Development of a Case Index for a Clinical Document Repository for Chronic Kidney Disease Management

By

Tuyet Thieu
B00105824
thieu@cs.dal.ca

Performed at
Medical Informatics, Dalhousie University

5849 University Avenue
Halifax, NS B3H 4H7

In partial fulfillment of the requirements of the Master of Health Informatics Program, Dalhousie University

Report of Internship for the period 6th September – 11th December, 2006

Date Submitted: 11th December, 2006

Supervisor: Grace Paterson, Clinician Informatician

(Signature :) _____

Date: _____

Acknowledgement and Endorsement

Tuyet Thieu has written this report in partial fulfillment of the requirements for the Master of Health Informatics Program at Dalhousie University. This report has not received any previous academic credit at Dalhousie University or any other institution.

I would like to thank Ms. Grace Paterson for her supervision on this internship and Ms. Deirdre Harvey for her support during this internship.

Tuyet Thieu

_____
Signature

_____
Date

Table of Contents

List of Figures

Executive Summary

The internship placement was supervised to build a case index for clinical documents repository in order to use as a portal website tool to teach medical resident students. This internship was performed from September 7, 2006 to December 9, 2006 and it had two objectives. The first objective is to learn how the standards of HL7 CDA improve data quality for research outcomes through the task of enhancing the existent discharge summary template by mapping the different diagnosis codes among standardized diagnostic systems: ICD-9, ICD-10-CA and SNOMEDCT. The second objective is primary and to focus on the use of HL7 CDA Version 2 for structure standardization, SNOMED CT for terminology standardization and topic maps for ontological standardization to develop a case index for a clinical document repository from hospital records and discharge summaries for the purpose of teaching medical students.

The first objective focused on the implementation of standardization and interoperability and allowed the intern to learn how medical terminologies and medical concepts are recorded differently among medical professionals, and how they are related to each other in a common ontology to enable sharing among health care providers. As the primary objective of this internship, the second objective is for the intern to learn the standards available to build a case index for clinical documents.

The conclusion has been draw and the recommendation has been made. One of the missions of health informatics is to inform and promote health information standards and technologies among healthcare communities and their stakeholders in term of management and sharing health information and knowledge to improve the outcomes of healthcare, research and teaching.

1 Introduction

The purpose of this internship was to support research into building a case index for a clinical documents repository for chronic kidney disease management. It was composed of three tasks:

- mapping lab test codes in French to the Systematized Nomenclature of Medicine Clinical Terms (SNOMEDCT) standard coding system in English version, diagnosis codes from International Classification of Disease (ICD) Version 9 into multiple vocabularies using ICD-10-CA, SNOMEDCT and Unified Medical Language System (UMLS) standard coding system;

- enhancing the PatientRx, a portal website tool, using a Health Level 7 (HL7) Clinical Document Architecture (CDA) Release 2 standard compliant XML Discharge Summary (DS) teaching template;

- developing case index using information and data from the Centre Informatisé de Recherche Évaluative en Santé et Systèmes de Soins (CIRESSS), university of Sherbrooke in Quebec and from the Capital District Health Authority (CDHA), Halifax, Nova Scotia;

The internship was instrumental in the production of a paper, "Semantic Indexing of Patient Cases in a Boundary Infostructure for eHealth" by Grace I. Paterson, Paul Fabry, Andrew M. Grant, Tuyet T. Thieu, Steven D. Soroka, Hassan Diab, Andriy M. Moshyk.

2 Description of the Organization

Medical Informatics, Faculty of Medicine, Dalhousie University, has major areas of focus in medical informatics including: biomedical-systems technology, health-information systems, medical-decision-support systems, image and signal analysis, statistical techniques and modeling, education and training, health-care management, and the human-machine interface. Medical Informatics Faculty has been in partnership with the Faculties of Computer Science and Business Administration, Dalhousie University since 2002 to begin the first Master of Health

Informatics program in Canada in the endeavor of using information technology to improve the outcomes of healthcare, research, teaching and health services administration.

3 Tasks Performed During the Internship and Lesson Learned

3.1 Mapping Lab Test and Diagnostic Codes

3.1.1 Background

There are many different coding systems available for vendors and hospitals to use. The decision makers consider many factors when choosing a coding system -- preferences, language of system users, electronic versions available, and organizations' budgets. When it comes to the need of sharing information among healthcare professionals, providers, researchers and other healthcare stakeholders, understanding how terms are coded and represented in different coding systems is important. Mapping from one coding system, ICD-9 as an example, into other coding systems, ICD-10 and SNOMEDCT for example, can be done by using Unified Medical Language System MetaMap Transfer (UMLS MMTx) and SNOCODE software to process natural language to capture the medical concepts and their semantic categories from raw data files. Since patients' data in this research came from CIRESSS and had been stored in a data warehouse database and converted to excel data files in French, there was no available software to map French to English. Thus, manually mapping lab data and diagnostic codes of ICD-9 coding system from French to English version ICD-10 and SNOMEDCT using UMLS as a switching language for different coding systems: ICD-9, ICD-10 and SNOMEDCT was the only way.

