TOOL-ASSISTED KNOWLEDGE TO HL7 V3 MESSAGE TRANSLATION

TOOL-ASSISTED KNOWLEDGE TO HL7 V3 MESSAGE TRANSLATION

By
PRIYA JAYARATNA, B.SC.

A Thesis
Submitted to the School of Graduate Studies
in partial fulfilment of the requirements for the degree of

Master of Applied Science
Department of Computing and Software
McMaster University

M.A.Sc. McMaster University, Hamilton, Ontario

TITLE: Tool-assisted knowledge to HL7 v3 message translation

AUTHOR: Priya Jayaratna, B.Sc. (University of Moratuwa, Sri Lanka)

SUPERVISOR: Dr. Kamran Sartipi

NUMBER OF PAGES: vi, 63

Abstract

Healthcare System Integration is an area of utmost importance in the overall eHealth strategy of Ontarios provincial government as well as the federal government of Canada. A large body of researchers from various governmental and non-governmental organizations are actively engaged in delivering solutions to integrate disparate healthcare information systems. The overall goal of these efforts is to provide a province-wide and nation-wide 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.

While the need for health information integration is clear, due to inherent complexities of the healthcare domain as well as health information standards such as Health Level 7 (HL7), completion of such projects within budget and time is not an easy task. The goal of this study is to understand and analyze the information architecture behind HL7 version 3 (HL7 v3) with the aim of simplifying healthcare system integration process. In this thesis, we present a novel framework for extracting HL7 v3 messages to represent healthcare transactions that take place in an integration scenario. We have developed a prototype tool based on semantic web (SW) technologies to support our approach. We also present three healthcare case studies to demonstrate our solution.

Acknowledgements

I extend my heartfelt gratitude and appreciation to the people who contributed towards the success of this study and the masters thesis. I would like to express my sincere thanks to my thesis supervisor, Dr. Kamran Sartipi, for providing all necessary guidance throughout my research and preparation of this thesis. I am greatly indebted to Dr. Kamran Sartipi for his immense help in planning and executing this work in time. My sincere thanks are due to the official referees, Dr. Poehlman (examiner) and Dr. Anand (Chair), for accepting to review this thesis. Finally, I would like to thank all those who supported me in conducting my research and compiling the thesis.

Contents

List of Tables

List of Figures

Chapter 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 [16, 31, 13]. However, compared with other business domains such as banking, telecommunication and media, the IT spending and adoption rates of the healthcare industry has historically been lagging behind [20].

The slow pace of IT adoption in healthcare can be attributed to a variety of factors. Historically, there has been a perceived lack of trust amongst healthcare professionals towards information technology. Doubts about cost vs benefits 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 health care systems due to prohibitive inial costs of implementation, lack of conclusive evidence of return on investment and information security and privacy concerns [33, 25].

However, the landscape of healthcare information technology has slowly been changing over the past decade. With educational institutions taking the lead in providing IT education to health professionals, there's 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 organizations such as the Canada Health Infoway[14], Agency for Healthcare Research and Quality [31] and Healthcare Information and Management Society (HIMSS) [20] shows tangible benefits of using information technology in healthcare. The USA Congressional Budget Office estimates that the use of electronic medical records could save the nation $12.5 billion over 10 years [15]. As a result, harnessing IT for betterment of health services is becoming an integral part of the heatlhcare 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 too 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. Ontario's provincial government in 2008 formed a new agency named eHealth Ontario entrusted with driving the province's IT strategy and allocated $2.1 billion for years 2008 through 2012 [1].

There's a flurry of activity in the field of healthcare informatics to help translate the change in attitude and commitment to tangible business results. Healthcare stake holders at all levels, from patients to practitioners to health authorities demand a variety of services from healthcare IT infrastructure. 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 [25]. In summary, there's a growing need for unification of information across application and organization boundaries in a secure and reliable manner.

Software System Integration is one of the most important areas of Health Informatics as it provides a way of combining information from varied sources making it richer as well as sharing of information amongst various actors in a healthcare scenario. A single visit to the physician by a patient may trigger all or some of the following activities: sending off a prescription to the pharmacy; pharmacist checking prescription history on the EHR system; physician ordering a laboratory test; physician receiving results; physician scheduling an appointment with a specialist and the specialist scheduling a surgery. To complete each subsequent task successfully, information may need to flow from the previous task or tasks. For example, for the physician to electronically book a laboratory test and receive results back efficiently, the physician's system and the laboratory system must be integrated. Thus in a healthcare setting, many remote systems performing disparate tasks and handling different types of information must be able to talk to and interpret each other.

Healthcare systems currently in use by different organizations are as diverse as the healthcare domain itself. Information exchange between different organizations is still mostly non-electronic, largely depending on telephone, fax and mail. Even where systems have been integrated, they are mostly point-to-point integrations where typically, each system would support a variety of interfaces to enable sharing information with different systems. This is obviously a rather non-scaleable and maintenanceintensive approach since any change in one system's interface would warrant all externals interacting systems to change. Thus the the importance of widely accepted and adhered-to standards in healthcare is immense to realize efficient and scalable integration among healthcare systems.

Healthcare system integration includes activities spanning across many software engineering disciplines. These include: data modelling, communication modelling, security and privacy design, interface development and software and hardware infrastructure design. This study focusses 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 thesis and the proposed solution.

1.1 Problem definition

Achieving seamless integration amongst heterogeneous healthcare systems is not an easy task. Many challenges such as information security, personal health information privacy, legal and regulatory concerns, scale-ability, performance, robustness and availability of information need to be addressed. One of the major barriers in implementing nationwide integrated solutions such as the EHR is the problem of interoperability [25]. 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) [19] is the 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 have 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 in order to perform

[1] SNOMED - Systematized NOmenclature of MEDicine - Clinical Terms (SNOMED CT) [30] and LOINC - Logical Observation Identifiers Names and Codes (LOINC) [24].

message design tasks independantly. 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 organized into a hierarchy of information models from which messages are progressively derived. These information models are described in detail in Chapter 3 on Standards and Technologies. 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 of healthcare represented by a D-MIM3.1.1. A D-MIM may further be refined into "topics". 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 RIM3.1 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. Examples of such inconsistencies are:

- Activate, Revise and Nullify Interactions of the Patient Billing topic of Account and Billing domain derive from the same topic under HL7, and the hence same R-MIM (Patient Billing Account Event). However, interactions representing the same actions (i.e. Activate, Revise and Nullify), for "Person" topic under Patient Administration domain derive from three different R-MIMs.

- 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 the 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 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.

1.2 Proposed solution

In this study 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* 4.2. The process consists of three stages:

Integration Requirements Analysis; Structured Transaction Generation and Mapping. Chapter 4 of this thesis describes the proposed process in detail.

The Mapping step of the process uses a search tool built on leading edge Semantic Web (SW) technologies [18] to automate mapping of structured healthcare transactions onto HL7 v3 Interactions (3.1.2).

The following 4 steps describe the development and usage of the tool:

1. Through careful analysis of HL7 Interactions and domain model, identify searchable metadata

2. Express healthcare transactions in a structured, machine-readable format using the above identified metadata

3. Create an Interaction store and associated metadata

4. Use structured transactions to search for matching Interactions

The proposed process improves efficiency and accuracy of HL7 v3 based integration projects by aiding system developers with little knowledge of the standard to extract appropriate messages to meet communication requirements. Thus our approach helps to solve the problem defined in the previous section.

1.3 Thesis contributions

This thesis presents an approach to simplify the process of identifying HL7 v3 messages required to represent real world healthcare scenarios. The approach relies on a reliable 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 Nima and Sartipi [17]. The open-source, semantic-web based prototype search tool is built upon Sesame RDF framework. It provides advanced semantic search facilities for identifying and browsing HL7 artifacts suitable for representing an input *structured healthcare transaction.*

The contributions of this thesis to the healhcare informatics field can be presented as follows.

