MEdical Terminology browsers: How usable are they for describing clinical archetypes?

Jesús Cáceres Tello¹, Miguel-Ángel Sicilia Urbán², Adolfo Muñoz Carrero¹, Carlos Rodríguez-Solano Nuzzi²

¹Carlos III Health Institute. Telemedicine and e-Health Research Area, Madrid, Spain
²Department of Computing Science, University of Alcalá, Alcalá de Henares, Madrid, Spain
{jesus.caceres, adolfo.munoz}@isciii.es, {msicilia, carlos.solano}@uah.es

Keywords: SNOMED CT, archetypes, usability.

Abstract: Clinical terminologies are a major concern in medical informatics, as they are key to provide medical systems with higher levels of interoperability. Large terminologies as SNOMED CT are gaining presence in practical applications. In a related but different direction, archetypes or data type templates are becoming widespread as interchange mechanisms for medical information. Archetypes support mapping to terminologies, in a process that is typically done by the experts developing the archetype. It has been argued that terminology browsers are not appropriate for the task of helping clinical experts in the mapping process. This paper reports usability studies on two popular SNOMED CT browsers when used as tools for mapping archetypes. Empirical findings suggest that problems are due to the exposure of internal details of the terminology in the interfaces and the fact that they do not provide clear distinctions on the kinds of entities encoded in SNOMED CT.

1 INTRODUCTION

The archetype formalism has been proposed for specifying models of elements of electronic healthcare records as a means of achieving interoperability between systems. Archetype-based systems have attracted an increasing attention as a meaningful way for the interoperability of heterogeneous systems (Wollersheim, Sari and Rahayu, 2009). However, some sort of mapping of archetype data elements with shared terminologies is required to guarantee a level of common semantics across archetypes and also with existing clinical systems. SNOMED CT (Systematized Nomenclature of Medicine Clinical Terms) is a large-scale comprehensive clinical healthcare terminology that is gaining widespread use and can be used for that binding.

The mapping process of archetypes with SNOMED terms consists essentially on associating a data element in the formal Archetype Description Language (ADL) to a SNOMED element or expression. While this process seems to be simple and straightforward, it entails several practical difficulties when approached from clinical experts.

These include the following:

- experts need to decide on the archetype elements to be mapped; due to the hierarchical nature of archetype expressions, this can be done at several points in the hierarchy
- SNOMED CT represents different kind of entities, and often a lexical matching is not enough to provide a correct mapping, as there is a need to know the hierarchy to which the term belongs and
- In many cases, the concrete entities are not directly available in SNOMED CT as concepts, but they can be expressed through coordinated post-expressions, so that it may represent a concept in an expression using a combination of two or more concepts. The use of post-coordinated expressions is justified by the same scale terminology SNOMED CT that do not cover all potential clinical concepts and ideas potentially represent. A good example of what is trying to say with the post-coordinated expressions could be next, it would represent the term "emergency appendectomy" where the concepts "appendectomy" and "urgency"
are covered by SNOMED CT but change the complex concept "emergency appendectomy" is not. The concept in this case would represented by the expression:

"80146002|apendiceptomía|:260870009|prioridad|=103391001|urgente|"

However, the language for using those expressions requires some training beforehand for the experts to master it. These issues make the mapping process difficult, and a part of the problems found can be hypotetised as coming from the usability of the tools for browsing SNOMED. In that direction, this paper presents an initial exploratory study to gather insights on the usability of SNOMED browsers for the particular task of archetype mapping. It should be noted that the present study is complementary to existing ones that look at the inherent complexity of SNOMED as a terminology (Duo Wei et al, 2008) or to the complexity and ease of understanding of archetypes (Borovicka et al, 2010). The aim of this research is obtaining insights that can be used as input for devising integrated SNOMED-archetype editors or browsers for doing the mapping. Following the direction of Balbo et al. (2008), usability studies for clinical software need to involve the users, which in this case are the clinical experts.

