MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190546

Implementation of a Terminology Server with SNOMED CT in Graph Databases

Elizabeth Silva Layes[a], Marcelo Bondarenco[a], Daniel Machiavello[a], Fabián Frola[b], Martín Lemos[b]

[a] *Health Informatics Department, Sociedad Médico Quirúrgica de Salto, Salto, Uruguay,*
[b] *Development Department, Sociedad Médico Quirúrgica de Salto, Salto, Uruguay*

Abstract

This paper described the implementation of a terminology server performed at a health institution in Uruguay, whose architecture is based on SNOMED CT using graph databases (NoSQL). The aim of this development was to create an intuitive terminological service, making the most of SNOMED CT's ontology, and which can be used from a clinical, statistical, management, decision support and research point of view, among others, with good performance.

Keywords:

Systematized Nomenclature of Medicine; Vocabulary, Controlled; Databases, Factual

Introduction

Terminological services (TS) are essential components in the development of an Electronic Health Record (EHR), since their implementation strengthens the semantic quality of the data recorded during the care act. These services are fundamental tools for the management of controlled vocabularies in different instances of clinical events, whether for use in statistical data, management, decision support and research, or other areas. SNOMED CT is a global clinical terminology that covers a wide range of specialties, disciplines, and medical and care requirements. With SNOMED CT, clinical information is recorded with identifiers that refer to formally defined concepts as part of the terminology [1]. SNOMED CT allows the recording of clinical information with appropriate levels of detail through the use of relevant clinical concepts. This terminology structure allows information to be entered using synonyms that adapt to local preferences, while recording information in a consistent and comparable manner. SNOMED CT is also an ontology, which contains a poly-hierarchy model, meaning that a code can be grouped into different categories, forming directed acyclic graphs (DAG). Nonetheless, these graphs are difficult to implement in their entirety using relational or object-oriented databases, resulting in a loss of a wealth of information. SNOMED CT's continued expansion will make it more complex and model consistency will be more difficult to assure. Moreover, consumers of data will increasingly demand improvements to query functionality to accommodate additional granularity of clinical concepts without sacrificing speed. The above information and comparative retrieval characteristics of relational databases, triplestores, and general graph databases [9] determined the use of a graph database.

The emergence of Big Data has spurred development of new technologies and platforms, for example, Neo4j [2] which is a high-performance NoSQL database. We chose Neo4j because it had the best score of the graph databases [8]. This platform has already been used in different industries such as financial [3], social networks [4], in health [5], among others. Neo4j is a graphics database whose basic structure is composed of nodes, relationships and attributes. The nodes are designated as starting nodes and termination nodes, and two such nodes are connected by a relationship (edge). This database allows the representation of the SNOMED CT ontology, without loss of information, which makes it one of the most suitable platforms for this type of work. SNOMED CT has already been used in different works in conjunction with Neo4j [6] [7], but it has never been used as a terminology server to support clinical history in real time. Concretely, the objective of this work is the implementation of a terminology server in a health institution in Uruguay, using SNOMED CT on a graph-oriented database, in order to create an intuitive TS, making maximum use of the characteristics of the SNOMED CT ontology, and also to ensure usability from a clinical point of view.

Methods

In the first step, a graph database containing the SNOMED CT concept model was created. The implementation of the TS was done using the RF2 files of the International Release, Spanish Edition and the Uruguay extension. Currently, the TS is running with the Release International 2018-04-34. In terms of the graph database, the Neo4j Desktop Version 1.0.22 is used. Neo4j is a graph database platform based on Java that supports transactions with ACID (atomicity, coherence, isolation, and durability) properties. The content of SNOMED CT was extracted from the RF2 files and loaded into Neo4j using a series of scripts that were written in Python. The scripts extracted the following information from the RF2 files to create the nodes: "id", "term", and descType. For the creation of relationships, the following data is used: "sourceId", "destinationId", "relLabel", "typeId", "term", "descType." The nodes were created for all the basic SNOMED CT concepts, and the edges were created for all IS_A relationships, which is a relationship from a child concept to a parent concept. All nodes contain every aspect of a SNOMED CT concept such as concept ID, status, and the description of the concept. Queries on the terminology server are made using the declarative language of Neo4j (Cypher). In this language the query defined makes a lexicographical search using the descriptions (without accents and in minuscule) stored in the nodes, then a recursive search is made through the edges, using the IS_A relations (see Figure 1) but in the opposite direction (from parent to child). Finally, as a result of the query, the description and the ID of the obtained concepts is returned.