- Devised a novel, well-defined process to guide translation of healthcare transactions to HL7 v3 Interactions
- Developed a prototype tool based on semantic web technologies to automate the process of identifying HL7 Interactions appropriate to represent healthcare transactions
- Re-categorized HL7 v3 Interactions based on their behavioral traits in a messaging context. These categories provided valuable metadata to be used by the proposed search and mapping tool
- Extended an approach by Nima and Sartipi [17] for formal representation of business scenarios and adapted it to represent healthcare transactions. Structured transactions are formalized representations of business transactions and can be processed and mapped to HL7 v3 Interactions by the proposed tool.
- Demonstrated the use of the framework and the tool with three real world healthcare case studies

1.4 Thesis overview

The remaining chapters of this thesis are organized as below.

Chapter 2: describes an overview of the related work in the area of healthcare informatics.

Chapter 3: describes various healthcare standards and technologies applicable to this study.

Chapter 4: presents a discussion of the proposed process for healthcare transaction to HL7 message translation.

Chapter 5: provides a description of the semantic search and mapping tool and associated technologies.

Chapter 6: provides formal definitions of concepts used in the thesis.

Chapter 7: presents a conclusion to the thesis and discusses opportunities for future research.

Appendix: provides a list of acronyms used.

1.5 Limitations of the proposed approach

Limitations of scope: The proposed process is limited in scope to solving the problem of identifying HL7 v3 messages for representing a given healthcare scenario. Automated message generation and data population are out of the scope of this study and are considerations for future research projects.

Limitations of the technique: The proposed semantic search and mapping tool compares different components of the structured healthcare scenarios generated by the user to metadata about HL7 v3 Interactions that are stored in the tool's repositories. The accuracy of this mapping depends on how well the generated structured transactions represent original healthcare transactions. The process of generating structured transactions is largely manual though supported by guidelines in our process. Therefore end result will be subject to human judgement to some degree.

Chapter 2

Related work

2.1 HL7 V2 Tools

Due to relative simplicity of HL7 v2 data model and message format, the process of building tool-support is less complex and straightforward. There are a number of widely used commercial support tools available for HL7 version 2.

7Scan [11] is a specialized browser and editor that finds, displays, edits and transmits text-based HL7 version 2 messages with ease. 7Scan is an ideal tool to develop, test, and maintain HL7 interfaces. 7Scan also can be used as an endpoint simulator to send and receive messages with any HL7 interface being developing. 7Scan assists users to understand HL7 v2 messages by converting the coded, flat structured messages into hierarchical structures with user friendly field definitions.

7Edit [10] 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-segments[1] and message structures can be

[1]Z segments contain clinical or patient data that the HL7 Standard may not have defined in other

customized 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 Analyzer* [26] (formerly NeoBrowses)offers a multi-view interface simplifying how a programmer can view, edit, test, validate, and repair any HL7 message, increasing productivity. HL7 Analyzer provides users the ability to view and quickly understand any HL7 message, test conformance to an HL7- or user-defined message profile, simulate an HL7 message exchange to and from any healthcare application, test and debug healthcare applications TCP/IP interface and test and debug HL7 messages. Thus, out of the HL7 v2 tools we have analyzed, NeoTool's HL7 Analyzer offers the broadest spectrum of services to the user.

2.2 Research in eHealth

2.2.1 HL7 Based Integration Research

The HL7 v3 mapping process proposed in this paper is continuation of work carried out by Yarmand and Sartipi [35]. Their proposed model for message standardization is based on guidelines set forth by Canada Health Infoway[14]. 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.* [22] discuss an HL7 v2 based integration project to establish interoperability between a hospital information

areas. Z segments can be inserted in any message at any time, and Z segments can carry any data.

system (HIS) and a Picture Archiving 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 [23] is a far more advanced, full-fledged, open source healthcare messaging integration engine developed by WebReach, Inc., a health care 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, 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.

2.2.2 Electronic Health Record (EHR)

Health Information Management Systems Societys (HIMSS) 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. Included in this information are patient demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data, and radiology reports. 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 interfaceincluding evidence-based decision support, quality management, and outcomes reporting [3].

The latest and the greatest research and development efforts in electronic health is in the arena of integrated EHR. Currently Canada Health Infoway (Infoway) is spearheading projects to realize a Service Oriented Architecture (SOA) based, shared Electronic Health Record system in Canada leveraging HL7 v3. EHR Infostructure (EHRi) [21], 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. Infoway funds 276 projects across Canada, each of which is aimed at realizing a key piece of the overall EHR architecture. These projects include Registries, Diagnostic Imaging Systems, Drug Information Systems, Laboratory Information Systems, Interoperability projects and Infostructure projects.

openEHR [4] is an international not-for-profit Foundation, working towards making the interoperable, life-long electronic health record a reality and improving health care 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. Chen and Klein [28] describe implementation of a open source reference information model for openEHR project. openEHR has also designed and developed a template based EHR system called *Julius* that was integrated with existing EHR systems [27]. They claim that the use of templates in combination with user defined variables provides the clinical user groups with a set of tools to author variable definitions and templates, and to share the definitions of both, facilitating interoperability. It is also claimed to be a highly flexible and adaptive system since its runtime behaviour is driven by templates and variable definitions created and maintained by the end users, the medical professionals themselves. The Julius system has been implemented, tested and deployed to three health care units in Stockholm, Sweden with positive feedback from users.

2.2.3 Structured Scenarios

We have studied an approach proposed by Nima and Sartipi [17] 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, behavior 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's an increasing trend towards modernizing legacy healthcare IT infrustructure. Our research is concentrated on standard-based integration of legacy systems leveraging emerging technologies such as Web Services, SOA and ESB [22, 32]. Our mission is to contribute towards legacy system interoprability by providing guidelines, well-defined processes and tool-support to improve complexity, return on investment (ROI) and turnaround time of HL7 v3 standard-based integration projects.

Chapter 3

Healthcare standards and technologies

This chapter introduces healthcare standards and various technologies that are relevant to our research.

3.1 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. The term HL7 also refers to the standards created by the HL7 organization.

HL7 version 2.1, originally created to support hospital workflow was improved at version 2.6 to realize 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. Further, the standard lacked precision and consistency seriously limiting its scalability. Thus in the late 90's there was a clear demand for a cleaner, more precise standard. 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 *Reference Information Model (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 domain encapsulate domain concepts and terms. HL7 message refinement process describes how message types are derived from core RIM classes.

The strategy for development of version3 messages and related information structures is based upon the consistent application of constraints to a base specification called the HL7 Reference Information Model (RIM) and HL7 Vocabulary Domains and upon the extension of those specifications to create representations constrained to address a specific health care requirement. Constraints are applied on appearance, cardinality, type and vocabulary sets of base classes and attributes in a top down manner to recursively derive progressively specialized information structures.

3.1.1 HL7 v3 Information Model

The HL7 information modeling process recognizes three interrelated types of information models. Each of the model types uses the same notation and has the same basic structure. The models differ from each other based on their information content, scope, and intended use. The following types of information models are defined:

- *Reference Information Model (RIM):* The RIM is used to express the information content for the collective work of the HL7 Working Group. It is the information model that encompasses the HL7 domain of interest as a whole. The RIM is a coherent, shared information model that is the source for the data content of all HL7 messages. As such, it provides consistent data and concept reuse across multiple information structures, including messages. The RIM is intentionally abstract allowing it to represent the richness of the information topics that must be shared throughout the health system. The principles underlying this abstraction are discussed in detail in the introduction to the RIM.

