/
            hl7/schema#Interaction"/>
        <hl7:interactionClass rdf:resource=
            "http://localhost:8080/hl7/schema#Record"/>
        <hl7:context>OBS</hl7:context>
    </rdf:Description>
</rdf:RDF>
```

8 Conclusions

Increasingly, governments of many countries including Canada are recognising 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, well-defined approach to support message selection activity of communication design phase of HL7 v3 system integration projects. We presented a behaviour-based classification for HL7 v3 interactions that allows us to relate them to real life healthcare transactions via a novel search and mapping tool. We described the construction of this approach using SW technologies and we demonstrated its usage with the help of real life healthcare scenarios.

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 integration 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. During future research, we also intend to improve the context mapping mechanism of the tool to use NLP, which will add further value to the tool. Our current research has added the capability of user-assisted instantiating of HL7 v3 messages using message schemas editing and a developed application.

References

7Edit, HL7 website, available at http://www.7edit.com/home/index.php (accessed on 10 December 2011).

7Scan, HL7 browser website, available at http://www.7scan.com/ (accessed on 10 December 2011).

Apache Lucene, Full-text search engine website, available at http://lucene.apache.org/java/docs/ (accessed on 10 December 2011).

Arroyo, S., Lara, R., Ding, Y., Stollberg, M. and Fensel, D. (2004) 'Semantic web languages – strengths and weaknesses', *International Conference in Applied computing (IADIS04)*, pp.23–26.

Bousquet, C., Trombert, B., Kumar, A. and Rodrigues, J.M. (2008) 'Semantic categories and relations for modeling adverse drug reactions towards a categorial structure for pharmacovigilance', *Proceedings of the AMIA Annual Symposium*, November, Vol. 6, pp.61–65.

Canada Health Infoway, Canada Health Infoway website, available at http://www.infoway-inforoute.ca/lang-en (accessed on 15 August 2010).

Chaudhry, B. et al. (2006) 'Systematic review: impact of health information technology on quality, efficiency, and costs of medical care', *Annals of Internal Medicine*, May, Vol. 144, No. 10, pp.742–752.

Congressional Budget Office, Congressional Budget Office cost estimate, available at http://www.cbo.gov/ftpdocs/99xx/doc9968/hr1.pdf (accessed on 15 August 2010).

Department of Health and Human Services, website, available at http://healthit.hhs.gov/ (accessed on 15 August 2010).

Dezhkam, N. and Sartipi, K. (2008) 'Knowledge transformation from task scenarios to view-based design diagrams', *International Conference on Software Engineering and Knowledge Engineering (SEKE 2008)*, July, San Francisco Bay, USA, pp.26–32.

eHealth Ontario, available at http://www.ehealthontario.on.ca (accessed on 10 December 2011).

Fensel, D. (2000) 'The Semantic Web and its languages', *Journal of IEEE Intelligent Systems*, November/December, Vol. 15, No. 6, pp.67–73.

Health Level 7, website, http://www.hl7.org (accessed on 10 December 2011).

Healthcare Information and Management Systems Society (HIMSS) (2008) 'Electronic health records: a global perspective', Technical report, HIMSS.

HIMSS, webpage with definition of EHR, available at http://www.himss.org/ASP/topicsehr.asp/ (accessed on 10 December 2011).

Infoway, Infoway EHR Infostructure, available at https://www.infoway-inforoute.ca/lang-en/about-infoway/approach/investment-programs/infostructure (accessed on 10 December 2011).

Jdet.com, SNOMED browser, available at http://www.jdet.com/ (accessed on 20 September 2008).

Liu, B., Li, X., Liu, Z., Yuan, Q. and Yin, X. (2008a) 'Design and implementation of information exchange between HIS and PACS based on HL7 standard', *International Conference on Information Technology and Applications in Biomedicine (ITAB 2008)*, May, pp.552–555.

Liu, B., Li, X., Liu, Z., Yuan, Q. and Yin, X. (2008b) 'Experiences with mirth: an open source health care integration engine', *International Conference on Software Engineering (ICSE'08)*, May, Germany, pp.649–652.

LOINC, website, available at http://www.loinc.org (accessed on 10 December 2011).