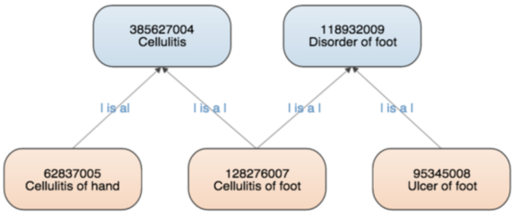

Figure 1 - (SNOMED IS_A Relationship) Obtained from https://confluence.ihtsdotools.org/

The Terminology Server is accessed by clients through a Glassfish application server that uses SOAP or REST protocols depending on the web services invoked. The server consults a Lexicon (under development) in the first instance (a linguistic tool with the morphological variations and grammatical uses of words), subsequently accessing the Terminology Server the Bolt protocol (see Figure 2).

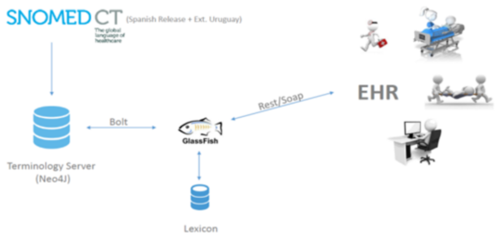

Figure 2 - Architecture

Figure 3 below shows the record form of the EHR, in which the TS is accessed through a web service to obtain terms associated with findings that will be linked to the patient as personal or family background. At present the TS is in production for use by internal medicine physicians.

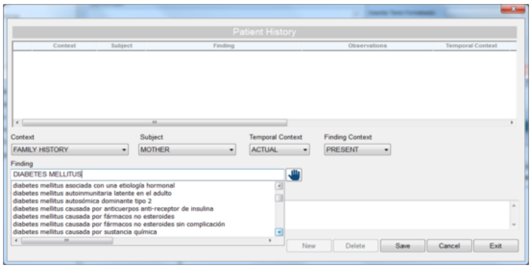

Figure 3 - Background Record Form of the EHR

Terminology Server Features

In the present example (Figure 3), the query to the TS is made from the EHR and includes the following features:

- Analysis of the text entered (diabetes mellitus) by the lexicon.
- Search the set obtained from the lexicon in the database.
- Return of the found terms and the associated coding that was requested when making the query.

Conclusions

The development of a TS with a graph database architecture using SNOMED CT as the basis of the data model allows more comprehensive queries navigating through the database in a hierarchical manner than a relational database. We can navigate through the ontology choosing the levels of specificity in the terms retrievals, and the integration of extensions for use in a particular country can be included thanks to the characteristics of SNOMED CT. It is also noteworthy that the response times to the information are adequate, with an average of 500 milliseconds for the visualization of the collection of terms in the user interface of the EHR.

References

[1] “SNOMED CT Implementation”.https://confluence.ihtsdotools.org/display/DOCTIG/2.+SNOMED+CT+Implementation. [accessed August 5, 2018].

[2] Neo4j. https://neo4j.com/. [accessed August 5, 2018].

[3] G.C.G. Van Erven, M. H., R.N. Carvalho, Á. Rocha, A. Correia, H. Adeli, L. Reis, S. Costanzo. "Detecting Evidence of Fraud in the Brazilian Government Using Graph Databases." Recent Advances in Information Systems and Technologies, WorldCIST 2017, 570.

[4] Nettleton, D. "Data Mining of Social Networks " *Computer Science Review*, **7** (2013), 1–34.

[5] Rajkumar Tekchandani, R. B., Maninder Singh. "Code clone genealogy detection on e-health system using Hadoop." *Computers & Electrical Engineering* **61** (2017), 15-30.

[6] W. Scott Campbell, J. P., James C. McClay, Praveen Rao, Dhundy Bastola, James R. Campbell. "An alternative database approach for management of SNOMED CT and improved patient data queries." *Journal of Biomedical Informatics* **57** (2015), 350-357.

[7] W. Scott Campbell, J. P., James R. Campbell (2015). "Graph database approach to management and use of SNOMED CT encoded clinical data." https://confluence.ihtsdotools.org/download/attachments/12780169/5_GraphDB_2015.pdf.

[8] "DB-engines ranking of graph dbms.", http://db-engines.com/en/ranking/graph+dbms. [accessed August 5, 2018].

[9] D.R. Schlegel, J. P. B., P.L. Elkin. "Comparing small graph retrieval performance for ontology concepts in medical texts." VLDB Workshop on Big Graphs Online Querying, Springer (2015), 32–44.

Address for correspondence

Corresponding author: Elizabeth Silva. Email: elizabeth.silva@smqs.com.uy
Corresponding author: Marcelo Bondarenco. Email: mbondarenco@smqs.com.uy