3.1.2 Process of Mapping Data in Excel format to MySql Database

Patients' data from CIRESS were stored in data warehouse database and converted to files and organized in worksheets. To extract the lab tests of five different patients, a formula to lookup test names in French (TITLE_CHAMP) and map them with test codes of SNOMEDCT.

2006 Fall Health Informatics Internship Report

Figure 1: The original lab data in French.

Figure 2: The lab test data after the mapping and cleaning redundant and irrelevant lab test data

**2006 Fall Health Informatics Internship Report**

Figure 3: The conversion of data from Microsoft Excel worksheet to Mysql table for the patients' lab test

Figure 4: The result of manually mapping diagnostic codes in Mysql table

2006 Fall Health Informatics Internship Report

3.1.3 Objective and Lessons Learned

From more than 1007 patients' records stored in Excel data files in French, 150 lab test codes and 100 diagnostic codes were extracted and mapped using Excel formulas. The mapping process used UMLS version 2006-AA as a switching language for ICD-9-CM, ICD-10, and SNOMEDCT. Other tools included SNOMEDCT Clue Browser and SNOMEDCT web browser provided by National Cancer Institute to generate SNOMEDCT description files for each diagnostic code after the mapping. The intern encoded lab tests and diagnostic codes into the four code systems, UMLS, SNOMEDCT, ICD-10-CA and ICD-9 and learned basic knowledge about the different semantic structures, hierarchies and coding guidelines of these coding systems.

3.2 Enhancing the PatientRx, a Portal Website Tool to Generate HER for CKD Patients

3.2.1 Background

The PatientRx, a portal website tool owned by Medical Informatics Faculty, was created to record, retrieve health information and data of individual patient and generate a patient electronic health record using HL7 CDA Release 2 standards. PatientRx v.1 was designed for direct patient data entry. The enhancement, PatientRx v.2, was enhanced to enable populating a longitudinal EHR from a CIRESSS extract represented using Microsoft Excel. To support case indexing of clinical data and its reuse by multiple communities of practice, PatientRx v.2 needed to support the ICD9 coding used in CIRESSS and the SNOMED CT and ICD-10-CA coding used in the Chronic Kidney Disease discharge summary.

In order to build a case index for clinical document repository for CKD managements, after the mapping task, the longitudinal patients' data needed to be reconstructed as an electronic health record (EHR) in XML format. The first version of PatientRx was created and represented patients' information and diagnostic codes using the local codes set by Canadian Organ

Replacement Registry (CORR), the national information system for renal dialysis, organ transplantation, organ waiting lists, and organ donation. In the enhanced version, SNOMEDCT and ICD-10 were used to record data of laboratory test and diagnostic codes in patients' electronic health records.

3.2.2 Process of Enhancing the PatientRx Version 2 Interfaces

After the mapping and converting patients' data from Microsoft Excel to MySql database formats, Cold Fusion was used as a scripting language to modify interfaces and Altova XMLSPY was used to modify the Extensible Stylesheet (XSL) files to generate patients' clinical documents in XML format for the enhanced PatientRx Version 2.