Massachusetts Technology Collaborative (2003) 'Advanced technologies to lower the cost of health care and improve quality', Massachusetts Technology Park Corporation.

Nagy, M., Preckova, P., Seidl, L. and Zvarova, J. (2010) 'Challenges of interoperability using HL7 v3 in Czech healthcare', *Book: Studies in Health Technology and Informatics*, Vol. 155, pp.122–128, eBook, IOS Press, available at http://booksonline.iospress.nl/Content/View.aspx?piid=16851.

NeoTool HL7 Analyzer, website, available at http://www.corepointhealth.com/products/hl7-analyzer/hl7-analyzer (accessed on 10 December 2011).

openEHR, website, available at http://www.openehr.org/home.html/ (accessed on 10 December 2011).

OpenRDF.org, 'Sesame RDF Query Language', available at http://www.openrdf.org/doc/sesame/users/ch06.html (accessed on 10 December 2011).

OpenRDF.org, Sesame RDF framework website, available at http://www.openrdf.org/index.jsp (accessed on 10 December 2011).

Pakhomov, S.V., Pedersen, T., McInnes, B., Melton, G.B., Ruggieri, A. and Chute, C.G. (2010) 'Towards a framework for developing semantic relatedness reference standards', *Journal of Biomedical Informatics*, Vol. 44, No. 2, pp.251–265.

Pesquita, C., Falcao, A.O., Lord, P. and Couto, F.M. (2009) 'Semantic similarity in biomedical ontologies', *Journal of PLoS Computational Biology*, 31 July, Vol. 5, No. 7, p.e1000443, ePublish.

Rasmussen, J.E. and Bassøe, C.F. (1993) 'Semantic analysis of medical records', *Journal of Methods Inf. Med.*, February, Vol. 32, No. 1, pp.66–72.

Reichle, M., Bach, K. and Althoff, K.D. (2011) 'Knowledge engineering within the application-independent architecture SEASALT', *International Journal of Knowledge Engineering and Data Mining (IJKEDM)*, Vol. 1, No. 3, pp.202–215.

Rong Chen, G.E. and Klein, G.O. (2007) 'Julius a template based supplementary electronic health record system', *BMC Medical Informatics and Decision Making*, online publication, May, available at http://www.biomedcentral.com/1472-6947/7/10 (accessed on 10 December 2011).

Rong Chen, G.K. (2007) 'The open EHR java reference implementation project', *MEDINFO 2007*, IOS Press, pp.58–62.

Roth-Berghofer, T. and Forcher, B. (2011) 'Improving understandability of semantic search explanations', *International Journal of Knowledge Engineering and Data Mining (IJKEDM)*, Vol. 1, No. 3, pp.216–234.

SNOMED CT, website, available at http://www.ihtsdo.org/snomed-ct (accessed on 10 December 2011).

Southern California Evidence-based Practice Center (2006) 'Costs and benefits of health information technology', Technical report, Agency for Healthcare Research and Quality.

Stanford Infolab, 'Representing UML in RDF', available at http://infolab.stanford.edu/~melnik/rdf/uml/ (accessed on 10 December 2011).

Vishwanath, A. and Scamurra, S. (2007) 'Barriers to the adoption of electronic health records: using concept mapping to develop a comprehensive empirical model', *Health Informatics Journal*, Vol. 13, No 2, pp.119–134.

W3C RDF Primer, available at http://www.w3.org/TR/2004/REC-rdf-primer-20040210/.

W3C, 'Web Ontology Language', available at http://www.w3.org/2004/OWL/ (accessed on 10 December 2011).

WordNet, website, available at http://wordnet.princeton.edu/ (accessed on 10 December 2011).

Yang, T.H. (2006) 'A scalable multi-tier architecture for the National Taiwan University hospital information system based on HL7 standard', *International Symposium on Computer-Based Medical Systems (CBMS'06)*, Germany, pp.99–104.

Yarmand, M. and Sartipi, K. (2008) 'Standard-based data and service interoperability in eHealth systems', *International Conference on Software Maintenance (ICSM'08)*, pp.87–196.

Yousefi, A., Mastouri, N. and Sartipi, K. (2009) 'Scenario-oriented information extraction from electronic health records', *IEEE International Symposium on Computer-Based Medical Systems (CBMS'09)*, pp.1–5.