Translation of SNOMED CT – Strategies and Description of a Pilot Project

Gunnar O KLEIN[a] and Rong CHEN[b]

[a] *Department of Microbiology, Tumour- and Cell Biology, Karolinska Institute, Sweden*
[b] *Department of Biomedical Engineering, Linköping University, Sweden*

Abstract: The translation and localization of SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms) have been initiated in a few countries. In Sweden, we conducted the first evaluation of this terminology in a project called REFTERM in which we also developed a software tool which could handle a large scale translation with a number of translators and reviewers in a web-based environment. The system makes use of existing authorized English-Swedish translations of medical terminologies such as ICD-10. The paper discusses possible strategies for a national project to translate and adapt this terminology.

Keywords: SNOMED CT, Terminology server, Translation, Swedish, Tool

1. Introduction

There is currently interest in using SNOMED CT in a number of countries outside of the English speaking countries. The new IHTSDO – International Health Terminology Standards Development Organization includes Denmark, Sweden, the Netherlands and Lithuania from the formation but possible translation and use is discussed in several other countries. For a few large languages such as Spanish and German, the College of American Pathologists together with other partners provided first versions of translations to these languages but these have as yet to be used on a large scale.

When the first version of SNOMED CT was released in 2001 [1], we started a pilot project called REFTERM in Sweden with the following objectives:

- To train a core group of experts on SNOMED CT design principles and characteristics
- To make a first evaluation of the suitability for general use in our country for various purposes
- To develop a Terminology Server as a tool for future applications using SNOMED CT at runtime and for a Translation system
- To develop principles and pilot tools for a large scale translation and to perform a limited test translation

The results of the assessment published in Swedish [2], contributed to the government decision in 2006 to make Sweden one of the founding members of the IHTSDO.

This paper describes the tools developed and a number of general principles and issues related to the translation and localization tasks which would be applicable irrespective of the tools used.

2. Methods

2.1. Requirement Analysis

The translation system development started with a review of requirements for using SNOMED CT for various purposes based on a literature review and interviews with selected experts. The detailed characteristics of translation tool were developed in co-operation with Mikael Nyström at Linköping University.

2.2. System Design

The system consists of four sub-systems:

2.2.1. REFTERM Terminology Service

This service provides a dedicated storage service for the SNOMED CT content together with the locally generated translated terms. It was constructed to generate different built-in relationships. The intended use of this was to support the review and translation system but in addition, this could be a platform for clinical applications where this would be integrated with various new or legacy EHR systems.

2.2.2. REFTERM Translation Service

This subsystem provides the functionality needed for translation and review. It is based on the Terminology Service and uses the Translation Aid.

2.2.3. Semi-automatic Translation Aid

This sub-system utilizes pre-loaded mappings from English terminologies to Swedish translations. We had access to ICD-10, MESH and ICPC. If complete identity was found between a previously translated term and the SNOMED term, the translation was used automatically as a first proposal (to be reviewed). During the translation or review processes, the concept description[1] is used to find a candidate translation. The search algorithm is based on partial match. Each proposed translation is recorded with information about the source description, and the name of the mapping used. For a further discussion on this see Nyström et al [3].

2.2.4. Web Application with User Management

This is the web front-end, which exposes the back-end logic to the end users. It also provides the service to manage the users of the application, e.g. add new users with assigned role(s).

2.3. Implementation tools

The system is built with J2EE technologies and well-established Open Source products. The presentation layer is implemented as Java Server Pages and Servlets using a Java web application framework called Struts. It runs within the Servlet container Tomcat as part of the application server JBoss at runtime. The application persistence layer utilizes an object relation mapping tool called Hibernate. It uses MySQL for data storage and retrieval. The client part of the application can be any web browser.

---

[1] Description is the name used in SNOMED CT for the terms of the concepts.

3. Results

3.1. Main advantages, of SNOMED CT

From our analysis it was clear that there would be several important possible uses of SNOMED CT in our country, and the following characteristics were noted:

- SNOMED CT provides an advanced and in some regards good structure which was expected to provide a basis for automatic reasoning
- SNOMED CT provides a great deal of content relevant for EHRs, in areas also where we have had no previous national terminology
- The SNOMED structure provides for terms in multiple languages for the same concept. The English language would facilitate international biomedical scientific co-operation and publication from clinical sources originating in Swedish and also the use of international decision support systems. The potential ability to use multi-linguality to support eHealth across borders was emphasized and we also recognized the possibilities for providing automatic translations to a number of different languages preferred by immigrants in Sweden where 20 % of the population has a foreign background.

3.2. Requirements on the translation

a) In order to get high quality, speed and a wide acceptance for the translation, it is essential to include a large number of domain experts from different fields. Thus, a web based system is needed and a system to manage the process to allocate small sets to different translators.

b) Translations need review and approval in several iterations.

c) Many complete terms, sub phrases and words are already adequately translated in other terminologies and should be used, such as ICD-10.

d) Because of the systematic nature of the SNOMED CT, it will be possible to reuse fully specified names, many sub phrases and words in the translation of other concepts if the system can provide these as suggestions.

e) The meaning of the preferred term must be preserved and this is given by the defined relationships. Therefore these must be visible to the translator and reviewers, in addition to the term itself.

f) A term that is in line with the general national recommendations for medical language should be the preferred term (in Sweden as defined by the Language Committee of the Swedish Society of Medicine). In most cases, this will be a term already in common use.