- *Domain Message Information Model (D-MIM):* A D-MIM is a refined subset of the RIM that includes a set of class clones, attributes and relationships that can be used to create messages for a particular domain (a particular area of interest in healthcare). The D-MIM is used as a common base upon which all R-MIMs within a domain are built. The rules for the refinement process by which the D-MIMs and R-MIMs are created are discussed in detail in the specification on Refinement and Localization.

- *Refined Message Information Model (R-MIM):* The R-MIM is a subset of a D-MIM that is used to express the information content for a message or set of messages with annotations and refinements that are message specific. The content of an R-MIM is drawn from the D-MIM for the specific domain in which the R-MIM is used. The R-MIM may include clones of selected classes with alias names specific to the perspective of the message(s) to be derived. The R-MIM represents the information content for one or more abstract message structures, also called Hierarchal Message Definitions (HMDs).

The information models provide structure for message specifications. HL7 v3 messaging specifications derived from HL7 RIM are as follows:

Hierarchical Message Description (HMD): HMD is a tabular representation of the sequence of elements (i.e., classes, attributes and associations) represented in an R-MIM. Each HMD produces a single base message template from which the specific message types are drawn.

Message Type: Message Types are derived from R-MIM's to represent the actual information that need to be communicated. Each Message Type is intended to represents a certain set of related data fields. For example, message type Observation Request (Message ID POOB_MT210000UV) includes fields such as Author, Subject, Placer, Observation Code and Status. Message Types are enveloped with transmission and control related information to give context to the data captured in a message. This enables each message type to be reused for different purposes with the same data requirements. For example, Observation Request message type is used with different control information to request, cancel, revise and complete and observation.

3.1.2 HL7 v3 Interactions

HL7 defines *Interactions* as a unique association between a specific message type (information transfer), a particular trigger event that initiates or "triggers" the transfer, and the Receiver Responsibilities (in terms of 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. 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 healthcare transaction must be mapped to the correct set of Interactions and not a Message Type. Therefore, Interactions form the heart of the proposed process for HL7 message extraction.

3.2 *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. The current SNOMED CT version contains close to 283,000 active concepts, 732,000 active terms and 923,000 active relationships, making it the most comprehensive standard terminology system in the world.SNOMED CT is recommended by HL7 Organization as a terminology standard for clinical data exchange.

3.3 *LOINC*

It is a database of codes representing terms used primarily in the Laboratory and Observation areas of healthcare. LOINC 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 [24]. LOINC has been identified by the HL7 Standards Development Organization as a preferred code set for laboratory test names in transactions between health care facilities, laboratories, laboratory testing devices, and public health authorities. Unlike SNOMED, LOINC codes are not organized in any symmetrical or hierarchical manner, thus making the codes arbitrary.

3.4 Resource Description Framework (RDF)

RDF [18, 34] 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. Efforts are underway to RDF-enable object oriented modeling tools such as UML [5] creating potential for the tool to be integrated with UML tools in the application design phase of integration projects. Further, since RDF uses URI's, other resources such as documentation, schema, xml files etc. which are related to an artifact, can be retrieved from the web server with ease. By using RDF, any future changes to the HL7 information models and the artifact metamodel can be accommodated with minimum effort, whereas storing artifact metadata in a relational database would require the underlying data model to be updated.

In recent years a number of Semantic Web (SW) languages such as Web Ontology Language (OWL) [8], Ontology Inference Layer (OIL) [18] and DARPA Agent Markup Language (DAML) [12] 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.

3.5 Sesame framework

Sesame [6] 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 inferencing and query language Sesame Query Language (SeRQL) [7] to query and find implicit information in RDF schema and data.

3.6 Apache Lucene

Apache Lucene [2] is a high-performance, full-featured text search engine library written in Java. It is a technology suitable for nearly any application that requires full-text search, especially cross-platform. In our tool, we leverage the capabilities of Lucene to provide word and phrase mapping between textual medical scenarios and keywords in our database.

Chapter 4

Proposed approach

This chapter describes an approach that simplifies translation of healthcare transactions to HL7 v3 Interactions with the use of a novel tool. Following sections detail the proposed process and underlying concepts.

4.1 Extracting HL7 v3 metadata

The purpose of the above mentioned tool is to aid integrators to map healthcare transactions with HL7 v3 Interactions most appropriate to communicate their content and context. For this, specific relationships between real-world healthcare transactions and Interactions needed to be established 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.

Default metadata associated with HL7 Interactions are the D-MIM's (domains) and R-MIM's (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 thesis, there are inconsistencies among information hierarchies of different domains. Using our exhaustive knowledge of HL7 v3 information models, we developed the following pieces of metadata to drive the search tool.

4.1.1 Interaction Context

Using a holistic view of HL7 information model, we re-classified the domains and sub-domains of original HL7 v3 model in a more intuitive 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 finalized, each Interaction was associated with a single *Context*. *Context* acts as a key piece of metadata in the search tool. The complete list of Contexts along with their associated D-MIM's is given in Table 4.1.

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
Specialized Care and Professional Services	Care Provision (REPC_DM000000UV)	specialists, physiotherapy, psychology, counseling
Patient Health Condition	Care Provision (REPC_DM000000UV)	patient medical conditions
Family/Surgical History	Care Provision (REPC_DM000000UV0	family history, surgical history
Discharge Report	Care Provision (REPC_DM000000UV)	discharge report
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 - 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

Table 4.1: HL7 v3 Contexts, their description and associated D-MIMs.

Context	HL7 Domain and D-MIM	Description
Payment Notice	Claims and Reimbursements (FICR_DM000001UV)	payment notice
Statement of Financial Activity	Claims and Reimbursements (FICR_DM000001UV)	financial statement
Immunization	Immunization (POIZ_DM000000UV)	vaccination, substance administration, immunization
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
Regulatory Affairs	Personnel Management (PRPM_DM000000UV)	healthcare regulations
Healthcare Provider Registry	Personnel Management (PRPM_DM000000UV)	healthcare provider registry
Health Organization Registry	Personnel Management (PRPM_DM000000UV)	healthcare organization registry
Pharmacy - Patient Medication Record	Pharmacy (PORX_DM000000UV)	patient medication report, medication profile
Pharmacy - Device Dispensing	Pharmacy (PORX_DM000000UV)	medical device dispensing
Pharmacy - Dispensing	Pharmacy (PORX_DM000000UV)	prescription, dispensing
Pharmacy - Supply	Pharmacy (PORX_DM000000UV)	pharmacy supply
Pharmacy - Non-prescription medication statements	Pharmacy (PORX_DM000000UV)	non prescription medication, over the counter medication
Pharmacy - Contraindications	Pharmacy (PORX_DM000000UV)	contraindications
Device Tracking	Registries (PRRG_DM000000UV)	track devices
Provider Tracking	Registries (PRRG_DM000000UV)	track provider
Case Management	Regulated Reporting (PORR_DM000000UV)	case management, outbreak management, investigation
Scheduling	Scheduling (PRSC_DM000000UV)	appointment, schedule

Table 4.2: HL7 v3 Contexts, their description and associated D-MIMs (Contd.)

4.1.2 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 the *Interaction Classification Model.* The classes in the model and their descriptions are given in Table 4.3. The *Class* of an Interaction is the next key piece of metadata that would drive our tool. The Interaction Classification Model hierarchy sub-classes Interactions at three levels.

At Level I, an Interaction is sub classed 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.

At Level II, Initiator Interaction can further be classified into *Command, Query* and *Notification. Command* refers to an Interaction ordering the receiver to perform a task. *Query* represent requests for information. *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 sent to acknowledge its receipt in general and at times, the accept/reject status by the receiver. *InformationR* represent Query results or information sent as response to a Command requesting data.

