Master's Thesis planning report

Improving the Expression repository prototype for handling updates to SNOMED CT.

Medhanie Weldemariam

March 15, 2016

1.Author

My name is Medhanie Weldemariam, and I will be the author of this thesis project. I am a master's degree student in computer science at Linköping University since September 2014. My master's degree profile is computer science with broad interest in the area of data science and software developments.

Project Advisor:

Mikael Nyström

Project Examiner:

Daniel Karlsson

2. Preliminary thesis title

The preliminary title for this thesis project is *Improving the Expression repository prototype for handling updates to SNOMED CT.*

3. Preliminary problem description:

The SNOMED CT, Systematized Nomenclature of Medicine Clinical Terms, is a standardized and all inclusive vocabulary for clinical terminologies which covers a broad range of clinical specialties, disciplines and requirements. It allows to store and reuse a structured clinical information in a meaningful and easily understandable way. It also allows the reduction of specialty and geographical boundary effects that arise from use of different terminologies or coding systems by different clinical departments and countries.

In order to get the above mentioned and more benefits, the SNOMED CT terminologies shall be integrated into the electronic health record, thus developing a local expressions repository is necessary to manage the post-coordination expressions. A SNOMED CT enabled electronic health record can additionally provide the advantage of storing relevant health information in a consistent and common way of representation. Further more, it can benefit the society to identify the key health issues, controlling health of the society and keep updated to the major changes in clinical practices. It also improves the communication among health care centers by providing an accurate access to relevant information, reducing the risk of duplications and errors.

IHTSDO, International Health Terminology Standards Development Organization, the owner for the SNOMED CT ontology, updates its release at least twice a year, and every time a new release is published, the local expression repository shall also get an update so that it achieves consistency and interoperability by following latest and the same standard of content encoding as the original source.

Moreover it is necessary to analyze how the SNOMED CT terminologies could be used to represent the complex clinical terms accurately and in a computable form and also efficiently retrieve and analyze encoded problem expressions to drive automated decision support and reporting. The term *expression* refers here, to the SNOMED CT concepts, used to express an instance of clinical information stored in electronic health records. And the expression repository is used to record and

manage SNOMED CT post-coordinated expressions in the local electronic health record. Keeping this local expression repository updated enables to query a latest, predictable and indexable post-coordinated expression. The system, which is currently in use, requires a lot of human interventions in order to manage for any changes from the SNOMED CT in the local repository, and this has resulted in consumption of time, cost and energy. Thus, this project is aimed at extending the already existing system by designing and implementing a new feature prototype that can handle the update management of the expressions from the SNOMED CT terminologies in the local electronic health record database. The outcome of this prototype would hopefully simplify the the current working method.

This study will provide a well documented analysis of the general rules of possible changes that can occur to the expressions as well as a proposed prototype that could automate the into the local expression repository.

4.Preliminary approach

As discussed above, this project is an extension to an already existing system. It will require a good understanding of the release format structure, in order to produce a good prototype that can manage the changes. Additionally more research work and skill development is also expected on the tools and technologies that are going to be used to accomplish the work. Evaluating the implementation and writing a thesis report also time-consuming tasks and requires a considerable time.

This project work is going to be of multiple phase work. Each of the above phases are assumed to be planned in more detail at the beginning of the first days of the respective phase. Each phases will have their own tasks to be completed and the mid-time evaluation will probably be beginning of the second phase.

A study on the SNOMED CT releases will be a starting point to evaluate the release format 2 (RF2) . A prototype consists of rules that manages the the updates must be constructed. And more focus will be required on the phases for design and implementation of the constructed prototype.

The general approach to accomplish the work will involve the following phases:

- Pre-study phase,

- Prototyping,

- Implementation,

- Report and

- Presentation.

4.1. Pre-study phase

This phase is going to be the first phase of the work. During this phase, the proposed literatures and others will be analyzed to get an insight or deep understanding of the release format 2 in each version. A study on the existing system will also be done during this phase, to investigate how it is working currently. A system that handles storing new expressions is already in use, and this system

is going to be used as an existing or related work to what is being planned to be done. This thesis work needs to extend the current system with a new functionality that manages updates. This phase of the work will lead to capture the main changes that has been made so far. Tasks such as resource/information search, documenting introduction, background and theory sections of the report are included here. In general, this phase is going to be a lot of reading with some documenting work.

4.2. Prototyping phase:-

The second phase of the work will be construction of the prototype. The proposed prototype will produce a set of general rules on the RF2 structure.

4.3. Implementation phase:-

Once prototype that demonstrates the feasibility of the RF2 structure is at hand, the implementation work can begin. During this phase of the work, the design for the update rules will be implemented using Java programming language. The implemented prototype will be tested and evaluated by considering sample test cases of on some of the updates domains.

4.4. Structuring report phase:-

This phase of the project work will be dedicated into writing the report, reviewing previous chapters, structuring the document and proof-reading.

4.5. Presentation phase:-

The presentation phase is the smallest phase as compared to the others. During this phase, the presentation, demonstration and oppositions will be done. By this time, most of the work including the report must be ready and no major work will be added to the work. However, other small tasks such as testing and fixing small bugs, organizing documentation etc could possible be done.

5.Literatures and references

The immediate references to this project work will be articles and SNOMED CT document library from located at http://ihtsdo.org/fileadmin/user_upload/doc/. Additional literatures and references suggested by supervisor and examiner will also be used considered as the main resource. Moreover, the existing system and any available related sources codes will also be taken as an input to the work.

