

The Generic Clinical Ontology Viewer

Jon Patrick¹, Bahram Vazirnezhad^{1,2}, Peter Budd¹, Yefeng Wang¹, Yuzhong Cheng¹

¹Health Informatics Research Laboratory, School of Information Technologies, University of Sydney, NSW, Australia

²Biomedical Engineering Department, Amirkabir University of Technology, Tehran, Iran
{jonpat; bahram; pbud3427; ywang1; yche5936}@it.usyd.edu.au

Abstract: *Clinical thesauruses, terminologies, ontologies and classifications (TTOCs) are becoming an important infrastructure for information technologies in the health sector. Recently many medical organizations have developed their own medical or clinical TTOCs and in the last few years, many tools have been developed to browse them. However the tools cannot be used to browse and visualize any other TTOC which deviates from the original in any aspects of data structure. The aim of this project has been to build a generic web-based viewer for all of the existing clinical TTOCs. Their common characteristics need to share the general properties and structures of relationships between concepts, an ordered organization of concepts in a parent-child paradigm arranged in a set of top level general categories. This generic viewer is built on a terminology server platform that provides a set of general functions which respond to requests for common information for the available TTOCs. As we receive newer versions of any of the existing TTOCs and implement them in our terminology server database they will be installed and automatically appear for use on the viewer; making access, browsing and visualizing of them possible with the various tools and search algorithms in the generic TTOC viewer.*

Keywords: Medical Ontology, Information technology

1. Introduction

With the large amounts of clinical data being generated by high-throughput analysis methods, medical vocabularies and ontologies are becoming increasingly important in modeling units of information for ease of search, retrieval and clinical notes analysis. Many medical communities have created their own locally available ontologies, and the interfaces to query these ontologies tend to vary from group to group. We saw the need for a centralized location to perform queries on medical ontologies that would offer a lightweight web-accessible user interface. This has been developed in Health Informatics Research Laboratory (HITRL), and offers a free, web-based generic viewer for scrutinizing medical thesauruses, terminologies, ontologies and classifications (TTOC) called “The Generic TTOC Viewer” accessible from the HITRL web site [1].

According to Gruber [2], an ontology is an explicit specification of a conceptualization. The term “conceptualization” is defined as an abstract, simplified view of the world, which needs to be represented for some purpose. It contains the objects, concepts, and other entities that are presumed to exist in some area of interest, and the relations that hold among them. The term “ontology” is borrowed from philosophy, where ontology is a systematic account of existence. For knowledge-based systems what “exists” is exactly that which can be (and has been) represented. TTOCs can be enriched with role relations among concepts and each concept can have various attributes related to it. Furthermore, each concept can have concept-instances attached to it, ranging from a few to thousands. In general, while medical TTOCs minimally have a hierarchy of concepts forming a taxonomic tree of classes and sub-classes, they tend to have poorly defined sub-classes and relationships with idiosyncratic cross references. But in order to provide them with a common consistent structure as defined above, we have organized the concepts in each of the TTOCs loaded in our viewer into a systematic ontological hierarchy effectively converting them into pseudo-ontologies.

There are a lot of ontology viewers for medical and clinical domain. However, almost all of them are designed to view just a specific kind of ontology without enough generality to deal with other ontologies or terminologies of different kinds. For example for SNOMED CT which is the most comprehensive ontology in the medical domain around 30 browsers are identified by authors at the time including either desktop software or web-based browsers; however none of these are capable of dealing with another clinical terminology. Sixteen of these were inspected and compared in [3]. These are CaTTS, CliniClue, CLIVE, EdBrowse, FDB Sphinx, HealthTerm, LexPlorer, Mycroft, NCI Terminology Browser, OntoBrowser, OpenKnoME, Protégé-OWL, SNOB, SnoFlake, the UMLS Rich Release Format (RRF) Browser and the Virginia Tech Browser. The most popular one is CliniClue [4]. Another

browser of this kind has been previously developed by HITRL which is a free, web-based browser specifically designed for SNOMED [5].

