Developing a Search Algorithm and a Visualization Tool for SNOMED CT

Anthony Masi, Ankur Agrawal
Department of Computer Science, Manhattan College
Riverdale, New York, USA

Abstract—**With electronic health records rising in popularity among hospitals and physicians, the SNOMED CT medical terminology has served as a valuable standard for those looking to exchange a variety of information linked to clinical knowledge bases, information retrieval, and data aggregation. However, SNOMED CT is distributed as a flat file database by the International Health Terminology Standards Development Organization and visualization of data can be a problem. This study describes an algorithm that allows a user to easily search SNOMED CT for identical or partial matches utilizing indexing and wildcard matching through a graphical user interface developed in the cross-platform programming language Java. In addition to this, the algorithm displays corresponding relationships and other relevant information pertaining to the search term. The outcome of this study can serve as a useful visualization tool for those looking to delve into the increasingly standardized world of electronic health records as well as a tool for healthcare providers who may be seeking specific clinical information contained in the SNOMED CT database.**

Keywords—SNOMED CT; electronic; health records; visualization; search algorithm; GUI

I. Introduction

Systematized Nomenclature of Medicine Clinical Terms, commonly referred to by the acronym SNOMED CT, is a clinical healthcare terminology which serves as a foundation for the construction and development of electronic health records [1]. SNOMED CT is owned, developed, and distributed by the International Health Terminology Standards Development Organization (IHTSDO), a not-for-profit organization administrated by 27 countries and growing [2, 3].

The increasing popularity of the SNOMED CT terminology can be attributed to some of the general benefits that come with being maintained by the international community. It is an international standard, benefitting from multilingual support, and allowing the foundation and construction of electronic health records to take place across the globe [4]. In addition to this, SNOMED CT provides a standardized terminology spanning the various health care domains. Standardization allows for medical professionals to communicate along a mutually understood channel, eliminating ambiguity and redundancy among patients' health records [5]. Redundancy is reduced by aggregating the synonyms of a preferred term and pointing them at a universal concept identification code. The inherent structure of the terminology as a whole is highly scalable, allowing it to be extended effortlessly.

SNOMED CT serves as a valuable tool for clinical use as well as capturing medical information at the appropriate level for its application in the healthcare sector. Additionally, data can be transcribed by multiple physicians to the same patient thus allowing for consistency across multiple healthcare platforms. Through this consistency, SNOMED CT can ensure error-free maintenance of electronic health records by increasing the chances of correct interpretation of data. Standardizing the transcription of electronic health records make them a more viable option for portability, as a set of terms utilized in one office can be interpreted accurately by another healthcare provider utilizing SNOMED CT's unique concept identification number. This convention allows for the swift searching of clinical information, linking a single concept code to multiple synonyms as well as concept codes directly linked in the concept table hierarchy of SNOMED CT database. Overall, the terminology improves efficiency among healthcare professionals and enables a time-proof history of medical information to be kept. The benefits that SNOMED CT provides to the medical community warrant a user-friendly interface to harness its full power.

SNOMED CT is continuously evolving with a new version released every six months by IHTSDO in January and in July. The latest SNOMED CT release (January 2015) is available in an International format as a zip compressed archive [6]. Upon decompression, the contents of the SNOMED CT terminology are spread out across a span of various directories of text files. Considering the immense size of the terminology, this distribution is inherently useless as is, freezing Windows 7's default text editor Notepad upon access. As a viable alternative, SNOMED CT specific browsers [7] are needed to explore the contents of the terminology. One such browser is the stand-alone application CliniClue Xplore [8] developed by Clinical Information Consultancy. However, according to the CliniClue web page, this distribution has been discontinued and support is no longer provided. With the discontinuation of CliniClue Xplore creating a void in the field of SNOMED CT compatible desktop browsers, this study intended to create a viable second option for end-users seeking a graphical user interface in order to browse and manage the SNOMED CT terminology.