Figure 5: Diagnostic codes using CORR coding system in PatientRx Version 1
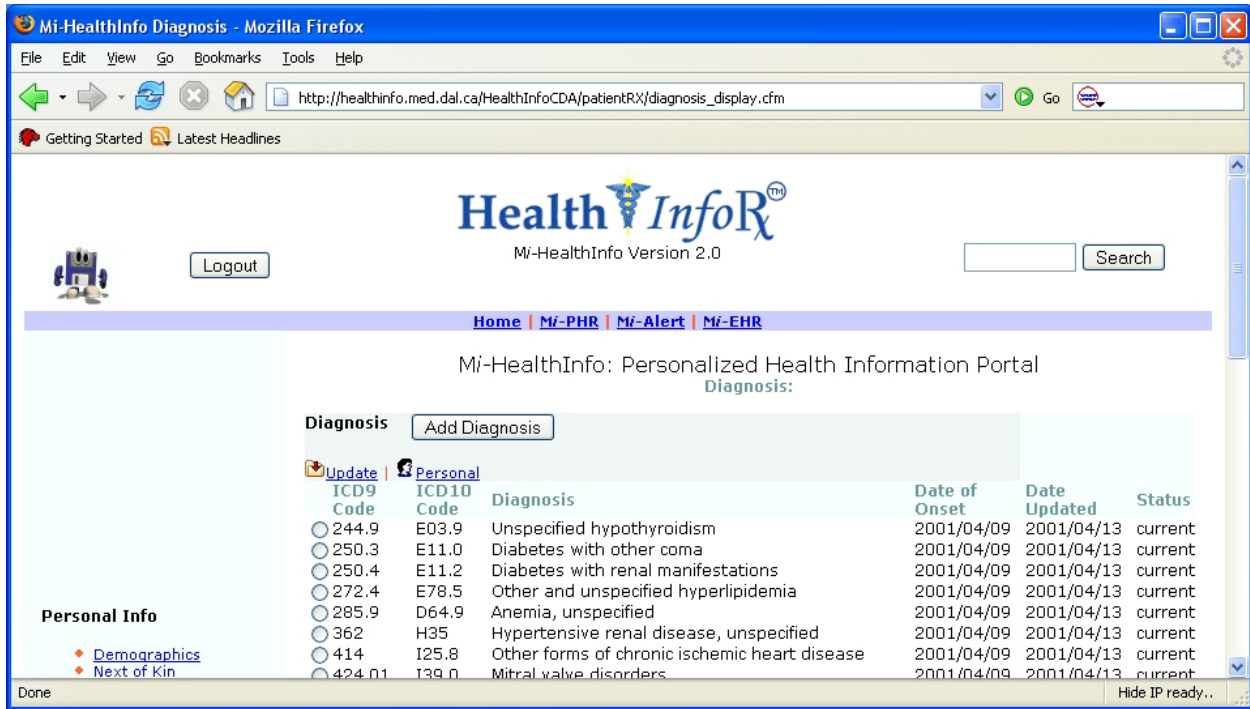
Figure 6: Diagnostic codes using ICD-9 and ICD-10-CA mapping in the enhanced PatientRx Version 2
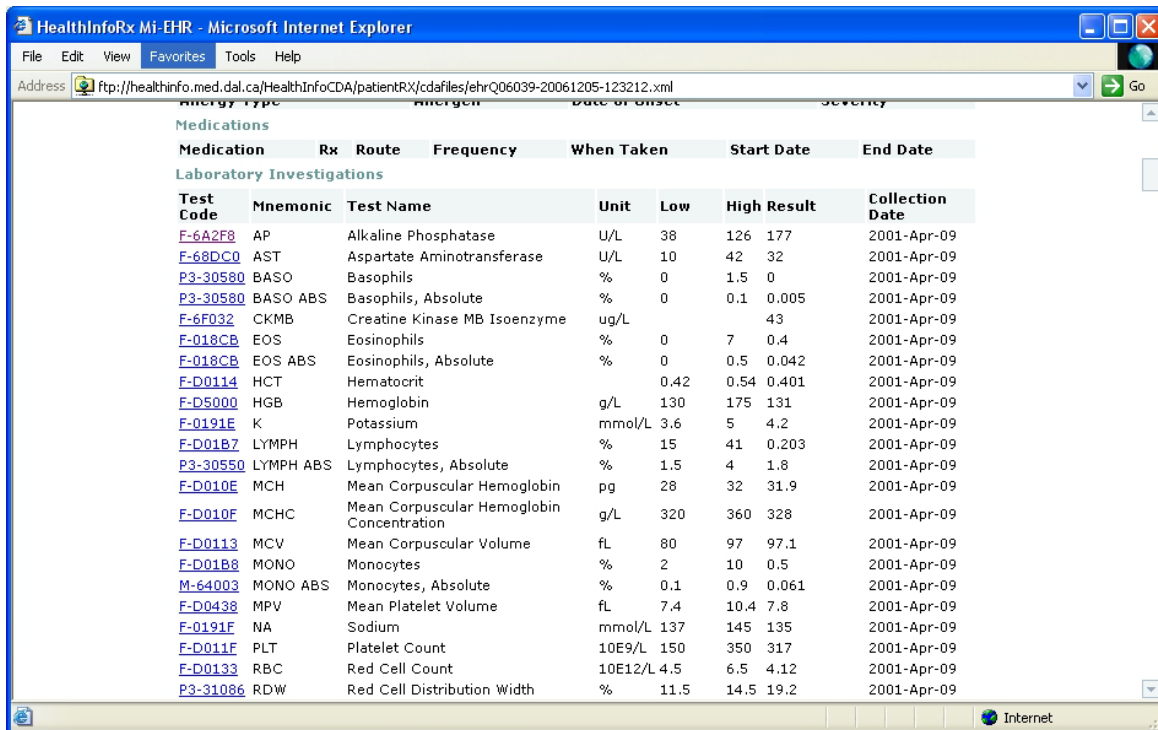
Figure 7: SNOMEDCT lab test codes encoded in the enhanced PatientRx Version 2