The contribution and ongoing aim of this project is to create and introduce a generic clinical TTOC browser which operates over a range of clinical TTOCs and effectively provides the most popular operations such as browsing, search, and visualization of the concepts regarding their parent-child relationships and retrieving all concept information.

This viewer is supported by a Terminology Server which is a database of a predefined structure which stores all TTOCs. Currently SNOMED CT, APACHE, NIC and NOC are available for viewing. In the coming months we expect to have more terminologies available through the same mechanism. We are preparing to deliver numerous editions of SNOMED CT and ICD 10AM.

2. Clinical TTOCs currently loaded

The universal facet of TTOCs that we are endeavoring to capture through this service consists of these aspects:

- All TTOCs have some minimal form of hierarchy.
- The root of the hierarchy carries the name of the TTOC.
- The uppermost level of the hierarchy beneath the root is labeled with meaningful names that can be used as categories.
- Every code or concept has a description which is retrieved by the terminology server.
- Every concept has parents and potentially children organized in a tree structure.
- Retrieval of concepts provides presentation of all the information directly related to a concept.

Except for SNOMED CT, other terminologies lack the structure of an ontology. We have converted the structure of these terminologies into pseudo-ontologies. A more mathematical definition of pseudo-ontologies, loaded into our Terminology Server is presented below, which is similar to the definition from [6]. A pseudo-ontology is a triple $O = (C, S, \text{and } IS_A)$ where:

- (1) $C = \{c_1, c_2, \dots, c_m\}$ is a set of concepts, where each concept c_i refers to a certain medical knowledge element,
- (2) $S = \{s_1, s_2, \dots, s_n\}$ is a set of slots or pieces of information attachments, where each slot s_i could refer to:
 - a. A property of a concept: a value of a simple type such as Integer, String,
 - b. A set of synonyms: popular synonyms and descriptions of each concept
- (3) $IS_A = \{IS_A_1, IS_A_2, \dots, IS_A_p\}$ is a set of inheritance relationships defined between concepts. Inheritance relationships carry subset semantics and define a partial order over concepts, organizing concepts into one or more tree structures.

2.1 Converting NIC & NOC and Apache IV to a pseudo-ontology

The process of converting NIC & NOC and Apache IV to a pseudo-ontology consists of recognizing and classifying knowledge elements in them and then organizing these knowledge elements into tree like structures. This includes adding identifying names for knowledge elements and removing ambiguous labels or altering them to be unambiguous. In Apache IV we have separated undifferentiated sub-trees.

For each concept in these terminologies Fully Specified Names (FSNs) have been constructed and multiple descriptions have been differentiated. Then we have made links between FSNs of each concept and relevant ancillary information. These set of knowledge elements have then been loaded into distinct appropriate tables in our database which are connected together in a consistent way for proper information retrieval.

2.2 Introduction to the TTOCs

The TTOCs which are currently loaded into the generic medical TTOC viewer are described below.

2.2.1 Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT)

SNOMED CT is the most comprehensive medical ontology available as an infrastructure for IT in healthcare. SNOMED CT is a systematically organized computer processable medical terminology covering most areas of clinical information such as diseases, findings, procedures, microorganisms, anatomy etc. It provides a consistent means for indexing, storing, retrieving, and aggregating clinical data across specialties and sites of care. It also helps organize the content of medical records, reducing the variability in the way data is captured, encoded and used for clinical care of patients and research. SNOMED is more than a simple medical terminology but rather it is a comprehensive ontology with more than 360000 medical concepts and over 1 million relationships [7].

2.2.2 Nursing Interventions Classification (NIC)

The Nursing Interventions Classification (NIC) is a comprehensive, research-based, standardized classification of interventions that nurses perform. It is useful for clinical documentation, communication of care across settings, integration of data across systems and settings, effectiveness research, productivity measurement, competency evaluation, reimbursement, and curricular design. The Classification includes the interventions that nurses do on behalf of patients, both independent and collaborative interventions, both direct

and indirect care. An intervention is defined as "any treatment, based upon clinical judgment and knowledge that a nurse performs to enhance patient/client outcomes." [8]. For the purposes of loading NIC into our Terminology Server we have converted it into a pseudo-ontology. The modifications we have had to make to it will be reported elsewhere.