(1) SNOMED CT Starter Guide

(2) Vendor Introduction to SNOMED CT

(3) SNOMED CT ® Technical Implementation Guide

(4) SNOMED CT Expression Constraint Language Specification and Guide

(5) SNOMED CT Compositional Grammar Specification and Guide

6.Time plan:

This thesis project requires at least twenty (20) full weeks of time to complete the work. The twenty weeks plan is supposed to have maximum 8 hours of work per a day, and this might include break time such as weekends and any possible holidays, thus some extra days are included to compensate the missing hours. For more, refer to the time table and Gantt chart at the end of this document..

7.Project Risk Management

7.1 Risk:- Author become unavailable due to a problem for unknown period of time.

Probability:- Low (1)

Impact:- Serious (3)

Avoid Risk:- Follow the time plan seriously and continue when available.

Risk Mitigation:- Stop the project until the problem is solved and make a re-plan and extend the tasks.

7.2 Risk:- Examiner or supervisor become unavailable due to a problem for unknown period of time.

Probability:- high (3)

Impact:- tolerable (2)

Avoid Risk:- Follow the time plan seriously and make sure that meetings and supervisions are made in advance.

Risk Mitigation:- Keep in touch through email.

7.3 Risk:- Proposed solution is not acceptable/ didn't work as planned.

Probability:- Moderate (2)

Impact:- tolerable (2)

Avoid Risk:- Plan additional pre-study and prototype construction.

Risk Mitigation:- Deliver a reflection report describing what attempts and efforts are made and specify the reason why it did not work as planned.

		Name	Duration	Start	Finish
1		⊟PRE–STUDY PHASE	52 days	3/8/16 8:00 AM	5/18/16 5:00 PM
2		Project plan	13 days	3/8/16 8:00 AM	3/24/16 5:00 PM
3		Phase–1(Pre– study) Plan	5 days	3/25/16 8:00 AM	3/31/16 5:00 PM
4		Literature review	8 days	4/1/16 8:00 AM	4/12/16 5:00 PM
5		Problem formulation	4 days	4/13/16 8:00 AM	4/18/16 5:00 PM
6		Introduction (documenting)	4 days	4/19/16 8:00 AM	4/22/16 5:00 PM
7		Background (documenting)	5 days	4/25/16 8:00 AM	4/29/16 5:00 PM
8		Study on the existing system	7 days	5/2/16 8:00 AM	5/10/16 5:00 PM
9		Theory (documenting)	5 days	5/11/16 8:00 AM	5/17/16 5:00 PM
10		mid–thesis review	1 day	5/18/16 8:00 AM	5/18/16 5:00 PM
11		⊟PROTOTYPING	25 days?	5/19/16 8:00 AM	6/22/16 5:00 PM
12		Phase-2 (Prototyping) Plan	1 day?	5/19/16 8:00 AM	5/19/16 5:00 PM
13		Construction of RF2 rules	8 days?	5/20/16 7:00 AM	5/31/16 5:00 PM
14		Modeling rules	10 days?	6/1/16 8:00 AM	6/14/16 5:00 PM
15		Method (documenting)	6 days?	6/15/16 8:00 AM	6/22/16 5:00 PM
16		⊟IMPLEMENTATION	27 days?	6/23/16 8:00 AM	7/29/16 5:00 PM
17		Phase-3 (Implementation) Plan	1 day?	6/23/16 8:00 AM	6/23/16 5:00 PM
18		Study of tools and technologies	5 days?	6/24/16 8:00 AM	6/30/16 5:00 PM
19		Algorithm implementation	16 days?	7/1/16 8:00 AM	7/22/16 5:00 PM
20		Testing	3 days?	7/25/16 8:00 AM	7/27/16 5:00 PM
21		Evaluating	2 days?	7/28/16 8:00 AM	7/29/16 5:00 PM
22		⊟REPORT	23 days?	8/1/16 8:00 AM	8/31/16 5:00 PM
23		Phase-4 (Report) Plan	1 day?	8/1/16 8:00 AM	8/1/16 5:00 PM
24		Results (documenting)	3 days?	8/2/16 8:00 AM	8/4/16 5:00 PM
25		Discussion (documenting)	3 days?	8/5/16 8:00 AM	8/9/16 5:00 PM
26		Conclusion (documenting)	3 days?	8/10/16 8:00 AM	8/12/16 5:00 PM
27		Grammar Proof-reading	13 days?	8/15/16 8:00 AM	8/31/16 5:00 PM
28		References Proof-reading	13 days?	8/15/16 8:00 AM	8/31/16 5:00 PM
29		Structuring thesis report	13 days?	8/15/16 8:00 AM	8/31/16 5:00 PM
30		Others	13 days?	8/15/16 8:00 AM	8/31/16 5:00 PM
31		⊟PRESENTATION	10 days?	8/18/16 8:00 AM	8/31/16 5:00 PM
32		Phase-5 (Presentation) Plan	4 days?	8/18/16 8:00 AM	8/23/16 5:00 PM
33		Final presentation	1 day?	8/24/16 8:00 AM	8/24/16 5:00 PM
34		Reflection Report and correction	4 days?	8/25/16 8:00 AM	8/30/16 5:00 PM
35		Others	1 day?	8/31/16 8:00 AM	8/31/16 5:00 PM