2006 Fall Health Informatics Internship Report
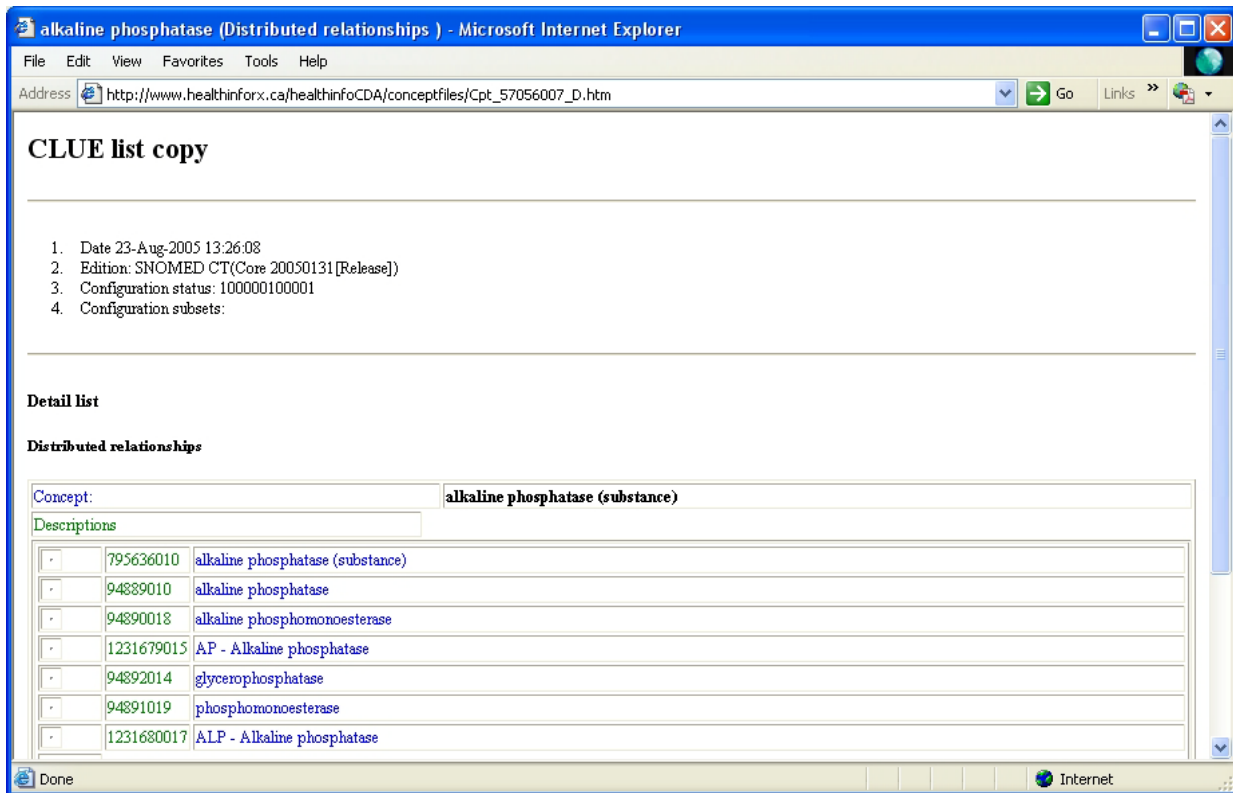
Figure 8: SNOMEDCT description file for test code F-6A2F8, Alkaline Phosphatase

3.2.3 Objective and Lessons Learned

Enhancing the discharge summary HL7 CDA compliant template using the existent PatientRx application is one of objectives of this internship. The intern has learned the process of using HL7 CDA Release 2 standards to build a compliant discharge summary template and the use of Extensible Stylesheet (XSL) and Cold Fusion for scripting language as authoring tools. The intern has also learned the strategy of how to bridge knowledge gap in developing a CDA template.