At Level III, *Command* and *Notification* is classified in to 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 sub-categorized 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.

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 action.***
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.
Create	Order receiving system to create a record.
Delete	Order receiving system to delete a record.
Information	Order receiving system to send some information.
Reactivate	Order receiving system to reactivate a record.
Record	Order receiving system to record information contained in the message body.
Replace	Order receiving system to replace previously sent information.
Resolve	Order receiving system to resolve duplicate records.
Retract	Order receiving system to discard previously sent information.
Resume	Order receiving system to resume a previously suspended operation.
Status	Order receiving system to provide status of an operation.
Suspend	Order receiving system to suspend a previously activated operation.
Update	Order receiving system to revise a record.
Other	Order receiving system to perform activities other than what's listed above.
Notification	***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.
Authorize	Notify receiving systems of an authorize operation.
Cancel	Notify receiving systems of an cancel operation.
Complete	Notify receiving systems of an complete operation.
Create	Notify receiving systems of an create operation.
Delete	Notify receiving systems of an delete operation.
Information	Notify receiving systems of information asynchronously.
Reactivate	Notify receiving systems of an reactivate operation.
Record	Notify receiving systems of an record operation.
Replace	Notify receiving systems of an replace operation.
Resolve	Notify receiving systems of an resolve operation.
Retract	Notify receiving systems of an retract operation.
Resume	Notify receiving systems of an resume operation.
Status	Notify receiving systems of a the status of an operation.
Suspend	Notify receiving systems of an suspend operation.
Update	Notify receiving systems of an update operation.
Other	Notify receiving systems of any other operation.
Response	**Respond to a command, query or notification.**
Acknowledgement	***Acknowledge the receipt of a message indicate if commandnotification is accepted for processing.***
Received	Acknowledge that a particular message was received.
Accepted	Inform that the receiver accepts to process a CommandQueryNotification.
Rejected	Inform that the receiver rejects to process a CommandQueryNotification.
Information	***Response to a command to send information/query.***
Summary	Summary information response.
Detail	Detailed information response.

Table 4.3: Definition of the classes in the Interaction Classification Model.

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 4.2.

HL7 v3 Information Model has specifications for over 900 Interactions in its 2009 January ballot [19]. In this study, we have categorized over 600 of these Interactions according to the concepts described above. A portion of the classification is presented in Table 4.4. These Interactions along with metadata of their classifications are stored within the repositories of our tool for searching.

4.2 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 Nima and Sartipi's [17] 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 *"Op-*

Interaction	Level I Concept	Level II Concept	Level III Concept	
Abort Care Provision(REPC_IN004610UV01)	Initiator	Notification	Abort	Care Provision
Activate Care Provision(REPC_IN004110UV01)	Initiator	Notification	Activate	Care Provision
Append Care Provision(REPC_IN004211UV01)	Initiator	Notification	Update	Care Provision
Complete Care Provision(REPC_IN004410UV01)	Initiator	Notification	Complete	Care Provision
Correct Care Provision(REPC_IN004210UV01)	Initiator	Notification	Update	Care Provision
Nullify Care Provision(REPC_IN004810UV01)	Initiator	Notification	Retract	Care Provision
Replace Care Provision(REPC_IN004913UV01)	Initiator	Notification	Replace	Care Provision
Report Care Provision(REPC_IN004014UV01)	Initiator	Notification	Report	Care Provision
Resume Care Provision(REPC_IN004510UV01)	Initiator	Notification	Resume	Care Provision
Suspend Care Provision(REPC_IN004310UV01)	Initiator	Notification	Suspend	Care Provision
Find Care Record Candidate Query(QUPC_IN041100UV01)	Initiator	Query	Detail	Care Record
Find Care Record Candidate Response(QUPC_IN041200UV01)	Response	Information	Detail	Care Record
Get Care Record Profile Query(QUPC_IN043100UV01)	Initiator	Query	Summary	Care Record
Get Care Record Profile Response(QUPC_IN043200UV01)	Response	Information	Summary	Care Record
Get Care Record Query(QUPC_IN040100UV01)	Initiator	Query	Detail	Care Record
Get Care Record Response(QUPC_IN040200UV01)	Response	Information	Detail	Care Record
Abort Care Transfer Request Notification(REPC_IN002620UV01)	Initiator	Command	Abort	Care Transfer
Care Transfer Promise(REPC_IN003130UV01)	Response	Acknowledgement	Accepted	Care Transfer
Care Transfer Request(REPC_IN002120UV01)	Initiator	Command	Update	Care Transfer
Notify Aborted Care Transfer Request(REPC_IN002610UV01)	Initiator	Notification	Abort	Care Transfer
Notify Care Transfer Promise(REPC_IN003010UV01)	Initiator	Notification	Information	Care Transfer
Notify Care Transfer Request(REPC_IN002110UV01)	Initiator	Notification	Other	Care Transfer
Notify Revise Care Transfer Request(REPC_IN002210UV01)	Initiator	Notification	Update	Care Transfer
Reject Care Transfer(REPC_IN002040UV01)	Response	Acknowledgement	Rejected	Care Transfer
Replace Care Transfer Promise(REPC_IN003930UV01)	Initiator	Notification	Replace	Care Transfer
Revise Care Transfer Request(REPC_IN002220UV01)	Initiator	Command	Update	Care Transfer
Add allergy/intolerance request accepted(REPC_IN000013UV01)	Response	Acknowledgement	Accepted	Allergies
Add allergy/intolerance request refused(REPC_IN000014UV01)	Response	Acknowledgement	Rejected	Allergies
Update allergy/intolerance request accepted(REPC_IN000021UV01)	Response	Acknowledgement	Accepted	Allergies
Update allergy/intolerance request refused(REPC_IN000022UV01)	Response	Acknowledgement	Rejected	Allergies

Table 4.4: A portion of HL7 v3 Interactions along with corresponding Context and Interaction Class metadata.

eration" 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:

$$Transaction : \{Message\}^{1..N} \tag{4.1}$$

$$Message : \{Actor\}^{2..N} + Operation + \{Data\}^{1..N} \tag{4.2}$$

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 4.1.

Now it's possible to incorporate the above concepts into a search tool that will map healthcare transactions to Interactions. The search tool facilitates input of 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 Chapter 5.