There is an increasing number of SNOMED browsers available, either commercial or free. And they differ significantly in the way of searching elements and presenting the structure of relationships inside SNOMED to users. Rogers and Bodenreider (2008) inspected 17 different browsers out of 23 identified and extracted a common set of characteristics.

The guiding exploratory hypotheses of the present study are the following:

- Are SNOMED browsers usable for supporting the task of archetype mapping?
- How are SNOMED browser users approaching the search and navigation process for the concrete task of searching for terms mapping particular archetype entities?

It should be noted that the cognitive task required for mapping an archetype term requires an understanding of the context of the archetype element selected, and this in turn requires understanding the position of SNOMED terms inside the taxonomy.

The rest of this paper is structured as follows. Section 2 provides background information on the archetype approach and the role of terminology mappings in providing semantics to archetypes, along with a review on studies related to the usability of SNOMED or similar tools. Then, the methods for the exploratory study are described in Section 3. Results and discussion are provided in Section 4 and finally, Section 5 is devoted to conclusions and outlook.

2 BACKGROUND

The archetype approach to the interoperability of healthcare systems is based on the notion of “two-level” modelling (Beale, 2002).

This concept is given by the idea of obtaining a complete ontological separation between the information model and the model of knowledge. On the one hand data from standards such as UNE-EN ISO13606 or terminology databases of different nature as SNOMED CT or LOINC, and on the other hand, the knowledge generated by this information in healthcare.

![Two-level modelling](image)

Figure 1: "Two-level" modelling (based on Beale, 2002)

Thus, we understand how all that information is stored in the system and does not vary over time, as are data on the patient's health. On the other hand we understand as knowledge the collection of facts accumulated over time to form a series of concepts that allow us to interpret the above data and who can be in constant evolution and change without being spoiled already stored data.

Using the UNE-EN ISO13606 we can see that the information object per excellence in the norm is called “extract” while the knowledge is represented by the model of archetypes (Muñoz et al, 2007).

The challenge to achieve interoperability in this sense is able to represent all possible data structures properly, in terms of health records are concerned. A feature in this context is the variability and complexity of clinical data sets, templates and ultimately in the way of representing data. The
archetypes emerge as a possible formal definition of possible compositions of model components for each clinical service references found in particular. In this sense an specific archetype or restricts the hierarchy of subclasses of a record component within a larger structure such as the extract, EHR Extract defining their names, optionality, cardinality, data types and ranges of these and even defaults for some of its components.

An archetype is expressed by the ADL language provides a formal means of expressing instances of data based on an information model which can be read by both humans and by machines. An ADL document expresses the serialization of one or more complex objects, where each one is a hierarchical attribute-value, where each attribute is the name of a real or implicit object that can be represented by a concept in a SNOMED CT as a terminological system.

The current practice of binding archetypes to terminological systems consists on the explicit declaration of coded terms or concepts. For example, the archetype for body weight (archetype with identifier openEHR-EHR-OBSERVATION.body_weight.v1) in the OpenEHR CKM repository contains the following term binding section.

term_bindings = <
  "SNOMED-CT" = <
    items = <
      ["at0000"] = <[SNOMED-CT (2010) :: 363808001]>
      ["at0001"] = <[SNOMED-CT (2010) :: 190975009]>
      ["at0002"] = <[SNOMED-CT (2010) :: 307294006]>
      ["at0003"] = <[SNOMED-CT (2010) :: 222913008]>
      ["at0004"] = <[SNOMED-CT (2010) :: 228910000]>
      ...
    >
  >
>

3 METHODS

This study was developed in response to obtaining three specific objectives:

a. A usability study of each of the tools proposed
b. As a complement to usability study of thinking aloud protocol collected thoughts, feelings and opinions of participant while performing the test of experience.
c. An experience study where the user makes a series of field tests on coding concepts using SNOMED CT.