3.3 Developing Case Index for Clinical Document Repository for CKD Using Information and Data from CIRESSS and CDHA

3.3.1 Background

"Chronic kidney disease is a progressive condition that results in significant morbidity and mortality. Because of the important role the kidneys play in maintaining homeostasis, chronic kidney disease can affect almost every body system. Early recognition and intervention are essential to slowing disease progression, maintaining quality of life, and improving outcomes." [1] The HealthInfoCDA research project was conducted by Ms. Grace Paterson with the aims at providing an educational intervention to medical residents treating patients with kidney disease secondary to diabetes and hypertension. In this research, the Problem Based Learning (PBL) approach was proposed, designed and tested as the instructional teaching method because of its popularity in medical education worldwide. PBL cases were automatically generated from and promoted the use of Electronic Health Record (EHR) which derived from actual patient problems paired with clinical solutions as an educational resource depicting a clinical context and information resources available for medical students to learn both clinical sciences and problem-solving skills [2].

However, there is an essential issue about the semantic interoperability among electronic health records when in the same patient record, there are different clinical pragmatic patterns recorded by different healthcare providers. The semantic interoperability approach is not only support disease management but also for life-long learners in medical education process. "An infostructure may bridge the semantic gap caused by different ways of capturing clinical data in records, especially one composed of boundary objects. Boundary objects are pragmatic constructions that do the job required of them. They are implementations of medical language and information architecture standards used in electronic health records (EHR)." [3] The task of constructing a semantic indexing using the boundary infostructure of 18 case instances from CIRESSS and CDHA was the primary objective of this internship.

3.3.2 Process of Developing Case Indexing for CKD Clinical Document Repository

Most of patients diagnosed with chronic kidney disease have a long history of their medical records from the time they first were diagnosed and treated with the kidney disease during their illness. The lab tests required for case indexing for kidney disease were chosen and grouped by test methods as the markers of disease progression over time periods. The number of visits of patients and lab tests' results were grouped according to the time of admission in and of discharge from the facilities where they had been seen by medical professionals. The calculation for Glomerular Filtration Rate (GFR) and CKD stage calculations were manually entered using Modification of Diet in Renal Disease (MDRD) calculation embedded in CKD template designed for HealthInfoCDA project. The diagnostic codes for each case were extracted from each visit of the patient.

Table 1 – Semantic Case Indexing by Subject [3]

Case	GFR	CKD Stage	Anemic (ICD_Dx, Ferritin, % Sat, Hgb)	Proteinuria
Q06039-1			<ICD-9=285.9> <Hgb normal→low>	
Q06039-2			<Hgb low>	
Q06039-3	5	5	<ICD-9=285.8,285.9> <Hgb low>	3+
Q61372-1			<Hgb normal>	
Q61372-2			<Ferritin normal> <Hgb low→normal>	
Q61372-3	6	5	<% saturation low> <Hbg normal>	
Q61372-4	7	5	<Hgb low→normal>	
Q61372-5	7	5	<Ferritin high> <Hgb low>	
Q61372-6	6	5	<Ferritin high> <% saturation low> <Hgb low>	3+
Q61965-1	6	5	<Ferritin normal> <% saturation low> <Hgb low>	
Q34501-1	12	5	<Ferritin high> <Hgb low>	negative
Q34501-2	13	5	<% saturation low> <Hgb normal>	3+
Q52165-1	6	5	<ICD-9=285.9> <ferritin normal→high> <%saturation normal> <Hgb low>	
Q52165-2	5	5	<Hgb normal→low>	
Q52165-3			<Hgb low>	
Q52165-4			<ICD-9=285.9> <% saturation high> <Hgb low>	
Q52165-5			< abnormal> <Hgb low>	
NS0093-1	7	5	<ICD-10-CA=D50.9> <Ferritin normal> <% saturation low> <Hgb low>	3+
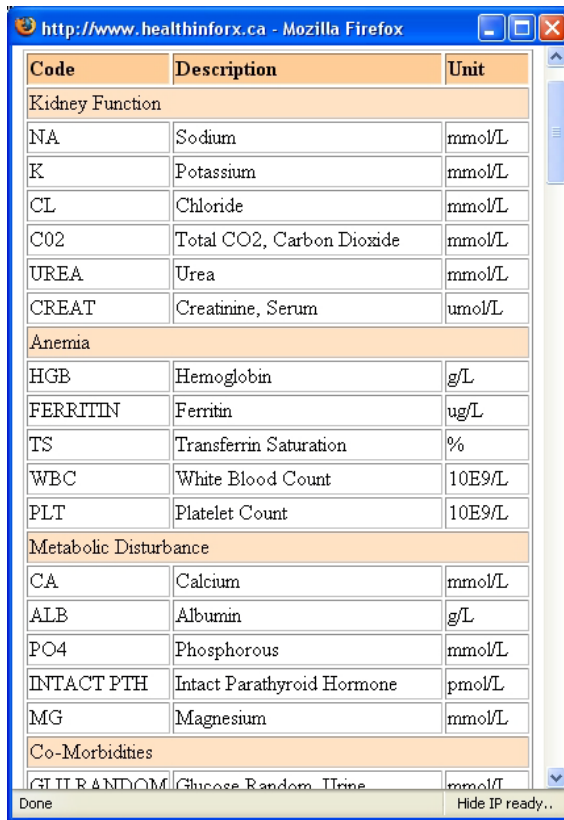
Figure 9: GFR and CKD Stages Calculations