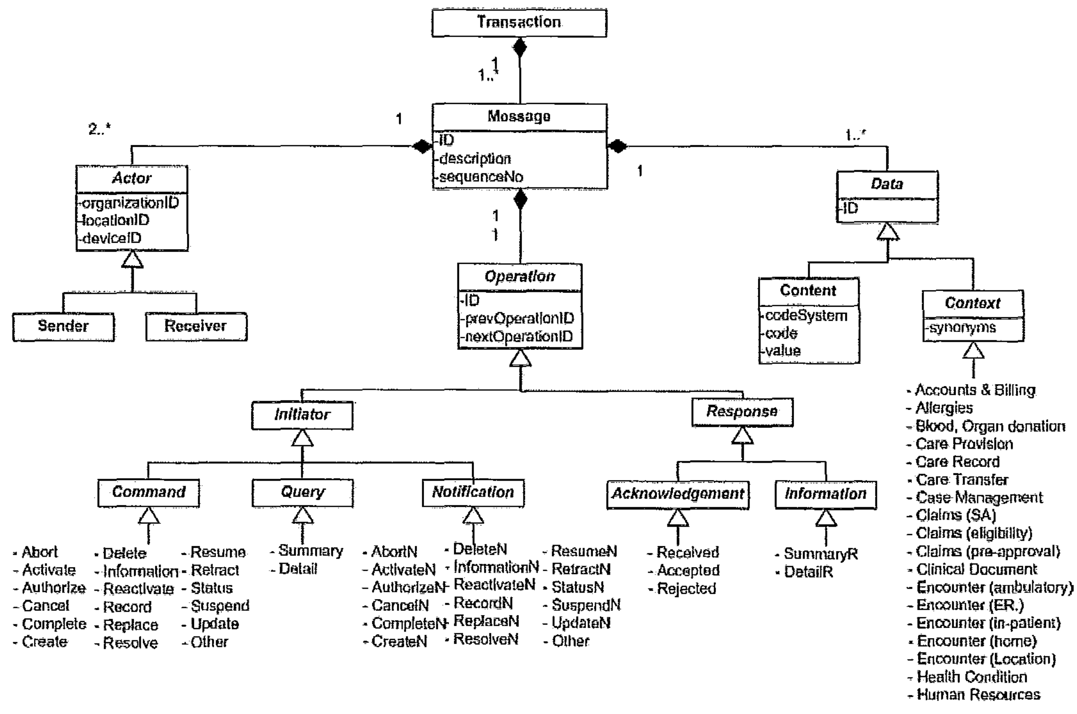

Figure 4.1: Healthcare transaction schema. The schema describes a healthcare transaction as a composition of Actors, Actions and Data.

4.3 Proposed process

We propose a 3-stepped process 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.

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:

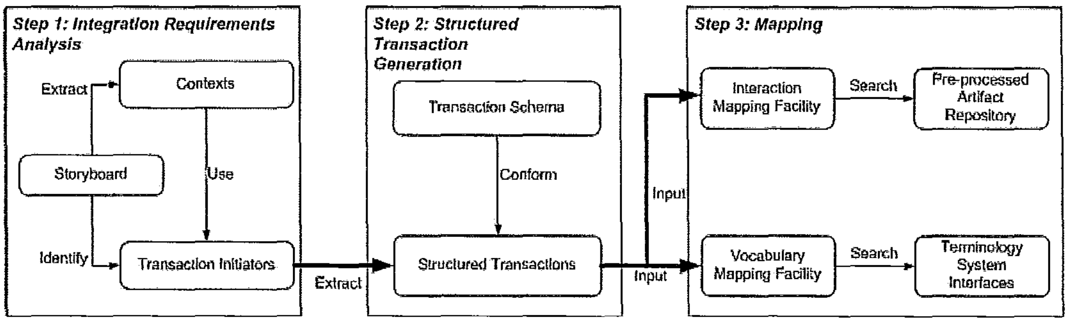

Figure 4.2: Proposed process for translating healthcare transactions to HL7 v3 Interactions

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 "Visit to Physician to Refill Prescription" given below 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 visists 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.*

As seen in the above example, typically, 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 following activities to impose structure on the information in Storyboards.

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 describing it. We have used Lucene [2] and SNOMED vocabularies to incorporate as many cognitive synonyms and phrases to describe each Context. 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

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.

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 in structured format as follows:

EMR sends request for patient medication history.

Actor: EMR

Action: QueryDetail

Context: Pharmacy - Patient Medication Record

Content: medication history

EMR requests for patient allergies.

Actor: EMR

Action: QueryDetail

Context: Allergy

Content: patient allergies

EMR updates EHR with patient medication.

Actor: EMR

Action: CommandUpdate

Context: Pharmacy - Patient Medication Record

Content: patient medication

EMR updates EHR with allergies.

Actor: EMR

Action: CommandUpdate

Context: Allergy

Content: adverse reaction

EMR sends prescription request to pharmacy to dispense.

Actor: EMR

Action: CommandOther

Context: Pharmacy - Dispensing

Content: prescription

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 4.5: Storyboard *Medication Refill* - transaction initiators and corresponding interactions mapped to them.

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 has 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 4.5 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 HIS 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

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

Action: QueryDetail

Context: Health Condition

Content: health record

HIS requests EHR to update demographic information.

Actor: HIS

Action: CommandUpdate

Context: Patient Administration

Content: demographic information

HIS sends emergency record to X's Primary care physician.

Actor: HIS

Action: NotificationInformation

Context: Emergency Encounter

Content: emergency record

HIS sends emergency record to EHR.

Actor: HIS

Action: NotificationInformation

Context: Emergency Encounter

Content: emergency record

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)

Table 4.6: Storyboard *Emergency Encounter* - transaction initiators and corresponding Interactions best suited to represent them.

Table 4.6 gives transaction initiators and corresponding Interactions mapped to them:

Chapter 5

The tool

The proposed tool is open source and web-based. The architecture of the tool is illustrated in Figure 5.1 and described in the next section.

5.1 Architecture

The tool is a client-server, multi-layer application. It is composed of the following components:

- Web User Interface
- Search Controller
- Terminology System Interface
- Sesame Interface
- Lucene Interface
- Repository Layer (MySQL database)

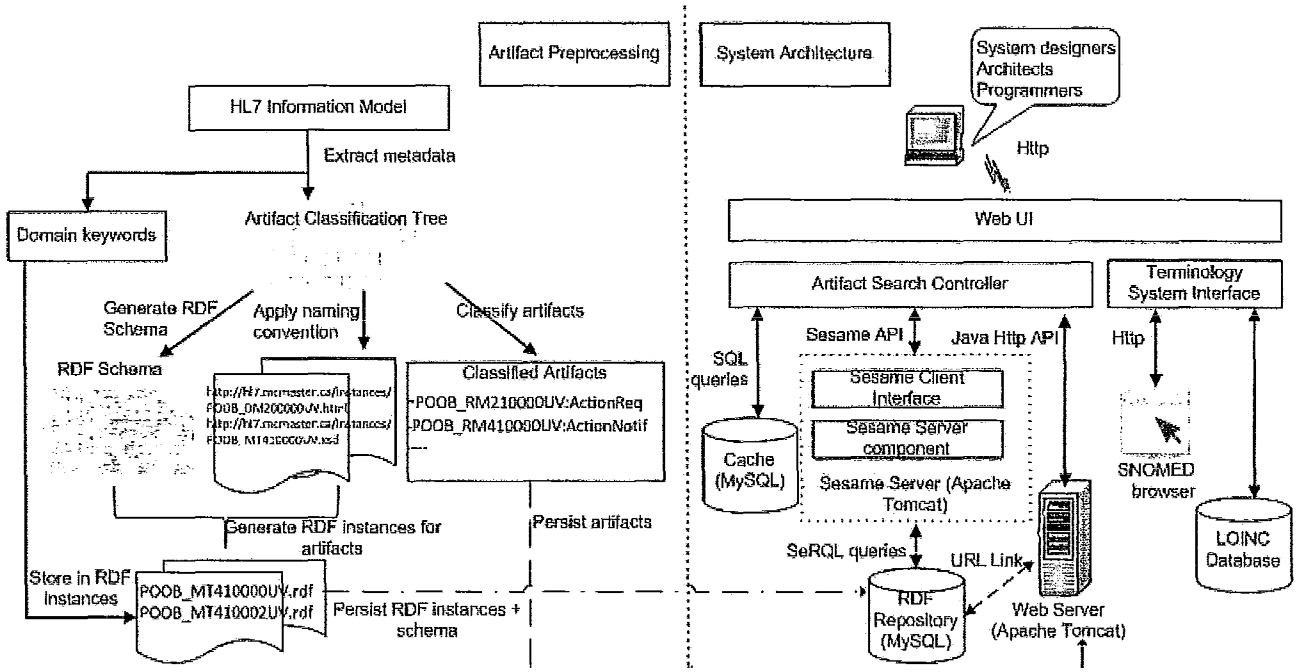

Figure 5.1: Tool architecture. Left half of the figure illustrates steps involved in preprocessing HL7 artifacts for persisting in artifact repositories. Right half illustrates the system architecture.