2.2.3 Nursing Outcomes Classification (NOC)

The Nursing Outcomes Classification (NOC) is a comprehensive, standardized classification of patient/client outcomes developed to evaluate the effects of nursing interventions. Standardized outcomes are necessary for creating consistent documentation in electronic records, for the development of nursing knowledge, and the education of professional nurses. An outcome is a measurable individual, family, or community state, behaviour or perception that is measured along a continuum and is responsive to nursing interventions. The outcomes are developed for use in all settings and with all patient populations and can be used across the care continuum to follow patient outcomes throughout an illness episode or over an extended period of care. Since the outcomes describe patient status, other disciplines may find them useful for the evaluation of their interventions [8]. For the purposes of loading NOC into our Terminology Server we have converted it into a pseudo-ontology. The modifications we have had to make to it will be reported elsewhere.

2.2.4 APACHE IV

APACHE IV is a two level classification used to assess the severity of illness in Intensive Care Units. Patients are evaluated by physiologic scores and chronic health status. Results of the evaluation can be used to estimate the mortality rate for patients in the ICU and during the hospitalisation [9]. For the purpose of being able to view APACHE through the generic viewer we have had to convert it into a pseudo-ontology. This means new names have been created for each of the sub-classes beneath the 10 top level categories. Also there are some ambiguities in the assignment of specific items to class labels so that they now reside in two different classes. Apache IV is converted into a pseudo-ontology to be loaded in to Terminology Server.

3. Terminology Server

The Terminology Server is responsible for the delivery of TTOC information to any information system or service. The Terminology Server has two major components to it; the Medical Natural Language Processing (MNLP) server and the Domain Knowledge Server (DKS). The main Terminology Server acts as a

routing server, diverting function calls to the specific sub-server to handle the function.

The MNLP server is responsible for all the MNLP processing done with the contents of the TTOCS. This includes identifying TTOC concepts in free text, stemming words, performing keyword searches through TTOCs, spellchecking and combining concepts together to add depth of meaning to concepts.

The DKS is responsible for all the queries that require knowledge about the structure of the TTOCs to answer them. These queries include such things as; what are the children of this concept, what are the parents, what are the attributes, what are the common ancestors between these two concepts, what are the leaves of a subgraph, what is the subgraph, what is the top level category of this concept, what is a concept in a different TTOC that means the same thing as this one, etc. The DKS can perform aggregations of concepts, as well as groupings of similar concepts and mappings of concepts between different TTOCs.

The Terminology Server is a standalone server that systems access via a TCP network connection. Function calls are performed using a text based API that allows the Terminology Server to be used on any platform or in any programming language, as well as being able to be "bolted-on" to any existing clinical information system to allow the indexation of patient data for use in large scale analytical queries and data retrieval.

The Terminology Server has a myriad of built-in functions for the management of the TTOCs and allows updating TTOCs while still retaining access to the previous versions. The TTOCs are all version controlled so that the information and medical concepts identified using an older version are not overwritten when a new version of a TTOC is released.

4. Viewer Web Service

The generic TTOC Viewer is provided as a web-based interface between the Terminology Server and users. This web service allows for the viewing the contents of TTOCs implemented in the Terminology Server; search for the concepts of interest; retrieve all of the relevant information of a concept e.g. synonyms and descriptions, parents, children, attributes with their possible values and relationships. This web service also has the functionality of supporting navigation through the TTOC structure along the IS-A (parent-child) relationships. This is based on a JavaScript tool to walk along the concepts with IS-A relationships. The web service has been designed in a format to view all of the TTOCs through a single generic mechanism by describing the TTOCs through common structures. The web service relies on the functions of the Terminology Server to answer the search queries for each TTOCs.