g) Agreement of the experts was not reached concerning the issue of the "Fully Specified Names" of SNOMED CT. Some, like the authors, claimed that the systematic nature of these, often with four or five words, should be translated, while other experts suggested that shorter Swedish forms should be found or created. Note that in Swedish and some other related languages it is customary to create new terms by joining complicated multiword concepts into one long word. Making a literal word for word translation of systematic fully specified names may enhance clarity on the meaning of the concept in a hierarchy but may be considered rather clumsy for everyday use as a preferred term.

h) Both one preferred term and synonyms would be useful although it is agreed that it does not make sense to translate all synonyms in English if there is no corresponding use in the national language. Similarly, it may be necessary to

include additional synonyms from the national language with no correspondence in English.

i) In some areas, particularly related to care administration, there will be a need to enter terms and concepts as a national extension where there is no corresponding content in the international version.

j) A translation system must have a full handling of versions

3.3. Implementation

3.3.1. The REFTERM terminology service

This basic service was built by one of us, RC, and was found to perform satisfactorily for a multi-user translation use even with rather basic hardware.

3.3.2. The REFTERM translation service

The translation service system built contains the following functions:

- Managing users with passwords and roles
- Allocating translation and review tasks to identified users based on selecting subsets of the SNOMED CT
- A translator can perform the following for each concept presented:
  Select type of description (Fully Specified, Preferred or Synonym). Translate one or all the descriptions of the concept while preserving the type of descriptions. It is also possible to change the type of the description. For example, if a Concept has a preferred term and 3 synonyms, it is possible to create a preferred term in Swedish that is based on one of the synonyms.
  In the translation task, one or more proposals for a complete or partial translation is generated by the system (see below). This proposal can be used as the basis for typing the proposed translation. An accompanying note can also be given to motivate the selected proposal if deemed necessary.
- One or several (possibly many) reviewers will be assigned to review the translation result. The reviewers can mark the translation as accepted or rejected with an optional reason. Possibly several iterations of reviews and translations are needed if agreement is not reached immediately.
- An administrator may accept the translation and thus publish the result.
- Semi-automated translation aid. Existing translations of ICD-10, MESH and ICPC were used to generate completely automatic first translations and proposals for a semi-automated process when there was only a partial sub phrase match. This subsystem provided more than 9000 reliable complete translations with an exact match of term, most of these coming from ICD-10.

4. Discussion

Translation and localization of SNOMED CT is an important task in several countries and although there is much in common compared to translating other international terminologies such as the WHO family of classifications, SNOMED requires other tools and strategies. Our tools are available as open source and may be of benefit to such projects, at least as a low cost starter. It is our hope that through international co-operation these tools can be developed further.

We believe it is essential to involve as much as possible a country's health professionals to obtain maximum acceptance of the new terminology. International experience with e.g. the German version, made only by professional translators provided by CAP, which met with resistance, emphasizes this [4]. Preliminary Danish experiences where a professional translation agency has translated a large part of SNOMED seems to indicate a similar problem. The Swedish government has decided to follow the Danish example to a large extent and has acquired the Danish translation software and has contracted a Swedish bureau of translators. Some preliminary experience indicates that the translations provided do not match current medical terminology. It is possible that contracting of professional translators will lead to faster results, (four years are estimated in Sweden as in Denmark) but its long term acceptance remains to be seen. A well organized project involving a large number of qualified medical experts may possibly complete the task faster and to a lower direct cost. The Spanish translation, initially with a previous version, employed a combination of techniques [5]. Interestingly the IHTSDO members Netherlands and Lithuania have as yet not initiated a general translation to their languages.

In contrast to start translation of all the many terms, in no particular order, we believe that it may be of benefit to initially select minor parts of the terminology that will be used in real implementations, including all the referenced concepts which contribute to the definition of the selected ones. This will ensure that useful material is available early on to allow software suppliers to familiarize with the structure of SNOMED, and to present solutions to end users where SNOMED content is used. National strategies for introducing SNOMED CT should make great efforts to involve its industrial suppliers as early as possible.

5. Conclusion

We have developed an open source tool for large scale translation and for integration with clinical applications that aims to facilitate the adoption of SNOMED in various languages. Strategies for translation and localization should be carefully assessed before starting large expensive national projects.

References

[1] Spackman KA et al, eds. SNOMED Clinical Terms First Release. College of American Pathologists, Northfield, IL, USA, 2002.

[2] Klein, GO. Vad är SNOMED CT?, Rapport från projektet REFTERM, Karolinska Institutet 2005.Acknowledgements: This work was supported by the KKS foundation and the EU project Semantic Mining

[3] Nyström M, Merkel M, Ahrenberg L, Zweigenbaum P, Petersson H, Ahlfeldt H. Creating a medical English-Swedish dictionary using interactive word alignment. BMC Med Inform Decis Mak. 2006 Oct 12;6:35.

[4] GMDS. Positionspapier zur „Systematized Nomenclature of Medicine – Clinical Terms" (SNOMED CT) in Deutschland. www.gmds.de/pdf/publikationen/stellungnahmen/Positionspapier.pdf. Accessed 2008-09-01.

[5] Reynoso GA et al. Development of the Spanish version of the Systematized Nomenclature of Medicine: methodology and main issues. Proc AMIA Symp. 2000:694-8.

Email address for correspondence: gunnar.klein@ki.se