The web user interface is Java Servlet based and provides a user friendly environment for searching, browsing, navigating and exploring artifacts. 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 artifacts such as xml schema, documentation, xml sample instances, information models and other

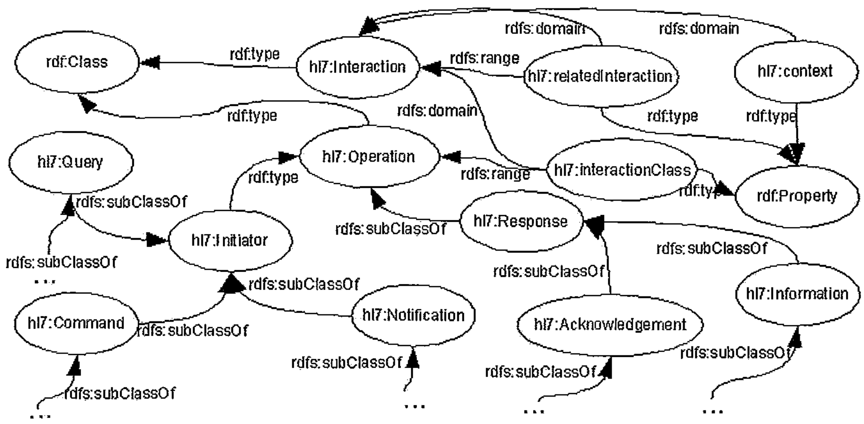

Figure 5.2: RDF Graph for HL7 artifact metadata model

representations for retrieval. The repository layer consists of MySQL databases for storing Contexts and synonyms and RDF instances. Right portion of Figure 5.1 illustrates the high level architecture.

5.2 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 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.

5.3 Search Controller

The 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 the popular open source full-text search engine Lucene [2]. Hence it interfaces with Lucene's HttpRepository API. The RepositoryConnector is responsible for connecting to the Sesame HL7 Store, constructing SeRQL[7] queries and invoking Sesame API to search for RDF instances in sesame repository. As described in Section 5.6, these RDF instances contain valuable metadata that helps create a link between transactions and HL7 Interactions.

5.4 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 [29] and a MySQL database of LOINC codes.

5.5 Repository Layer

The tool maintains a MySQL database with HL7 Contexts and their cognitive synonyms. The synonyms are based on Wordnet[9], 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 with actual Interactions are maintained as a separate website on the Apache Tomcat server.

5.6 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 artifacts 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). Left half of Figure 5.1 details the activities involved in offline artifact 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 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 is 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.

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 5.2 illustrates RDF graph of HL7 v3 Interaction metadata model. A part of the XML serialization of the schema is given below.


```
<?xml version="1.0"?>

<!DOCTYPE rdf:RDF [<!ENTITY xsd "http://www.w3.org/2001/XMLSchema#">]>

<rdf:RDF
```

```
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xml:base="http://www.mcmaster.ca/hl7/schema">
<rdfs:Class rdf:ID="Interaction"/>
<rdfs:Class rdf:ID="Operation"/>
<rdfs:Class rdf:ID="Command">
    <rdfs:subClassOf rdf:resource="#Operation"/>
</rdfs:Class>
<rdfs:Class rdf:ID="Activate">
    <rdfs:subClassOf rdf:resource="#Command"/>
</rdfs:Class>
<rdfs:Class rdf:ID="Authorize">
    <rdfs:subClassOf rdf:resource="#Command"/>
</rdfs:Class>
<rdfs:Class rdf:ID="Query">
    <rdfs:subClassOf rdf:resource="#Operation"/>
</rdfs:Class>
...
<rdfs:Class rdf:ID="Response">
    <rdfs:subClassOf rdf:resource="#Operation"/>
</rdfs:Class>
<rdfs:Class rdf:ID="Acknowledgement">
    <rdfs:subClassOf rdf:resource="#Response"/>
</rdfs:Class>
```

```
...

<rdf:Property rdf:ID="interactionClass">
    <rdfs:range rdf:resource="#Operation"/>
    <rdfs:domain rdf:resource="#Interaction"/>
</rdf:Property>
<rdf:Property rdf:ID="context">
    <rdfs:domain rdf:resource="#Interaction"/>
</rdf:Property>
</rdf:RDF>
```

A section of the RDF instance for Interation "Request to record subject observation" (Artifact 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>
```