The web service uses Common Gateway Interface technology to transfer information between the World Wide Web server and the CGI programs; CGI programs are written in the python programming language. There are also a number of other JavaScript functions which are mainly used to transfer information between the items and hidden fields in the query forms across various pages. Figure 1 shows the architecture of the web service and the existing links between the mechanisms.

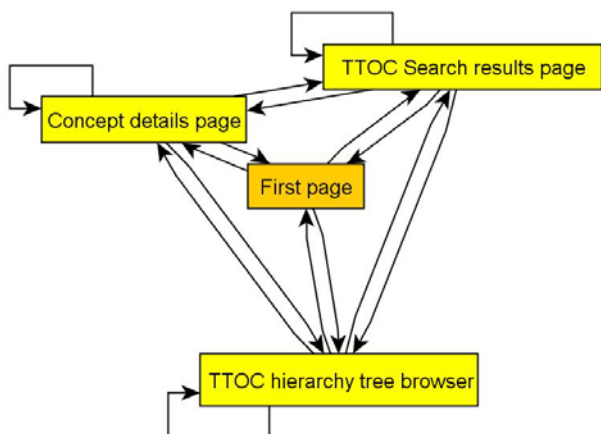


Fig. 1: The architecture layout of the web service module of the Generic TTOC Viewer

When the front page of the viewer is loaded the web service makes a query to retrieve the list of TTOCs which are currently installed in the Terminology Server. Then each TTOC will be shown with a radio button beside it. The user chooses any of the listed TTOCs then

a search can be performed using keywords or phrases which approximate the items of interest to the user. Two search criteria have been implemented: an NLP Matching criterion, and a Literal String Matching criterion. The NLP Matching criteria algorithm stems words in the input phrase and then looks for the occurrence of the stemmed words in the synonyms and descriptions of each concept to identify the nearest concepts. The Literal String Matching criterion uses the exact occurrence of the phrase in the synonyms and descriptions of the concepts to return them as the matched items.

Alternatively, in the case that a Concept ID (unique identifier) of a TTOC is known it can be used to go directly to the details page for that concept.

There are two icons on all pages which open new pages related to the root concept of the chosen TTOC. These links are especially useful for novice users, who don't have enough knowledge about the medical TTOCs and their contents to make a search using an appropriate keyword. Clicking on one of these icons (tree icon) will give a hierarchy tree with the chosen TTOC root concept at the root of the tree; and clicking on the other icon (details icon) opens the concept details page for the TTOC root concept.

If a keyword or phrase search is performed, a results page will be opened. Figure 2 is an example of search results page while searching for "Myocardial Infarction" within SNOMED CT version 20070731.

Figure 2 shows a prompt on the top of the search results page which shows the number of the returned items along with the name of TTOC which were chosen.

101 results found while looking for "myocardial infarction" in SCT version 20070731:

#	Concept ID	Full Name	Synonyms
1	113155009	Myocardial infarction education	Myocardial infarction education; Myocardial infarction education (procedure); Myocardial infarction education (regime/therapy)
2	129574000	Postoperative myocardial infarction	Postoperative myocardial infarction (disorder); Postoperative myocardial infarction
3	161503005	HIC: myocardial infarct at greater than 80	History of - myocardial infarct at age greater than 80 (situation); History of - myocardial infarct at age greater than 80; HIC: myocardial infarct at greater than 80; History of - myocardial infarct at age greater than 80 (context-dependent category); History of - myocardial infarct at greater than 80 (context-dependent category)
4	15990001	Acute myocardial infarction of posterolateral wall	Acute myocardial infarction of posterolateral wall; Acute myocardial infarction of posterolateral wall (disorder)
5	164865005	ECG: myocardial infarction	Electrocardiogram: myocardial infarction; ECG: myocardial infarction; Electrocardiogram: myocardial infarction (finding); ECG: myocardial infarction (finding)
6	164867002	ECG: old myocardial infarction	Electrocardiogram: old myocardial infarction (finding); ECG: old myocardial infarction; Electrocardiogram: old myocardial infarction; ECG: old myocardial infarction (finding)
7	164872006	ECG: myocardial infarct NOS	Electrocardiogram: myocardial infarct NOS (finding); Electrocardiogram: myocardial infarct NOS; ECG: myocardial infarct NOS; ECG: myocardial infarct NOS (finding)
8	1755008	Old myocardial infarction	Old myocardial infarction (disorder); Healed coronary; Old myocardial infarction; Healed myocardial infarction
9	194802003	True posterior myocardial infarction	True posterior myocardial infarction (disorder); True posterior wall infarction; True posterior myocardial infarction
10	194804002	Inferior myocardial infarction NOS	inferior myocardial infarction NOS (disorder); Inferior myocardial infarction NOS