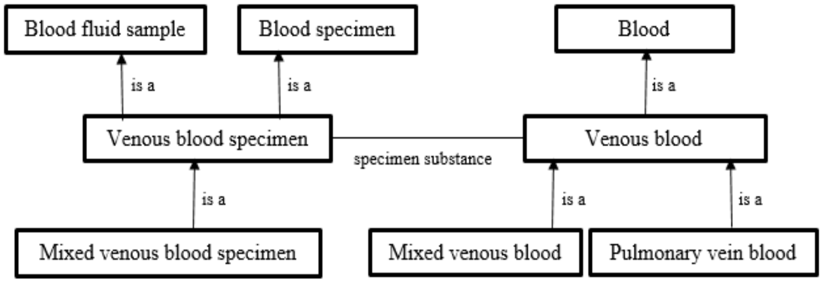


Fig.1. Various relationships of venous blood specimen concept

II. Background

SNOMED CT is a large and comprehensive clinical health terminology providing a standardized way to represent clinical concepts by different healthcare providers for the sake of interoperability [9]. The content of SNOMED CT is represented using three different components – concepts, descriptions and relationships. A concept represents a clinical entity of some kind and has a unique concept identifier. For example, *venous blood specimen* is a concept in SNOMED CT and is uniquely identified by its concept identifier as 122555007. SNOMED CT consists of over 200,000 such active clinical concepts organized into 19 hierarchies. The root concepts of these 19 hierarchies are shown in Table I.

A description represents a concept in a human readable form and has a unique description identifier. A description could be a preferred term, a fully specified name or a synonym. For instance, the concept *venous blood specimen* is a preferred term, *venous blood specimen (specimen)* is its fully specified name and *venous blood sample* is its synonym. The term within the parenthesis in a fully specified name is known as a semantic tag. In case of the concept, *venous blood specimen*, its semantic tag is specimen.

TABLE.I. SNOMED CT HIERARCHIES

Body structure
Clinical finding
Context-dependent category
Environments and geographical locations
Event
Linkage concept
Observable entity
Organism
Pharmaceutical/biologic product
Physical force
Physical object
Procedure
Qualifier value
Record artifact
Social concept
Special concept
Specimen
Staging and scales
Substance

A relationship represents a link of some kind between two concepts and is represented by a unique relationship identifier. A relationship could be hierarchical relationship (also known as parent-child or *is a* relationship) or lateral relationship (also known as attribute relationship). All concepts in SNOMED CT except for the overall root concept consist of at least one *is a* relationship. Figure 1 demonstrates one such concept with all of its relationships. As can be seen in Figure 1, the concept *venous blood specimen* from the *specimen* hierarchy has two parents – *blood fluid sample* and *blood specimen*. The concept is connected to its two parents using the *is a* relationship. The concept also has a child concept called *mixed venous blood specimen* and the two are also connected using the *is a* relationship. The concept is further connected to a concept *venous blood* from the *substance* hierarchy using a lateral relationship called specimen substance. The concept *venous blood* has a parent concept *blood* and two children concepts – *mixed venous blood* and *pulmonary vein blood*.

SNOMED CT also uses role-groups to organize the attributes and their values to create specific associations. For example, the concept *intraoperative transluminal femoral-popliteal angioplasty* has two role groups as shown in Table II. The concept can thus have two *procedure sites – direct* attributes with target values of *structure of femoral artery* and *structure of popliteal arter*y. Each of these two procedure sites then has its own procedure method and surgical approach.

TABLE.II. ROLE GROUPS OF THE CONCEPT INTRAOPERATIVE TRANSLUMINAL FEMORAL-POPLITEAL ARTERY

Group 1	**Group 2**
Procedure site Direct structure of femoral artery	Procedure site Direct structure of popliteal artery
Method Dilation repair - action	Method Dilation repair - action
Surgical approach Transluminal approach	Surgical approach Transluminal approach

SNOMED CT concepts are further defined as being either primitive or fully defined. A concept is primitive if its defining characteristics are not sufficient to uniquely distinguish it from other similar concepts. For example, the two concepts *disease outbreak* and *overexertion* are primitive concepts since both concepts have an *ISA* relationship to the *event* concept with no other defining relationships to distinguish them from each other and from other similar concepts.