For the usability study a questionnaire was designed to first collect a set of basic data of the participant, relating not only to his person, such as nationality, sex or year of birth, but also relative to career as the Medical Institution where he developed his job or years of experience in the healthcare environment. In this context, see different roles thought of health as primary care doctors, hospital doctors, nurses, pharmacy or documentary staff. The question of his previous experience with medical terminology was not considered.

The second part of the study itself the usability of the two applications by the user or participant. Usability is an important aspect of any software. Factors such as ease of installation or further learning and use of certain software are very important indicators when assessing the quality of a particular software application and its future. In this sense we discussed different types of questionnaires, SUS (Brooke, J. 1996), QUIS (Harper B, Norman K, 1993), CSUQ (JR Lewis, 1995), SUMI (van Veenendaal, 1998). Finally it was decided to conduct the study using the methodology SUMI (Software Usability Measurement Inventory). This type of questionnaires developed by the Human Factors Research Group (HFRG), University College Cork, in 1986 and allows a heuristic evaluation of usability of software according to the perceptions and attitudes of users who use it, resulting reliable indicator amounts of software quality by analyzing five key areas (Kirakowski, Corbett, 1993):

- **Efficiency**: refers to the sensation perceived by the user when interacting with the software as regards the speed and effectiveness that allows the program when performing a task.
- **Affect**: referring to the emotional state that occurs in the user mental stimulation that both positive and negative experiences when this interaction with the program.
- **Helpfulness**: refers to the sensations perceived by the user when interacting with the program's help in solving the operational problems that may be created.
- **Control**: referring to the perception of the user in the sense of who is setting the pace in solving tasks.
- **Learnability**: referring to the ease with which the user learns to know the new features of the product.

The experience study consisted of the search of SNOMED codes for the users to identify both the archetypes according to UNE / EN ISO13606 and all of data fields that make up. In this sense, a first methodological issue was that of the source of the archetypes. The CKM repository maintained by the OpenEHR foundation was used as it is the more
mature platform and it is also open and not restricted to some particular institution's viewpoint. The CKM repository manages reviews and publishes a wide range of clinical knowledge resources. The resources managed include archetypes, models, and metadata set of terms that define the clinical models and related resources. The openEHR approach proposes a clinician-led development of quality models or archetypes that can be reused in the future.

Another reason for choosing the repository maintained by the openEHR was that the archetype language of OpenEHR includes five kinds of entities that come from ontological analysis (Beale and Heard, 2007), namely Observation, Evaluation, Instruction, Action and Administrative. They are representing different types of care entries, and may be hypothesized to bring different requirements for the mapping:

- The observation archetypes will be responsible for establishing the data model for all information obtained by the researcher through measures or test or analysis of questionnaires designed to uniquely characterize the patient's system. An example of such archetypes could be "openEHR-EHR-OBSERVATION.indirect_oximetry.v1" which includes the method of monitoring blood gas measurements such as Spo2 and pTco2 by indirect, currently non-invasive, means.

- The evaluation archetypes are those that are more related to the cognitive basis of health care and establish the information model on the meaning of these observations obtained as diagnoses or assessments. An example of this kind of archetype may be "openEHR-EHR-EVALUATION.adverse.v1" which recording the presence of a harmful or undesirable response to an agent or substance including food, as determined by the clinician - excluding poisoning and abnormal use.

- The archetypes of instruction contain the set of instructions directly executable by the researchers' agents, be they people or machines, in order to take action including samples that are considered, like a biopsy. Used to pick the steps that must be taken in any clinical activity, including steps for cancellation or postponement. In this group of archetypes will be used for both standard and clinical processes that respond to a request ad hoc. An example of archetypes of instruction is "openEHR-EHR-INSTRUCTION.transfusion.v1" which includes instructions for transfusion.

- The action archetypes contain information about interventions that have occurred due to information obtained from the algorithms of instruction. An example of an action archetype may be "openEHR-EHR-ACTION.medication.v1" which describes the characteristics of a medication ordered by a physician which specifies which medication to take, when, for how long, etc.