Fig. 2: TTOC search results page. Clicking on the concept full names will open the concept details page and clicking on the tree icons beside full names will open the corresponding hierarchy tree with that concept in the middle of the tree.

The radio box related to the previously chosen TTOC is checked and set as the default TTOC in this page, but it can be changed anytime to make a query over another TTOC. In the search results page Concept ID, Fully Specified Name and Synonyms of the returned concepts are shown. Every page contains 10 items and users can proceed to more pages using the next/previous buttons above and below the results table. The Fully Specified Name for each concept has a link which directs users to the concept details page to get all the details related to that concept. There is also a tree icon beside each concept Fully Specified Name which opens the relevant Hierarchy tree browser with that concept in the middle of the tree; users can walk along the IS-A relationships using this hierarchy tree. This tree is a representation of Concepts regarding the IS-A Relationships which is the conventional name for the Relationship between a Parent Concept and its corresponding Child Concept.

By opening the concept details page, the information and details for the concept, the information about its synonyms and descriptions, the parents, children and relationships (if available) are retrieved from the terminology server. Figure 3 shows the concept details page for the concept “Respiratory Tract Infection” from SNOMED CT version 20070731. Some TTOCs have

more concept information in comparison with others, so that the structure and number of columns differs for concepts in different TTOCs. However a single algorithm prepares and prints out the table in a generic scheme. For example the columns *Concept Status*, *CTV3ID*, *SNOMED ID* and *Is Primitive*, in the *Concept Information* section in the figure 3, are unique for SNOMED CT. Also SNOMED CT has Relationships defined for concepts whereas concepts of other TTOCs don't have the section, *relationship*. Root concepts don't have the parent section and leaf concepts don't have a children section. The important issue regarding the algorithm is that a single generic algorithm without checking the position of the concept in the TTOC and only by retrieving and parsing the response to a request for information from the Terminology Server constructs the appropriate structure for the presentation table and presents the information entries in it.

There are links across the concept detail page for all of the parents, children and concepts with a relationship with the current concept where clicking on them will get a similar details page for those concepts. Alternatively by clicking on the tree icon beside these links a user can get the hierarchical IS-A tree browser for these concepts in a new window.