III. Method

Utilizing Java's Swing API [10], a form is created in order to contain the various controls necessary to interact with the backend database. This database is designed in Oracle's MySQL relational database management system [11]. The database consists of three tables: concept, description, and relationship. These tables are populated through a locally run MySQL script, using the flat text files supplied in the standard

SNOMED CT releases. Java's packaged SQL library is utilized to connect the frontend graphical user interface to this locally hosted database.

With the database of concepts, descriptions and relationships now accessible through a graphical user interface, the user is able to harness the full power of SNOMED CT terminology. Features of the front end include a find-as-you-type concept lookup algorithm, multiple search filters, and grouping based upon term relationships.

The find-as-you-type concept lookup algorithm is designed with convenience in mind. The textbox utilized as the search field is bound to two events, MouseClicked and KeyReleased. The MouseClicked event serves the purpose of clearing the search field when focus is acquired, allowing the user a clean slate in the case of multiple searches during one session. The KeyReleased event contains the heart of the algorithm, firing in the cases of: a) the textbox's value is greater than 0 and b) the last key pressed was not the backspace or delete key. This serves the purpose of avoiding unnecessary queries to the database while also speeding up the process of data retrieval.

In the case that the event does fire successfully, the algorithm then takes into account the search filter selection. There are a total of six possible filter combinations, choosing one option from each of the provided two combo boxes. The first combo box contains two options for the search method, exact match or partial match. The second combo box limits the amount of results to the top 5, 10, or all results. In the case of the partial match method, the algorithm branches down the path of querying the database utilizing MySQL's LIKE operator to determine possible result sets.

The initial result set returned contains possible term matches from the *description* table along with their classification in the SNOMED CT nomenclature. Terms are only returned if they are deemed non-deprecated in the *description* table. Terms can be a member of one of three subgroups: preferred terms, fully specified names, or synonyms. The algorithm parses the *description type* column from the *description* table in order to prepend a tag to each term in the output box of the "Concept Lookup" pane. These possible tags are shorthand denominations of the preceding subgroups and are labeled [FSN], [SYN], and [PREF] for fully specified names, synonyms, and preferred terms respectively.

At this point the algorithm awaits user input, idling until a *Mouse Clicked* event is detected in the provided list box reserved for search results or until a new search is performed. When the *Mouse Clicked* event is triggered, a more in depth search is executed on the selected term. This search involves parsing both the selected term and it's designation as a fully specified name, synonym or preferred name, returning a full set of clinical information from the database. This new result set includes the preferred term of the input, the fully specified name according to SNOMED CT, all possible synonyms, the SNOMED CT concept identification code, and a set of concept relationships.