```
12.</rdf:RDF>
```

Line 2 begins an rdf:RDF element. On the same line, there is an XML namespace declaration for the rdf:RDF start-tag. This declaration specifies that all tags in this content prefixed with rdf: are part of the namespace identified by the URI reference http://www.w3.org/1999/02/22-rdf-syntax-ns#. URI references beginning with this namespace are used for terms from the RDF vocabulary. Line 3 specifies the XML namespace declaration for the prefix hl7:. This specifies that the namespace URI reference http://localhost:8080/hl7/schema# is to be associated with the hl7: prefix.

Lines 4 and 5 define the URI of the resource this RDF instance describes which is "http://localhost:8080/hl7/ interactions/POOB_IN000001UV. Lines 6 and 7 indicate that this resource is an instance of class Interaction. Lines 8 to 10 specify some of the properties of this resource. Lines 8 and 9 indicate that the value of property interactionClass for the resource is http://localhost:8080/hl7/schema#Record. Line 10 specifies that the value of property "context" is "OBS" which is the abbreviated code for Context "Non-laboratory Observation".

Chapter 6

Formal definitions

In this chapter, we express the concepts that we use throughout this thesis as formal definitions. We provide a model for representing the Interaction Classification Model described in Section 4.1.2.

- As per our discussion in Section 4.1.2, an Interaction is sub-classed at three levels. Let Cl_1, Cl_2 and Cl_3 denote the sets of classes at levels 1, 2 and 3 of the Interaction Classification Model respectively.
- Let Cl_1 be the set of Level 1 sub-classes in the Interaction Classification Model.
- $Cl_1 = \{Initiator, Response\}$
- Let Cl_{2In} be the set of subclasses of class Initiator.

  $Cl_{2In} = \{Query, Command, Notification\}$
- Let Cl_{2Re} be the set of subclasses of class Response.

  $Cl_{2Re} = \{Acknowledgement, Information\}$

- We can define Cl_2 as follows:

  $Cl_2 = Cl_{2In}\ U\ Cl_{2Re}$

- Let Cl_{3Qu} be the set of subclasses of class Query.

  $Cl_{3Que} = \{Summary, Detail\}$

- Let Cl_{3Com} be the set of subclasses of class Command.

  $Cl_{3Com} = \{Abort, Activate, Authorize, Cancel, Complete, Create, Delete,$

  $Information, Re - activate, Record, Replace, Resolve, Resume,$

  $Retract, Status, Suspend, Update, Other\}$

- Let Cl_{3Not} be the set of subclasses of class Notification.

  $Cl_{3Not} = \{AbortN, ActivateN, AuthorizeN, CancelN, CompleteN, CreateN,$

  $DeleteN, InformationN, Re - activateN, RecordN, ReplaceN,$

  $ResolveN, ResumeN, RetractN, StatusN, SuspendN, UpdateN, OtherN\}$

- Let Cl_{3Ack} be the set of subclasses of class Acknowledgement.

  $Cl_{3Ack} = \{Received, Accepted, Rejected\}$

- Let Cl_{3Inf} be the set of subclasses of class Information.

  $Cl_{3Inf} = \{SummaryR, DetailR\}$

- We can define Cl_3 as follows:

  $Cl_3 = Cl_{3Com} \bigcup Cl_{3Not} \bigcup Cl_{3Qu} \bigcup Cl_{3Ack} \bigcup Cl_{3Inf}$

- We define the set of Interactions specified by HL7 v3 as the set denoted by S.

- Let e be any interaction in S.

  $e \, \varepsilon \, S$

- Let $Cl_1(e)$, $Cl_2(e)$ and $Cl_3(e)$ be level1, 2 and 3 classifications of Interaction e respectively.

- Then we can express that,

  $\forall e.e\varepsilon S \bullet Cl_1(e) = Initiator \Rightarrow Cl_2(e)\varepsilon Cl_{2In} \bigcap$

  $Cl_1(e) = Response \Rightarrow Cl_2(e)\varepsilon Cl_{2Re} \bigcap$

  $Cl_2(e) = Command \Rightarrow Cl_3(e)\varepsilon Cl_{3Com} \bigcap$

  $Cl_2(e) = Notification \Rightarrow Cl_3(e)\varepsilon Cl_{3Not} \bigcap$

  $Cl_2(e) = Queue \Rightarrow Cl_3(e)\varepsilon Cl_{3Que} \bigcap$

  $Cl_2(e) = Acknowledgement \Rightarrow Cl_3(e)\varepsilon Cl_{3Ack} \bigcap$

  $Cl_2(e) = Information \Rightarrow Cl_3(e)\varepsilon Cl_{3Inf}$

- Finally, we can express classification C of an Interaction e as

  $C(e) = Cl_1(e) \bigcap Cl_2(e) \bigcap Cl_3(e)$

Chapter 7

Conclusion

In this thesis we presented a novel, well-defined approach to support message selection activity of communication design phase of HL7 v3 based system integration projects. We presented a behavior-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 using semantic web 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 Enterprize Search and Knowledge Management applications.

During future research, we also intend to improve the context mapping mechanism of the tool to use Natural Language Processing which will add further value to the tool. At this time our tool does not offer an environment for automated message instantiation and schema editing. For future research, we will continue to improve the tool to provide such features to render Transaction-to-HL7 Message translation as seamless as possible. An end-to-end message translation capable of automating the entire process of message selection and instantiation would be highly advantageous to the healthcare system integration community.