2006 Fall Health Informatics Internship Report

Code	Description	Unit
Kidney Function		
NA	Sodium	mmol/L
K	Potassium	mmol/L
CL	Chloride	mmol/L
C02	Total CO2, Carbon Dioxide	mmol/L
UREA	Urea	mmol/L
CREAT	Creatinine, Serum	umol/L
Anemia		
HGB	Hemoglobin	g/L
FERRITIN	Ferritin	ug/L
TS	Transferrin Saturation	%
WBC	White Blood Count	10E9/L
PLT	Platelet Count	10E9/L
Metabolic Disturbance		
CA	Calcium	mmol/L
ALB	Albumin	g/L
PO4	Phosphorous	mmol/L
INTACT PTH	Intact Parathyroid Hormone	pmol/L
MG	Magnesium	mmol/L
Co-Morbidities		
GLU RANDOM	Glucose Random, Urine	mmol/L

Figure 10: Lab tests grouped by test methods

2006 Fall Health Informatics Internship Report

Figure 11: Numbers of visits for one patient from CIRESSS

2006 Fall Health Informatics Internship Report
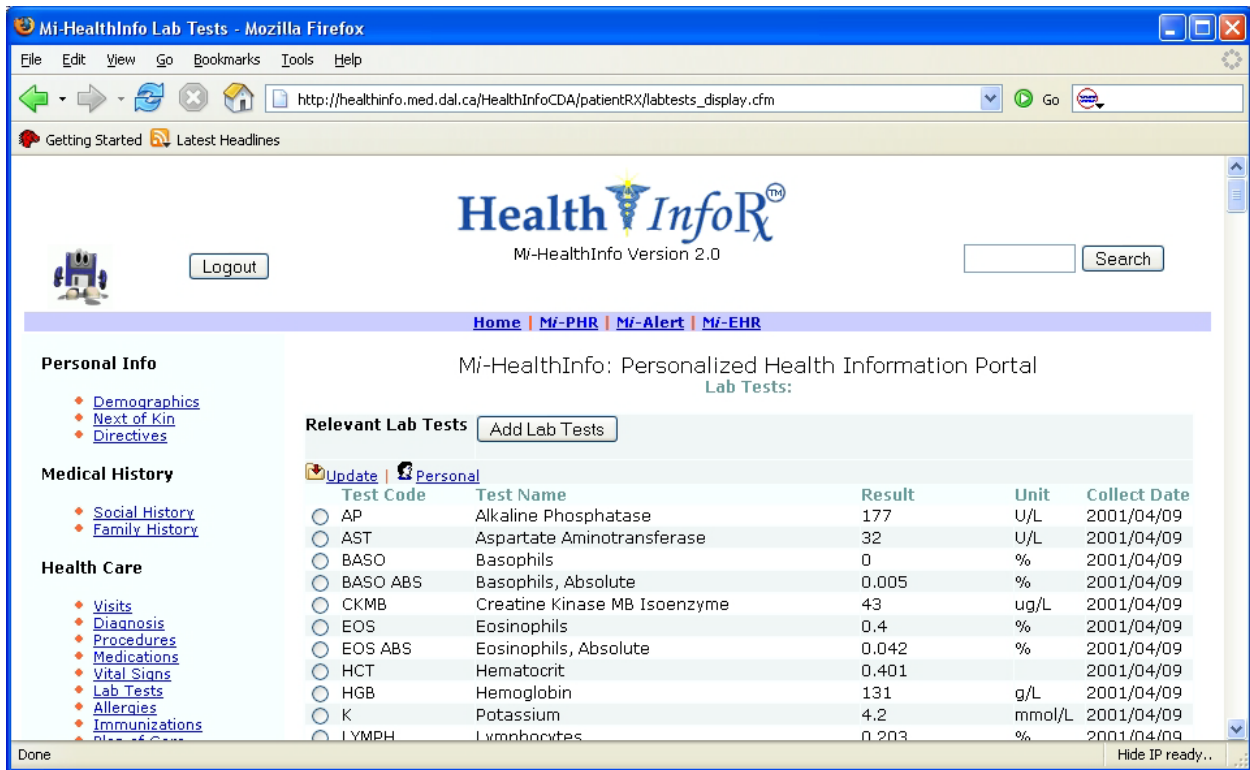
Figure 12: Lab tests results of one patient from CIRESSS in the visit from 2001/04/09 to 2001/04/2001

3.3.3 Objective and Lessons Learned

The task of building a case index gave the intern an opportunity to expose to the problem of clinical pragmatic patterns in patients' records and the semantic interoperability between the content and the information structures and explore the possible solution to this problem.