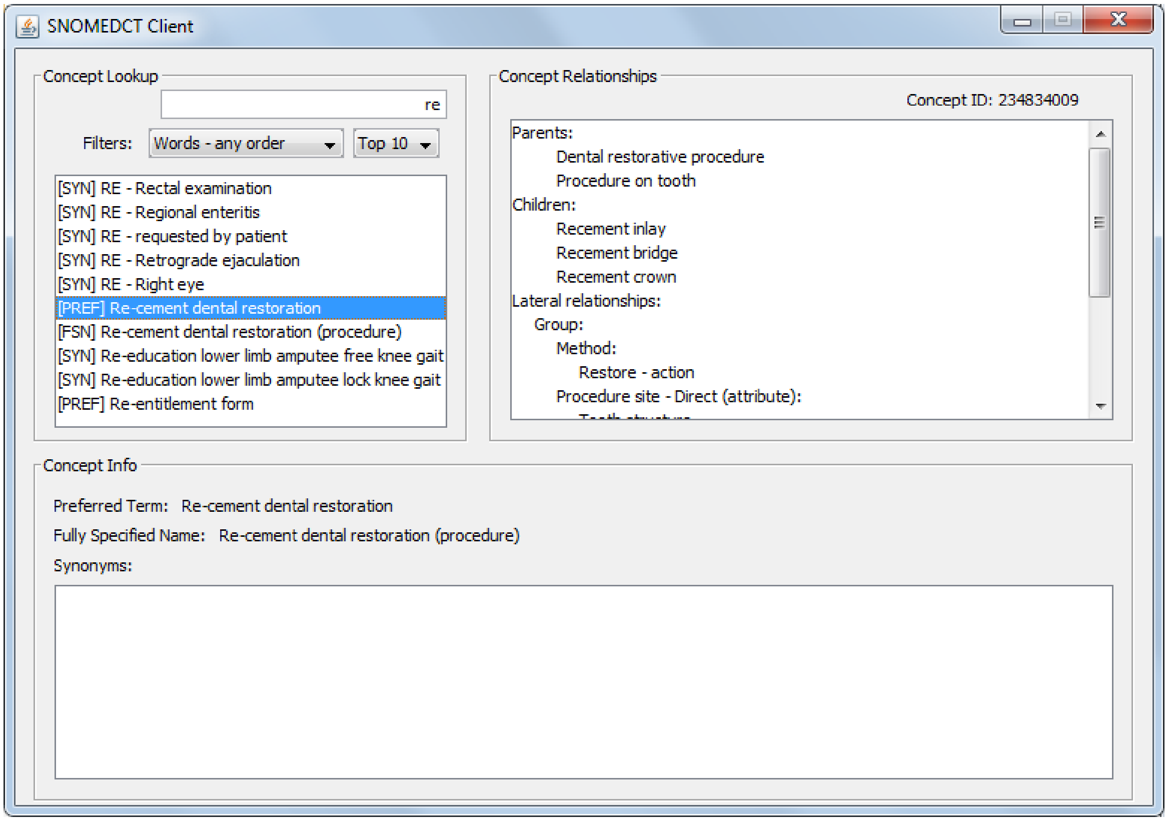
Concept relationships include parent-child relationships, also referred to as *is a* relationships, and lateral relationships, also referred to as attribute relationships. The algorithm groups these terms by their relationship to the input, subgrouping lateral relationships into unique sets as determined by their relationship role group value. Once execution of the algorithm is completed, the output that is presented in the user interface includes the searched term, which will be referred to as the input, the preferred term for the input, the fully specified name of the input, synonyms of the input, its concept identification number, and its parent, child, and lateral relationships to other terms in the SNOMED CT database.

IV. Results

The version of SNOMED CT used in this study is the January 2014 International Release, RF1. In this version there are 300,277 active concepts utilizing 783,757 active descriptions linked by 1,469,762 concept relationships**.** Figure 2 displays a screenshot of the user interface that implements the search algorithm discussed in the Methods section.

The graphical user interface as shown in Figure 2 is based upon three content frames. The upper left content pane is used to query the database utilizing a search box and a drop down filter. Another filter is present to establish an upper bound on the amount of search results. Results are outputted in the form of a clickable list to the central text area of the pane. The upper right pane is reserved for concept relationships. Populated upon selection of a term from the "Concept Lookup" pane, the concept identification number, parent terms, child terms, and lateral relationships are displayed. Lateral relationships are designed to establish a connection between terms that do not possess a parent-child relationship with each other. These relationships can be grouped together as defined in the SNOMED CT terminology. The bottom pane lists synonyms of the term selected by the user. In addition to these basic synonyms, the preferred term and fully specified name are also displayed in this bottom pane.

The screenshot in Figure 2 uses partial match search to find the preferred term *re-cement dental restoration*. Since the search term is "re", the algorithm displays all concepts with "re" as a part of their description. On the other hand, the screenshot in Figure 3 displays the identical term search filter, which returns results that match user input character for character. In this example, a search for the identical term for



SNOMEDCT Client

Concept Lookup

re

Filters: Words - any order | Top 10

[SYN] RE - Rectal examination
[SYN] RE - Regional enteritis
[SYN] RE - requested by patient
[SYN] RE - Retrograde ejaculation
[SYN] RE - Right eye
[PREF] Re-cement dental restoration
[FSN] Re-cement dental restoration (procedure)
[SYN] Re-education lower limb amputee free knee gait
[SYN] Re-education lower limb amputee lock knee gait
[PREF] Re-entitlement form