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.

Appendix

Following are screen shots of our implementation of HL7 support prototype tool. As part of future research we intend to improve the quality and features of the tool to represent our conceptual level design.

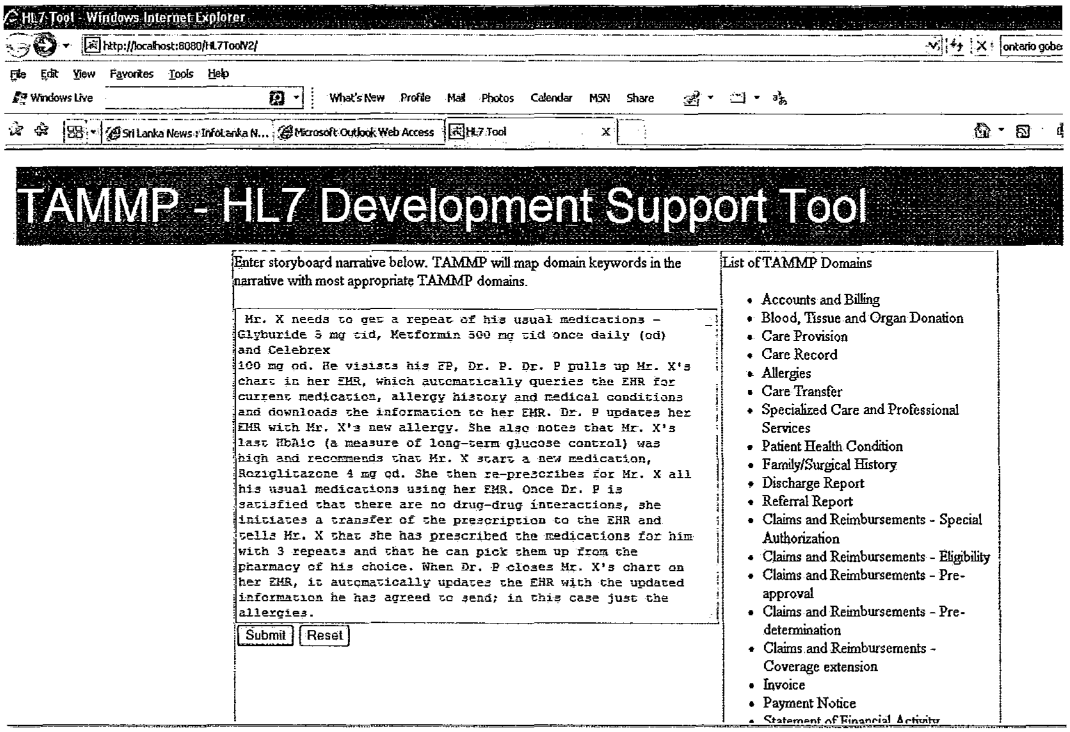

Figure 7.1: User enters storyboard as free-text.

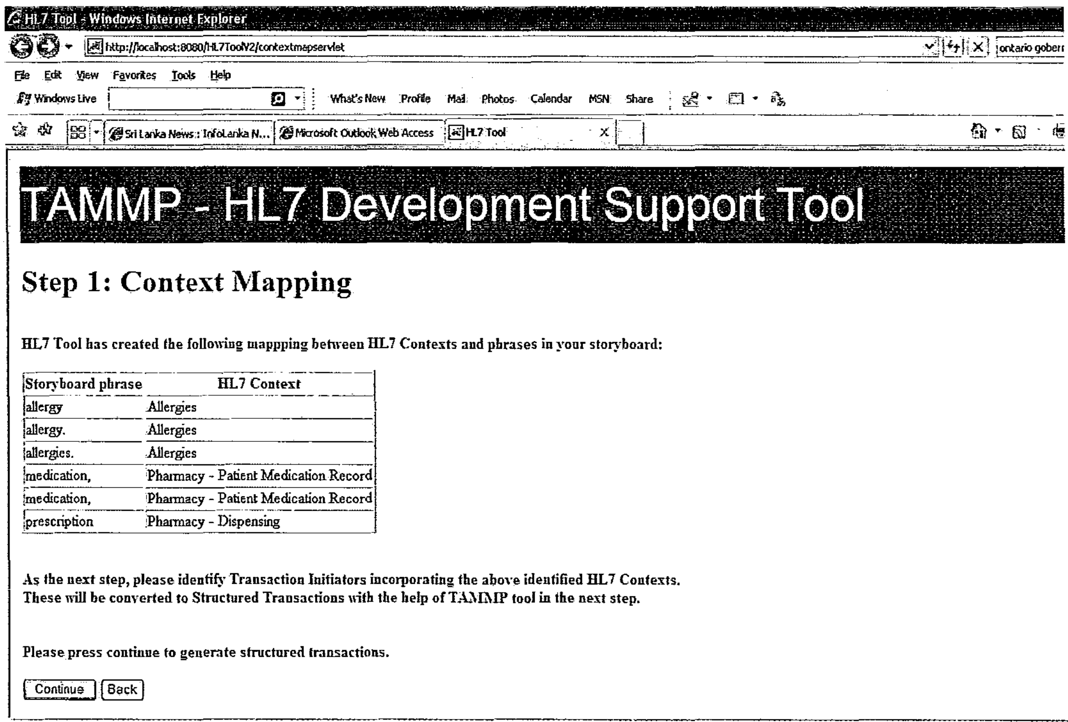

Figure 7.2: The tool maps storyboard text with HL7 Contexts.

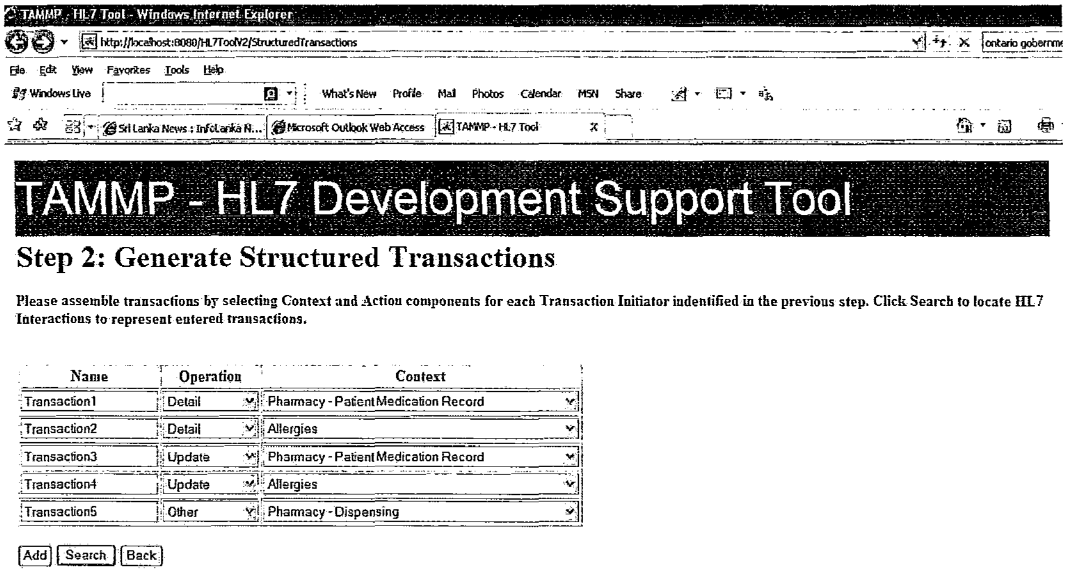

Figure 7.3: User enters structured transactions. The tool creates mappings between those and interactions in repository.

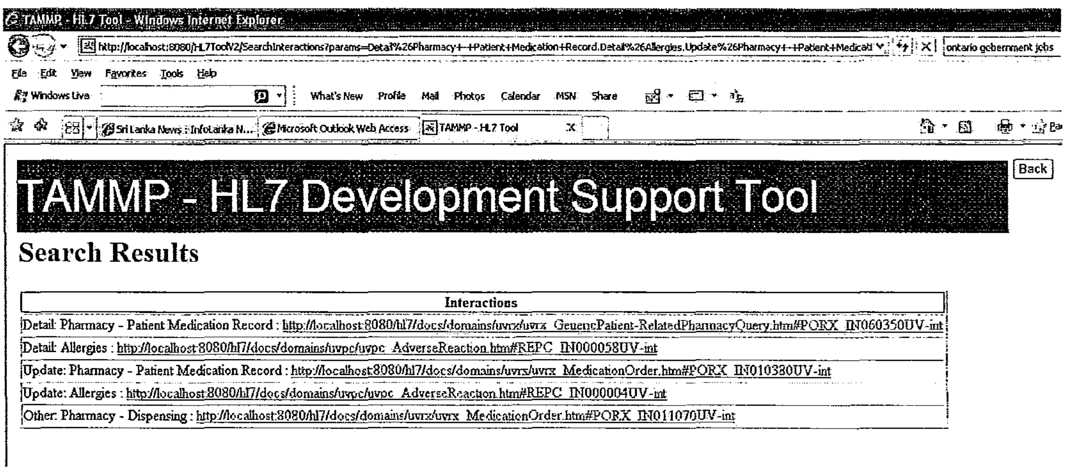

Figure 7.4: List of Interactions mapped to entered transactions.

Bibliography

[1] eHealth Ontario website
http://www.ehealthontario.on.ca.

[2] "Apache Lucene full-text search engine website.."
http://lucene.apache.org/java/docs/.

[3] "HIMSS webpage with definition of EHR.."
http://www.himss.org/ASP/topics_ehr.asp/.

[4] "openEHR website.."
http://www.openehr.org/home.html/.

[5] "Representing UML in RDF.."
http://infolab.stanford.edu/ melnik/rdf/uml/.

[6] "Sesame RDF framework website.."
http://www.openrdf.org/index.jsp.

[7] "Sesame RDF Query Language.."
http://www.openrdf.org/doc/sesame/users/ch06.html.

[8] "Web Ontology Language.."
http://www.w3.org/2004/OWL/.

[9] "Wordnet website.."
http://wordnet.princeton.edu/.

[10] 7Edit HL7 tool website.
http://www.7edit.com/home/index.php.

[11] 7Scan HL7 browser website.
http://www.7scan.com/.

[12] S. Arroyo, R. Lara, Y. Ding, M. Stollberg, and D. Fensel, "Semantic web languages - strengths and weaknesses." http://members.deri.at/ michaels/publications/sw-languages-IADIS04.pdf.

[13] Basit Chaudhry, Jerome Wang, Shinyi Wu, Margaret Maglione, Walter Mojica, Elizabeth Roth, Sally C. Morton and Paul G. Shekelle, "Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care," *Anals of Internal Medicine*, 2006.

[14] Canada Health Infoway website.
http://www.infoway-inforoute.ca/en/home/home.aspx.

[15] Congressional Budget Office Cost Estimate
http://www.cbo.gov/ftpdocs/99xx/doc9968/hr1.pdf.

[16] Department of Health and Human Services website.
http://healthit.hhs.gov/.

[17] N. Dezhkam and K. Sartipi, "Knowledge Transformation from Task Scenarios to View-based Design Diagrams," in *SEKE*, 2008.

[18] D. Fensel, "The Semantic Web and its Languages," *Intelligent Systems*, pp. 67–73, nov-dec 2000.

[19] Health Level 7 website.
http://www.hl7.org.

[20] Healthcare Information and Management Systems Society (HIMSS), "Electronic health records: A global perspective," tech. rep., HIMSS, 2008.

[21] Infoway EHR Infostructure.
http://www.infoway-inforoute.ca/en/WhatWeDo/Infostructure.aspx.

[22] 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*, pp. 552–555, may 2008.

[23] 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*, pp. 649–652, may 2008.

[24] LOINC website.
http://www.loinc.org.

[25] Massachusetts Technology Collaborative, "Advanced technologies to lower the cost of health care and improve quality," *Massachusetts Technology Park Corporation*, 2003.

[26] NeoTool HL7 Analyzer website.
http://www.corepointhealth.com/products/hl7-analyzer/hl7-analyzer.

[27] G. E. Rong Chen and G. O. Klein, "Julius a template based supplementary electronic health record system," *BMC Medical Informatics and Decision Making*, vol. 7, 2007.

[28] G. K. Rong Chen, "The openEHR Java Reference Implementation Project," *MEDINFO*, pp. 58–62, 2007.

[29] SNOMED browser.
http://www.jdet.com/.

[30] SNOMED CT Website.
http://www.ihtsdo.org/snomed-ct/.

[31] Southern California Evidence-based Practice Center, "Costs and benefits of health information technology," tech. rep., Agency for Healthcare Research and Quality, 2006.

[32] T.H.Yang, "A scalable multi-tier architecture for the National Thaiwan University hospital information system based on HL7 standard," in *CBMS, Germany*, pp. 99–104, 2006.

[33] Vishwanath, Arun and Scamurra, Susan, "Barriers to the adoption of electronic health records: Using concept mapping to develop a comprehensive empirical model," *Health Informatics Journal*, 2007.

[34] W3C RDF Primer.
http://www.w3.org/TR/2004/REC-rdf-primer-20040210/.

[35] M. Yarmand and K. Sartipi, "Standard-based Data and Service Interoperability in eHealth Systems," in *ICPC*, 2008.