4 Health Informatics Relevance

The internship was required to mainly focus on the research for methodologies to develop a semantic indexing for 18 cases from 1007 health records from CIRESSS and one health record from CDHA. The intern has gained the understanding of how the health information could be organized and efficiently and interoperably shared among the healthcare professionals and their communities. This research was conducted through discussions with the supervisor, the author of

the research. Investigations included clinical literatures findings for methodologies to construct the semantic indexing for chronic kidney disease cases.

"The challenge is to retrieve Quebec cases that exhibit the clinical pragmatic patterns embedded in an HL7 Template designed for the Chronic Kidney Disease Discharge Summary for Nova Scotian cases. Both groups collect data differently making it difficult to determine if they mean the same thing." [3] Different coding systems are used as preference by healthcare professionals/providers to record patients' data at the time in need to provide cares to patients. There were 1007 patient records from CIRESSS in French and from CDHA in English collected for this research. Besides the gap between different coding systems and clinical pragmatic patterns, there was a difference in language of these coding systems. In order to consolidate and construct the cases, the translation of information and data in CIRESSS health records from French to English and the mapping between two coding systems using UMLS as a switching language were necessary before they could be transformed to clinical documents as boundary objects.

These boundary objects were defined as resources of health information to be shared among the different communities of practice in the CHAMP model [figure 6] [4]

Health information transferring and knowledge management among different medical communities of practice is essential. The problem-base learning cases derived from healthcare

records can help medical students acquire knowledge that is better retained, more useable in clinical context and integrated from many relevant disciplines. If a standardized terminology-architecture interface is achieved through an HL7 CDA specification, it will serve to support education as well as clinical care. Education cases that focus on chronic disease teach longitudinal record keeping [2]. The problem-base learning method would have more advantages over traditional approaches when the semantic and pragmatic interoperability in the cases is achieved. Semantic indexing for problem based learning cases could be one of the solutions for the interoperability in health information management and sharing among communities of healthcare practices.

The appendix A is a supplement for this internship in details.

5 Semantic Interoperability and Clinical Pragmatic Patterns: Problems and Solutions

5.1 Problems

In the CDHA case, there was a narration: 'Hypertension: the blood pressure was elevated during the admission in the range of 140-220 systolic, 82-110 diastolic so hytrin 2 mg and cardiazem SR 180 mg BID was added. The norvasc was discontinued. He continues with acebutolol 200 mg BID.' One of the drug names was incorrectly spelled in the narration [3].

Serum Creatinine is one of the significant markers and is used to classify the stages of chronic kidney disease. Both GFR indicators and CKD stages are considered important parameters in the case indexing to determine how serious the patient's kidney is damaged and/or needed for dialysis.

Proteinuria is determined by three methods: 24-hour urine collection for protein and creatinine clearance, random urine test for albumin:creatinine ratio (mg/mmol) and albumin level by a dipstick [5]. There was no case matched for random urine test, neither for albumin level by a dipstick because there was no lab test recorded for albumin:creatinine ratio and Urine dipstick.

Unit differences in measurement for lab test are also another pragmatic problem when searching for matching cases, for example, the expression of Total Protein was mg/L in CHDA case and g/L in CIRESSS cases. In CMAJ article, the unit for albumin:creatinine ratio was "Random urine test for albumin:creatinine ratio (mg:mmol); normally < 2.8 for women, < 2.0 for men" [5]. In the AAFP article, the random urine test for albumin:creatinine ratio is expressed as Spot urine albumin-to-creatinine ratio. Normal values Men: < 17 mg of albumin to 1 g of creatinine Women: < 25 mg of albumin to 1 g of creatinine [1].   This indicates the peril of translating between information references that use different units of measure and different lexical variations when referencing the same test method (random spot urine test).

Different versions of the International Classification of Disease (ICD) are used for coded attributes, and different languages, French and English, are used for field descriptions and free text. ICD-9 with French version was used in CIRESSS and ICD-10-CA with English in CDHA. There was also a difference between ICD-9 which was published in 1975 and ICD-9-CM (Clinical Modification) which was published in 1986 by the National Center for Health Statistics (NCHS) and includes new categories compared to the original version.

5.2 Solutions

The Medications section of the HL7 Template for Chronic Kidney Disease Discharge Summary was designed and implemented in the way that a user can surely enter the correct medication/drug name through linkage to the Nova Scotia Drug Formulary.