 "Respiratory tract infection" from SCT version 20070731

Concept Information										
Concept ID	Text	Top Parent	Number of Parents	Number of Children	Hierarchy Level	Concept Status	CTV3ID	SNOMEDID	IsPrimitive	
275498002	Respiratory tract infection (disorder)	Clinical finding	2	10	5:6	0	XM1QW	D2-0009A	0	
Synonyms and Descriptions										
#	Description ID	Concept ID	Text	Description Type	Description Status	Initial/Capital Status	Language Code			
1	668374017	275498002	Respiratory tract infection (disorder)	3	0	0	en			
2	411488017	275498002	Respiratory tract infection	1	0	0	en			
3	411489013	275498002	RTI - Respiratory tract infection	2	0	1	en			
Parent Concepts Information										
#	Concept ID	Text	Top Parent	Number of Parents	Number of Children	Hierarchy Level	Concept Status	CTV3ID	SNOMEDID	IsPrimitive
1	50043002	Disorder of respiratory system (disorder)	404684003	2	47	4:5	0	H...	D2-00000	0
2	301810000	Infection by site (disorder)	404684003	2	62	4	0	Xa85n	DE-00017	0
Child Concepts Information										
#	Concept ID	Text	Top Parent	Number of Parents	Number of Children	Hierarchy Level	Concept Status	CTV3ID	SNOMEDID	IsPrimitive
1	50417007	Lower respiratory tract infection (disorder)	404684003	5	11	5:6:7	0	X1004	D2-00100	0
2	312149008	Fungal respiratory infection (disorder)	404684003	2	6	6:7	0	XaDcm	DE-4001E	0
3	371044001	Transient respiratory distress with sepsis (disorder)	404684003	2	0	5:6:7	0	XUUHJ	D2-00116	1
4	312133006	Viral respiratory infection (disorder)	404684003	2	14	6:7	0	XaDcT	DE-300A5	0
5	195647007	Acute respiratory infections (disorder)	404684003	3	4	5:6:7	0	H0...	D2-0004E	0
6	312117008	Bacterial respiratory infection (disorder)	404684003	2	8	6:7	0	XaDcB	DE-1004E	0
7	54150009	Upper respiratory infection (disorder)	404684003	2	14	6:7:8	0	Xa2o8	D2-00020	0
8	58554001	Empyema of pleura (disorder)	404684003	3	8	6:7:8	0	XUH0u	D2-80500	0
9	267284008	Obstetric pyemic and septic pulmonary embolism (disorder)	404684003	3	6	6:7:8:9:10:11:12	0	XE0xM	D8-40047	1
10	314042000	Infective pleurisy (disorder)	404684003	2	4	6:7:8:9	0	XaEW2	D2-00101	0
Relationships										
#	Target Concept ID	Relationship Type	Concept Set Values							
1	Structure of respiratory system	Finding site	{Respiratory system subdivision; Entire respiratory system; Pulmonary vascular structure; Respiratory tract structure; Gill; Lower respiratory system structure}							
2	Infectious agent	Causative agent	{Fungus; Arthropod-borne organism; Bacteria; Parasite; Prion; Virus; Opportunistic infectious agent; Opportunistic Infectious agent}							
3	Severities	Severity	{Mild; Moderate; Mild to moderate; Moderate to severe; Fatal; Severe}							
4	Episodicities	Episodicity	{First episode; Old episode; New episode; Ongoing episode; Other episode RCGP; Undefined episodicity}							

Fig. 3: The concept details page for the concept “Respiratory Tract Infection” from SNOMED CT.

As shown in figure 3, the *Concepts information* section of the table contains: the *Concept ID*, *Fully Specified Name*, *Top Category* which is the first level general parent of this concept and an immediate child of the root concept; number of parents and children; *hierarchy level(s)* which are levels that the concept is located on; *Concept Status* which is a field in the concepts table that specifies whether a concept is currently in use or not; *CTV3ID* which identifies the concept in the equivalent Read code; *SNOMED ID* is a field in the concepts table which contains the SNOMED RT identifier for the concept, and *Is Primitive* is a field in the concepts table that indicates whether a concept is Primitive or Fully Defined. If a concept is connected to the root via multiple paths there are multiple hierarchy levels separated by colons.

The *Synonyms and Descriptions* section in the table contains *Description ID*; related *Concept ID*; *Name*; *Description Type* which is a field in the description table that specifies whether a description is a fully specified name, preferred term, or synonym; *Description Type* which is a field in the descriptions table that specifies whether a description is in current use or not; *Initial Capital Status* which specifies whether the capitalization of the first character is significant, and *Language Code* which indicates the language and, optionally, dialect applicable to a row in the subset table, descriptions table based on ISO639-1 language coding.

The *Parent Concepts Information* and *Children Concepts Information* sections in the table contain the concept information of all of the immediate parents and children of this concept respectively.