Concept Relationships

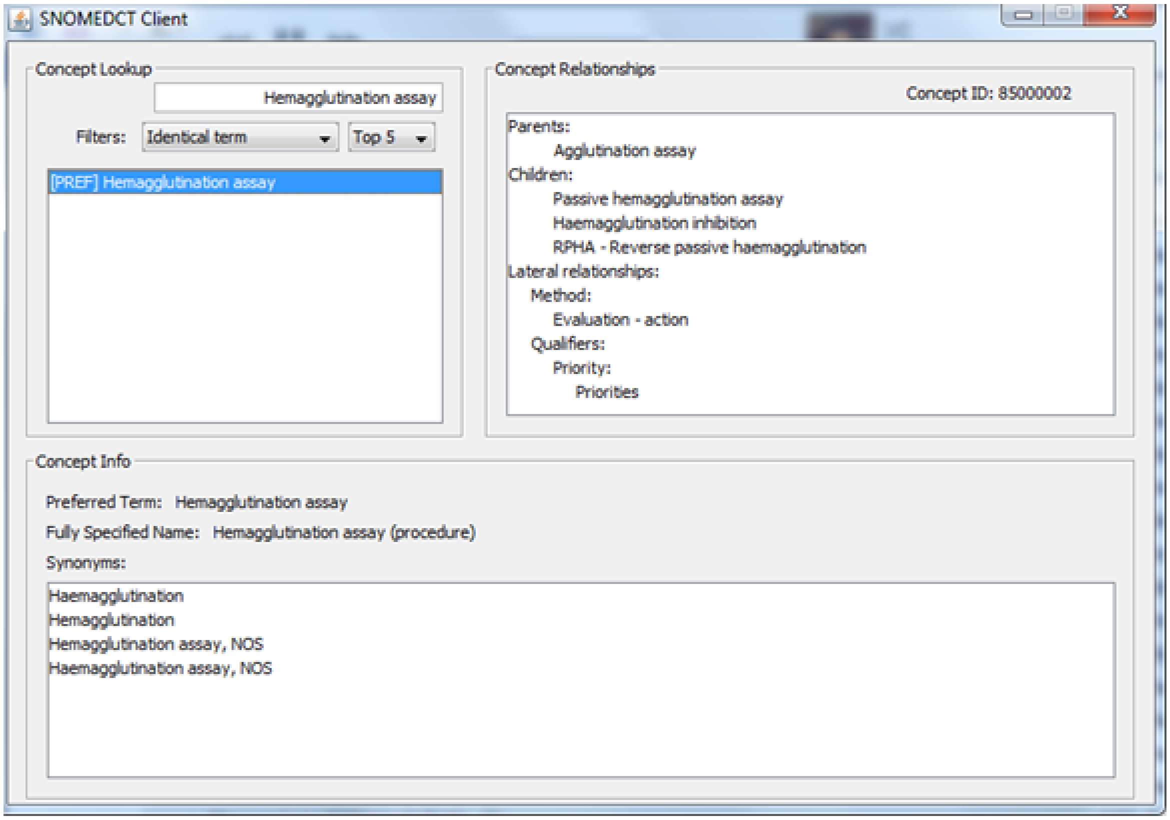
Concept ID: 234834009

Parents:
Dental restorative procedure
Procedure on tooth
Children:
Recement inlay
Recement bridge
Recement crown
Lateral relationships:
Group:
Method:
Restore - action
Procedure site - Direct (attribute):
Tooth structure

Concept Info

Preferred Term: Re-cement dental restoration
Fully Specified Name: Re-cement dental restoration (procedure)
Synonyms:

Fig.2. Screenshot of the user interface with the words of the searched concept in any order



SNOMEDCT Client

Concept Lookup

Hemagglutination assay

Filters: Identical term | Top 5

[PREF] Hemagglutination assay

Concept Relationships

Concept ID: 85000002

Parents:
Agglutination assay
Children:
Passive hemagglutination assay
Haemagglutination inhibition
RPHA - Reverse passive haemagglutination
Lateral relationships:
Method:
Evaluation - action
Qualifiers:
Priority:
Priorities

Concept Info

Preferred Term: Hemagglutination assay
Fully Specified Name: Hemagglutination assay (procedure)
Synonyms:
Haemagglutination
Hemagglutination
Hemagglutination assay, NOS
Haemagglutination assay, NOS

Fig.3. Screenshot of the user interface with exact match

hemagglutination assay displays one concept with the exact same concept description.