- The administrative archetypes will define data structure to record business events arising within an administrative context, like admission, bed hold, and so on. An example of this kind of archetypes can be "openEHR-EHR-ADMIN_ENTRY.admission.v1" used for Admitted patient only. It signals to the Beginning of patient's stay in a health care facility.

In order to not prolong the duration of the test was decided to choose two of the aforementioned archetypes according degree of use by clinicians in general. With this premise, the selected were observation and action archetypes as the most important types of archetypes to carry out the study. To increase the effectiveness of the study were performed between archetypes and software combinations for analysis so that it covers all cases with the number of samples taken as indicated in the table below:

<table>
<thead>
<tr>
<th>Users</th>
<th>Archetypes</th>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observ.</td>
<td>Action</td>
</tr>
<tr>
<td>1</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>2</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>3</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>4</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>5</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>6</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>7</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>8</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Table 1: Distribution of tasks among the participants in the study

So, this covered all possible combinations to prevent intoxicated data in the study results. the sequence began again with the user number nine.
Additionally SUMI questionnaire, to obtain data about the quality of the software used, we implemented a "thinking aloud" protocol (Lewis C.H., 1982) during the study of experience, at technical that allow data acquisition of cognitive activity user during the execution of a task (concurrent) or as this task is being remembered (retrospective) (Ericsson & Simon, 1993). Using this technique, considered at the group of "Oral Protocols", we can see cognitive and metacognitive strategies used by the informant (Jaspers, 2009). In this study, notes were taken during each respondent's participation in the study and have questions about possible improvements to the software and possible complications in the use of SNOMED in their daily work.

The study was limited to free browsers having at least the core SNOMED CT data browsing features identified by Rogers and Bodenreider (2008). After a review of existing free ones, the selection resulted in ClinIClue and Minnow.

ClinIClue browser was found by Elhanan et al. (2010) the most popular tool used to access SCT (54%, exceeds 100%), as well as the most preferred one (29%). In contrast, Minnow is a relatively center lease. Minnow was selected due to their wide differences in response time (as they are using Lucene indexes inside) and the use of more agile hierarchical browsing representations using profuse libraries. Even though search and traversal mechanisms can be considered similar, look & feel and interaction has some differences that make them good candidates for a contrast on these details.

Participants in the user test were asked to verbalize their intentions while performing the mapping tasks, while being observed by a researcher. Prompts were only provided if the user was unable to proceed. This was in combination with screen recording for subsequent analysis of the actions taken and the notes of the researcher. Users are asked to formulate any question or comments for discussion with the researcher after the event.

Each task was presented as a task of assigning codes SNOMED CT as the name itself of the above archetypes as well as the data fields they contain. In this case, the breakdown in subtasks was user initiated, as they were given freedom to decide from where to start and what archetype elements to map. The users were shown directly to have GUI software to infer what features they should use and where to find (the search for concepts, for example). At no time were shown the graphical view of openEHR CKM nor explained the concept of archetype simply told them, who requested it, that these were data structures relating to a particular medical intervention (observation, action, etc. ..) Thus a search was achieved directly without the confusion could be created at the time of making a light and quick explanation on the term archetype ADL or on codes relating to the definition of it.

After completing the session "think aloud" while performing the tests of experience, after each of them (the two archetypes in this study), participants were invited to perform the SUMI questionnaire online finishing the test.

Finally it was decided not to ask the informant for post-coordinated expressions as a possible solution in the SNOMED CT code assignment in the data that make up each of the archetypes selected as this would require prior training. Instead they were asked to users who could express the identification of an archetype as a combination of SNOMED CT terms rather than a single term.

4. RESULTS AND DISCUSSION

To do the study's tests, enabled a room with several computers each of whom had two browsers installed SNOMED CT that were analyzed (ClinIClue and Minnow). In addition, each team had an Internet connection to allow the user to each of the two online surveys, the part of usability by entering the personal and professional data were estimated and the SUMI questionnaire of 50 questions for further automatic processing. Each participant while performing the tasks was accompanied by an observer to collect interesting data using the protocol "Think Aloud" formerly mentioned.