The *Relationships* part contains all of the concepts with a relationship other than the IS-A type of relation; for example as it is shown in the figure 3, *Respiratory tract infection* is connected by the *finding site* to the *Structure of respiratory system* and the values for this attribute are shown in the *Concept values set* in the last column. Other attributes for *Respiratory tract infection* are *Causative agent*; *Severity* and *Episodicity* have possible values of each of these attributes are shown in the last column again.

The other visualization tool in the Generic TTOC viewer is the Hierarchical tree which is a browser tool to navigate along IS-A relationships of the organization of concepts. General concepts are at the top of the hierarchy tree. At each level down the hierarchy, concepts become increasingly specialized. In each TTOC concepts are arranged into Top-level hierarchies which are the most general concepts in the TTOCs. Each of these hierarchies subdivides into smaller sub-hierarchies. More specific concepts are related by IS-A relationships to their more general parents directly above them in the hierarchy. There is one concept from which the Top-level hierarchies descend called the Root concept. Each concept on the tree is shown by its Fully Specified Name

and Concept ID and clicking on the name will open the concept details page of this item in a new window. Clicking on the tree icon beside each name will construct a new tree with the parents of it on the first layer and this concept and other concepts originating from the same or common parents on the middle layer and children of all of these middle ones on the last layer branching from their own general concept. Figure 4 shows the hierarchical tree browser opened around the concept *Myocardial infarction* from SNOMED CT the concepts in other TTOCs can be viewed with trough the same mechanism.

You are now in level(s) 5:7:8:9:10 down from the Root.

IS A tree for "Myocardial infarction" from SCT version 20070731

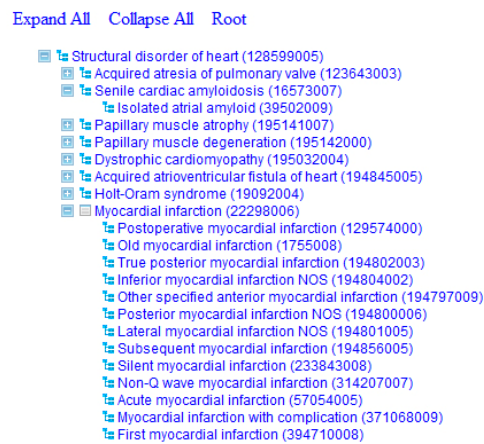


Fig. 4: The hierarchical tree browser opened around the *Myocardial infarction* from SNOMED CT.

As shown in figure 4 each of branches in the tree can be opened and closed by clicking on the plus (minus) icon beside each branch. The **Expand All** and **Collapse All** functions above the tree can open and close all of the branches respectively at the same time. Clicking on the *Root* on the top of the tree will direct the user to the root concept of the current TTOC.

5. Conclusions

In the recent years the need for automatic manipulation of the clinical data, especially clinical notes, produced by high throughput analysis systems is becoming increasingly important. Clinical thesauruses, terminologies, ontologies and classifications, are necessary to provide a consistent way to index clinical notes, for a proper storage, retrieval, communication between information systems and aggregating clinical data across specialties and sites of care. To avoid repeating the long terms, thesauruses, terminologies, ontologies and classifications, we adopted the mnemonic TTOC to capture all of these terms. Recently many medical organizations have developed their own medical

or clinical TTOCs and in the last few years, many tools have been developed to browse and visualize them, however their software cannot be used to browse and visualize any other TTOC which deviates from the original in aspects of purpose, structure and size.

So hence we saw the need for a generic viewer capable of browsing all of the existing clinical TTOCs; and implemented as a free web service to make a platform for uniform browsing of the clinical TTOCs. In this way comparison between clinical TTOCs would be possible having the same mechanism for making queries. The generic viewer can clarify the commonalities and differences between the clinical TTOCs. As we gain the newer versions of the TTOCs and implement them in our terminology server database they will automatically appear for use on the viewer; making access, browsing and visualizing of them possible with the various tools and search algorithms in the generic TTOC viewer.