V. Discussion

SNOMED CT is a large and comprehensive, multilingual clinical reference terminology that has a wide array of applications, the most important being it's use as a reference terminology in electronic health records. The data for SNOMED CT is provided in the form of flat files by IHTSDO. So it becomes necessary to have a tool that can be used to view the content of SNOMED CT with ease.

This study aimed at devising efficient algorithms that can be used to navigate through the contents of SNOMED CT. We developed a fast and efficient algorithm that can be used to search for any concept in SNOMED CT using both partial and exact match as discussed in the Results section. The concept that the user searches for is displayed along with its several properties including the parents and children of the concept, lateral relationships, information about role groups, the preferred term, fully specified name and synonyms if any. We also developed an intuitive and user friendly visualization tool in the form of a graphical user interface to display the results of the search algorithm. This interface can be used to get a clear and complete picture of a SNOMED CT concept. This will be a useful visualization tool for clinicians who use SNOMED CT to code patient data and for researchers who are working in the field of clinical informatics using SNOMED CT to fetch clinical data.

Future work will involve a study to bring the application online so that it can be accessed from anywhere with an internet connection. We will also bring in an ability to browse through the different releases of SNOMED CT as SNOMED CT is an evolving terminology with a new release every six months, in January and in July. We will also work on an algorithm that will offer the ability to make more complicated search queries such as being able to query the system to find all concepts that has the word heart in it, has at least two parents and has at most three role groups. Such queries can be a useful tool for researchers who are using SNOMED CT as their test bed.

VI. Conclusion

SNOMED CT is an important and widely used clinical reference terminology which has been gaining a worldwide acceptance and popularity for its use in electronic health records. SNOMED CT is also widely used by researchers in the medical informatics field as a test bed for various clinical applications trial which often requires them to look into the attributes and other properties of a concept. A SNOMED CT visualization tool such as the one presented in this study can be an important aid to the users of SNOMED CT by helping them navigate through its content with ease.

References

[1] SNOMED Clinical Terms; [cited 1 January 2015]. Available from: http://www.ihtsdo.org/snomed-ct.

[2] International Health Terminology Standards Development Organisation; [cited 1 January 2015]. Available from: http://www.ihtsdo.org.

[3] Members of IHTSDO [cited 1 January 2015]. Available from: http://www.ihtsdo.org/members.

[4] SNOMED CT Benefits; [cited 1 January 2015]. Available from: http://www.ihtsdo.org/snomed-ct/why-should-i-get-snomed-ct.

[5] Dougherty M. Standard terminology helps advance EHR. J AHIMA. 2003;74(10):59-60.

[6] SNOMED CT International Release Files; [cited 1 January 2015]. Available from: http://www.nlm.nih.gov/research/umls/licensedcontent/snomedctfiles.html.

[7] SNOMED CT Browsers [cited 1 January 2015]. Available from: http://ihtsdo.org/fileadmin/user_upload/doc/browsers/browsers.html.

[8] CliniClue Xplore: The Clinical Information Consultancy Ltd; 2011 [cited 2015 1 January]. Available from: http://www.cliniclue.com.

[9] SNOMED CT Starter Guide [cited 1 January 2015]. Available from: http://ihtsdo.org/fileadmin/user_upload/doc/download/doc_StarterGuide_Current-en-US_INT_20140222.pdf.

[10] Java Swing API [cited 1 January 2015]. Available from: http://docs.oracle.com/javase/8/docs/technotes/guides/swing.

[11] MySql [cited 1 January 2015]. Available from: http://www.mysql.com.