4.1 Participants

Users who participated in this study were all active and were from different health institutions including the University Hospital of Fuenlabrada, Clinic Hospital San Carlos, 12 de Octubre Hospital and 1 Hospital in Madrid and Principe de Asturias Hospital in Alcalá de Henares.

The study involved a total of 14 participants (8 women and 6 men), the cost-benefit ratio is lower applying the tests 3 to 5 users (Nielsen, 1993). The ages of the participants were besieged from 34 to 52 years the average age being 40 years, all of Spanish nationality.

As for the experience of participants in the health context, the average was 15 years and his previous experience with medical terminologies did not reach 33%.

Most participants were hospital doctors, as shown in the following graph reflects the different roles covered by the study:
As for the time spent in completing this test varied according to previous experiences of participants with clinical terminologies as well as browsers of this type. In mid-range was situated at 72 minutes in the realization of global test. Interestingly, in this case was that less time spent in conducting the test with the browser Minnow than the browser CliniClue as shown in the following graph:

![Figure 3: Time spent on the resolution test](image1)

As for dropout rates, there were 4 dropouts in the entire test. One drop was the start of the test and the other three were with the first browser, but they didn't the second so it was also considered dropout. Regarding the reasons behind these dropout, as indicated by the participants, was due, among other reasons, to the language and the time of the tests.

In all cases, participants expressed the importance of the idiomatic wall that they found on search the concepts in other language than their own. While some had experience with clinical terminologies, had always worked with the Spanish version of SNOMED CT. The following chart shows the importance of this indicator in the study.

![Figure 4: Reasons for dropout the study](image2)

### 4.2 Usability of the browsers

Regarding the usability of the browser, after collecting data from users in the SUMI questionnaire on each browser. The following table shows the results collected in the form SUMI specifying each of the subscales in the case of CliniClue browser.

<table>
<thead>
<tr>
<th>SD</th>
<th>D</th>
<th>NA/ND</th>
<th>A</th>
<th>SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>13</td>
<td>78</td>
<td>142</td>
<td>97</td>
</tr>
<tr>
<td>Efficiency</td>
<td>6</td>
<td>16</td>
<td>31</td>
<td>17</td>
</tr>
<tr>
<td>Affect</td>
<td>1</td>
<td>27</td>
<td>20</td>
<td>13</td>
</tr>
<tr>
<td>Helpfulness</td>
<td>2</td>
<td>11</td>
<td>29</td>
<td>23</td>
</tr>
<tr>
<td>Control</td>
<td>3</td>
<td>11</td>
<td>32</td>
<td>21</td>
</tr>
<tr>
<td>Learnability</td>
<td>1</td>
<td>13</td>
<td>30</td>
<td>23</td>
</tr>
</tbody>
</table>

*SD: Strongly Disagree, D: Disagree, NA: Neither Agree/Nor disagree, A: Agree, SA: Strongly agree*

Table 2: The results of SUMI questionnaire for CliniClue browser.

In this case, the highest levels were obtained in the Efficiency subscale, while the lowest corresponded to the subscales of Helpfulness and Control, i.e., users perceived a lack of friendliness in the search interface and of control in the search process, instead, assessed the performance aspects of the program as well as the level of satisfaction in their management (affected).
The following table shows the results collected in the form SUMI for the Minnow browser, as can be seen, are clearly superior.