The serum creatinine level itself is not sufficient to measure kidney function. The GFR calculation is more accurate to measure the creatinine clearance. The CKD Discharge Summary HL7 template with the embedded GFR calculation gives the user a practical way to classify and communicate the stage of kidney failure.

A set of primary diagnostic codes for kidney disease and its co-morbidities diagnostic codes are encoded using three coding systems: SNOMEDCT, ICD-10-CA and ICD-9-CM and linkable from the CKD Discharge Summary HL7 template for user to select. The lab test codes are also encoded in the template using SNOMEDCT coding system.

The semantic indexing of cases was constructed based on a constrained set of subject terms drawn from clinical pragmatic patterns and stored in the Document Class, an HL7 Infrastructure class for Structured Documents. It supports the retrieval of cases according to clinical pragmatic patterns that are embedded in the CKD Discharge Summary template for HL7 Clinical Document Architecture (CDA). [3]

6 Conclusion

A semantic index for a problem-based learning case is a solution for semantic terminologies and clinical pragmatic patterns when bridging between EHR and CBR representations of clinical data. A semantic foundation is required for case indexing to help ensure there is a single semantic interpretation for a statement.

Finding different clinical pragmatic patterns in different dataset in a medical context can help to improve the HL7 Template design. The template captures data that is also useful in the investigation of chronic kidney disease and the management for a chronic disease. Medical educators teach about the markers of disease progression, such as the urinary albumin:creatinine ratio for staging proteinuria or the GFR indicators for kidney failure stages, and health informaticians need to ensure which data should be captured in an EHR. The boundary infostructure shows promise as a tool for learning at the boundary between different jurisdictions and different communities of practice (clinicians, health informaticians, administrators, medical educators and patients). [3]

7 Recommendation

Health Level 7 (HL7)'s Clinical Document Architecture (CDA) Version 2 is well-known and internationally adopted. It provides a set of standards supporting the exchange of health information and data through a common reference information model, a common set of data elements, a common terminology, common data structures, and a common transport standard for sharing health information among healthcare communities and their stakeholders.

The research Semantic Indexing of Patient Cases in a Boundary Infostructure for eHealth is a continuous research to solve the problem of semantic terminology and clinical pragmatic patterns for problem-based case learning for the HealthInfoCDA project. The project was to develop the HL7 Template for Chronic Kidney Disease Discharge Summary using a portal website interface to teach resident medical students. It is strongly recommended that the CKD Discharge Summary HL7 CDA Template and the portal website tool PatientRx version 2 should be used as middleware applications for teaching purpose not only in medical education but also in health informatics program to bridge the gap between the academic knowledge to the implementation of HL7 Common Terminology Service for terminology information in healthcare system.

Since the standardization version of HL7 is still growing and being adaptable, this HL7 CDA Template for Chronic Kidney Disease Discharge Summary should be updated and improved to keep it up-to-date and compliant with the standard.

References

[1] Catherine S. Snively, M.D. and CeCilia Cutierrez, M.D., University of California, San Diego, School of Medicine, La Jolla, California, Chronic Kidney Disease: Prevention and Treatment of Common Complications, http://www.aafp.org/afp/20041115/1921.html

[2] Grace I. Paterson, Syed Sibte Raza Abidi, Steven D. Soroka, HealthInfoCDA: Case Composition Using Electronic Health Record Data Sources, http://healthinfo.med.dal.ca/HealthInfoCDA/PatersonAbidiSoroka.pdf

 [3] Grace I. Paterson, MSc, Paul Fabry, MD, Andrew M. Grant, MB, PhD, Tuyet T. Thieu, BCS, Steven D. Soroka, MD, MSc, Hassan Diab, Andriy M. Moshyk, MD, MClinSci, Semantic Indexing of Patient Cases in a Boundary Infostructure for eHealth

[4] Grace I. Paterson, Steven D. Soroka, Formative Evaluation of the Clinical Pragmatic, http://informatics.medicine.dal.ca/iSHIMR2006paterson.pdf

[5] C. Stigant, L. Stevens, A. Levin, Nephrology: 4. Strategies for the Care of Adults with Chronic Kidney Disease, http://www.cmaj.ca/cgi/reprint/168/12/1553.

Appendix A

Grace I. Paterson, MSc, Paul Fabry, MD, Andrew M. Grant, MB, PhD, Tuyet T. Thieu, BCS, Steven D. Soroka, MD, MSc, Hassan Diab, Andriy M. Moshyk, MD, MClinSci, Semantic Indexing of Patient Cases in a Boundary Infostructure for eHealth