6. Future Work

The generic Medical TTOC Viewer is operational but there are many enhancements and functionalities which are under development and will be released in future versions. These enhancements can be categorized into visualization functionalities, search algorithm, TTOC comparison, editing and assembling subsets for the TTOCs.

The current TTOC viewer has limited presentation functionality and in particular is poor at showing multiple pathways through a network to a given node. Building a visualization tool like a hyperbolic tree in order to browse the concepts in a user friendly platform along with their relationships to find out the surrounding concepts and providing a better understanding of the TTOCs is one of our future enhancements. Embedding all of the tools and visualization facilities in a single page mechanism in order to avoid opening multiple pages and make it possible to study the contents in a single glance and browse the TTOCS in a single page is also other future work.

Enhancements for the search algorithm would be to have a more sophisticated algorithm for search which takes into account the hierarchy and thus the semantic content; and also developing an advanced search function with the possibility of restricting the search within a subset of top level categories of each TTOC or to a limited depth.

In terms of TTOC comparison; developing a platform to compare the contents, structures and size of each of the TTOCs in order to find out their commonalities and differences would be worthwhile. Making an alignment tool between the existing TTOCs in order to find out the level of similarity in various parts (categories) of TTOCs is under investigation. A mapping mechanism between the TTOCs is another function where for example we can

find the nearest concepts of SNOMED to a notionally similar concept in APACHE.

For editing the TTOCs a mechanism to import concepts from other TTOCs into a specific TTOC is needed. Making a mechanism to define and build subsets, adding new concepts and editing or adding new descriptions (in any language using Unicode standards), or defining new relationships is a tool under development. This tool should have the ability to undo and also show the history of the modifications. Building a mechanism such as a change log to register the history of the changes which have been made would be an important future functionality. Integration with Protégé so as to move files between the Terminology Server and Protégé to support ontology verification is also under investigation. This implies creating an OWL output format. Other additional functionality would be adding a mechanism to tag favourite items or issues which each user is interested in. Then on the next login each user can retrieve his/ her favourite items. Building an interface to post comments of users in order to exchange ideas and interests and making connections between users who have similar interests can be useful in our future implementations.

Instances of concepts drawn from actual data in clinical information systems is of interest to the user, much in the way examples of language usage are presented in dictionaries. However, representing instances as nodes connected to a class is not always effective because of their great number, so other alternatives should be used, like presenting the instances of a selected class as a list within a separate window. This would be valuable especially for educational purposes and further developing the TTOCs themselves.

References

- [1] Health Informatics Research Laboratory, School of Information Technologies, University of Sydney, www.it.usyd.edu.au/~hitru
- [2] Gruber T. "A translation approach to portable ontology specifications, knowledge acquisition," *Special issue: Current Issues in Knowledge Modelling*, Vol 5, Issue 2, 199–220, 1993.
- [3] Rogers J. and Bodenreider O., "SNOMED CT: Browsing the Browsers," In proceedings of KR-MED 2008, Arizona, USA, 2008.
- [4] CliniClue (build 2006.02.30) www.cliniclue.com
- [5] SNOMED CT Terminology Server V2, <http://www.it.usyd.edu.au/~hitru/sct/A1.cgi>
- [6] Amann, B. and Fundulaki, I. "Integrating ontologies and thesauri to build RDF schemas," In proceedings of the Third European Conference on Research and Advanced Technology for Digital Libraries. pp. 234–253, 1999.
- [7] SNOMED CT. IHTSDO, Copenhagen 2007 www.ihtsdo.org
- [8] Center for Nursing Classification & Clinical Effectiveness http://www.nursing.uiowa.edu/excellence/nursing_knowledge
- [9] Knaus W., Draper E., Wagner D. and Zimmerman J. "APACHE II: a severity of disease classification system," *Critical Care Medicine*, vol. 13, pp. 818–829, 1985.