<table>
<thead>
<tr>
<th></th>
<th>SD</th>
<th>D</th>
<th>NA/ND</th>
<th>A</th>
<th>SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>20</td>
<td>90</td>
<td>122</td>
<td>102</td>
<td>16</td>
</tr>
<tr>
<td>Efficiency</td>
<td>7</td>
<td>23</td>
<td>18</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>Affect</td>
<td>1</td>
<td>17</td>
<td>19</td>
<td>28</td>
<td>5</td>
</tr>
<tr>
<td>Helpfulness</td>
<td>0</td>
<td>21</td>
<td>26</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>Control</td>
<td>6</td>
<td>17</td>
<td>31</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>Learnability</td>
<td>6</td>
<td>12</td>
<td>28</td>
<td>23</td>
<td>1</td>
</tr>
</tbody>
</table>

(*Strongly Disagree, Disagree, Neither Agree/ Nor disagree, Agree, Strongly agree

Table 2: The results of SUMI questionnaire for Minnow browser.

In this case, the levels are clearly higher than CliniClue. The highest levels corresponded to the subscales of friendliness and satisfaction in their use. Participants generally viewed the multimedia component of this browser and the arrangement of windows in the interface very useful and explanatory. Lower levels in this case corresponded to the lack of control and learnability.

As for the protocol "Think aloud" some conclusions could be quite significant. It is important to note at first thought this process to meet among other internal indicators the participant may manifest more strongly at the moment of conducting the test. The indicators were sought were:

- Complexity of the test.
- Comments on the proposed archetype and all the data it contains.
- Observations on the coding of these data.

Regarding the test itself, the perception of most participants (84%) was that it had not much complexity, also they agree in stating that they had found the test very complete. Regarding the duration of the same they all considered too long, answering the question asked at the end of the test by the observer.

In regard to the proposed archetypes, most of the comments received were related to the inability to find some of the proposed concepts. In some times this lack of results in the searches occurred in both browsers, at other times produced results in only one of them. The following chart shows the total number of concepts located in both a browser and in another by all members of the study.

Was also seen that the encoding proposed by the data user archetypes that make up the varied markedly raised which was immediately indicated by the study participants. The level of agreement not reached 10% for participants in the study raised a number of causes of which the most defended were the difference in the search engines of the two software or differences in the visualization of concepts as which they could infer dramatically in the final choice of concept. The chart below reflects the views of participants about the reasons for this mismatch.
Although in all cases there was this mismatch of codes, all participants agreed in their enthusiasm for the test performed and considering the use of terminology standards as a necessary step to achieve semantic interoperability of medical systems.

4 CONCLUSIONS

Using the methodologies used to evaluate the usability of the browsers tested is plausible. This study provides a better understanding of the usefulness of SNOMED CT terminology standard and their use in identifying the concepts of archetypes according to UNE / EN ISO13606. In this sense it has been shown using SNOMED CT codes in the identification of the concepts that make up a openEHR archetype in the identification of archetypes themselves with more general concepts in these cases.

This study has also proven the effectiveness of the protocol "think aloud" in the participant's registration information with the figure of an observer. Perceptions as cognitive methodology in the study participants when performing searches or impressions of the proposed project have been collected with this method.

Future work will extend the exploratory usability studies to a wider range of SNOMED CT browsers, and to search and browsing tasks following a finer grained approach in task descriptions. Rogers and Bodenreider (2007) description of features and cases for visualization, navigation and searching can be used as a catalogue for such more detailed study.

Any way this study is certainly relevant to government agencies and institutions as well as companies that are committed to interoperability of medical systems using terminologies standards such as SNOMED CT with clinical standards UNE/EN ISO13606.

ACKNOWLEDGEMENTS

This work has been supported by the project “Métodos y Técnicas para la Integración Semántica de Información Clínica mediante Arquetipos y Terminologías (METISAT)”, code CCG10-UAH/TIC-5922, funded by the Community of Madrid and the University of Alcalá.

Likewise, this research work has been partially supported by projects PI09-90110 - PITES - from "Fondo de Investigación Sanitaria (FIS) Plan Nacional de I+D+i", "Plataforma en Red para el Desarrollo de la Telemedicina en España" and FIS 08-1148 "Representación de información clínica en cáncer de mama mediante arquetipos".

REFERENCES


Annual Mid-Atlantic Human Factors Conference, (pp. 224-228), Virginia Beach, VA.


