Barriers to the Clinical Implementation of Compositionality

Lawrence K. McKnight, M.D., Peter L. Elkin, M.D.,
Philip V. Ogren, and Christopher G. Chute, M.D., Dr.P.H.
Mayo Clinic, Rochester, MN

ABSTRACT

Background. Compositional mechanisms for the entry of clinically relevant controlled vocabularies have been suggested as a possible solution to providing adequate descriptive precision while keeping term vocabulary redundancy under control. As of yet, there are no widely accepted term navigators that allow physicians to enter problem lists utilizing controlled vocabularies with compositionality.

Methods. We report on the results of a usability trial of 5 physicians using our most recent attempt at developing the Mayo Problem List Manager. We tested the implementation of an automated term composition, and hierarchical term dissection.

Results. Participants found acceptable terms 96% of the time and found automated term composition helpful in 83% of the case scenarios. There was significant confusion about the terminology used to describe compositional elements (kernel concepts, modifiers, and qualifiers) however participants used the functions appropriately. Speed of entry was universally stated as the limiting factor.

Conclusions. The variety of methods that our participants used to enter terms highlights the need for multiple ways to accomplish the task of data entry. Successful implementation of user directed compositionality could be accomplished with further improvement of the user interface and the underlying terminology.

INTRODUCTION

Development of a meaningful electronic medical record necessitates a deeper understanding of vocabulary construction and organization. Problem lists, recorded using a controlled vocabulary with concept level understanding, will provide the interface to implement clinical reminders, links to clinical guidelines, critical paths, reference material, and other decision support tools. Additional benefits of an EMR for purposes of outcomes research, epidemiology, quality improvement, and more effective cost management also depend on a more descriptive and standardized medical vocabulary. Several language projects, including the ongoing vocabulary project at Mayo, are addressing these concerns. It is becoming increasingly apparent, however, that as the descriptive precision of the vocabulary (support for synonyms, multiple ontologies, semantic relationships, and subsumption) increases, database maintenance and control of term redundancy has become a significant problem.

The current version of the Mayo problem list manager is implemented using a 3 tiered architecture using standard TCP/IP calls so that the system could be run over the Internet. At its base, the system uses an Oracle database loaded and performance tuned by Lexical Technology Inc.. The database is accessed by an RMI server also developed by Lexical. This middle layer is called Metaphrase. Metaphrase provides a list of suggested terms, related terms, and definitions from the UMLS given a text search string. The client interface was developed by the Mayo Clinic, and is written in Java for platform independence. Please see Figure 1 for a screen shot of this interface. The client screen accepts user input, sends this to the server and displays the results of the search matches along with definitions and related terms.
Since the last usability study of our problem manager, there have been extensive changes to the screen layout and input methods. In the past however, it has been identified that the usability of the term manager strongly affects how physicians view the vocabulary. In unpublished usability trials of this product, a strong association between the user interface and perceptions of vocabulary depth was observed. A trial comparing a complex and powerful interface was compared to a simple less powerful design. Users of the complex interface had significant difficulty accomplishing their tasks, and blamed the vocabulary itself. As a result of this experience, the problem list manager has undergone serial usability testing.

The results of a recent usability study of this product have been reported. This study demonstrated the willingness of physicians to use suggested terms from a controlled medical vocabulary rather than using their own wording. The study also demonstrated the benefit of having more specific related terms, and the ability to use abbreviations for terms. This study also identified several key problem areas. For example, the participants felt strongly that the composition line and search line need to be near each other for comparison purposes. This information lead to a redesign of the screen layout. Other features such as drag and drop were added to add consistency with other standard GUI platforms.

Other groups have demonstrated the benefits and limitations of problem list managers in a more general sense. Salmon, et al. suggest that the current implementation of problem oriented medical records are limited by the reluctance of physicians to exert additional effort in maintaining problem lists, as this activity clearly cuts into time working with patients. For this reason, we propose automated term composition to speed complex term entry.

Methods of automating term composition have recently developed. In this method, the original search text is then parsed in the client.
and matched to the returned suggested term list from Metaphrase. The Automated Term Composition (ATC) routine returns a subset of the suggested term list felt most likely to match the users input string. This technique has been tested by indexers on a large sample of problem lists, where automated term composition has been shown to be significantly better than a non-compositional technique at matching terms to original concept meaning.

Salmon also reports the need for effective ways to record related groups of problems that are linked and have the capability to record various levels of granularity needed by the different members of the health care team. To address this concern we propose a novel way of presenting problems as a hierarchical display that allows problems to accept modifiers and qualifiers as well as wholly subsumed problems. The hierarchy of the composition is determined automatically by a process we term automated term dissection (ATD). This compositional relationship is displayed below the user input for comparison, and may be edited as necessary prior to adding the problem to the problem list.

These new techniques have not been evaluated in a usability trial to date. Because of the strong interrelationship between interface characteristics and user perception of lexicon acceptability we performed the following usability study to evaluate the progress of our problem manager application.

METHODS

This study was performed in the Mayo Clinic usability lab following the traditional "talk out loud model" as described by Kushiruk and Wier. The Mayo Clinic Usability Laboratory has had extensive experience with the usability methodology. Five (5) Mayo staff physicians in the specialties of Internal Medicine or Family Medicine were randomly asked to participate in the test. After consent was obtained for videotaping, the participant was scheduled a time to come to the lab. As the participants entered the usability lab, they were greeted by a facilitator and given an overview of the testing method. The problem manager software was loaded for the participants, and limited instruction was given to orient the user to the functionality of the program. Next, the participants underwent a training scenario, where they were asked to do simple directed tasks. After completing the training scenario, participants were observed as they entered 4 moderately complex patient cases into the problem manager. In order to create time pressure, each participant was allowed 30 minutes to enter the cases, but entered only as many cases as time allowed. The case scenarios were designed to evaluate multiple characteristics of the physicians thinking and term entry process as well as their interaction with the computer interface. Each case scenario provided 4-6 concepts to enter. Some concepts were designed to test the use of subsumption within a single clinical problem. For example one case presents a man with AIDS and Cocciodes immitus pneumonia treated with Amphotericin B, who develops renal insufficiency. Participants were not instructed on which case elements to enter, only to enter what they felt to be important. After each scenario the participant was asked to evaluate aspects of program's usability and if they were satisfied with their final entry. If the participant finished the 4 structured case scenarios, he/she was asked to enter a case they have recently seen in clinic as a test of a more generalized problem entry.

RESULTS

<table>
<thead>
<tr>
<th>Question</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suggested terms available for the composition acceptable</td>
<td>11/13 (85%)</td>
</tr>
<tr>
<td>Automated term composition feature helpful</td>
<td>11/13 (85%)</td>
</tr>
<tr>
<td>Automated term dissection accurately represented problem(s)</td>
<td>8/13 (62%)</td>
</tr>
<tr>
<td>Final entry acceptable</td>
<td>11/13 (85%)</td>
</tr>
</tbody>
</table>

Table #1. Post case scenario questions.

The 5 participants were able to enter a total of 13 case scenarios and a total of 48 problem concepts into the problem manager.

Automated Term Composition (ATC). Participants accepted the ATC concept as indicated by their responses to questions following each case. In 11/13 (85%) of the case scenario participants felt the feature to be helpful. Interestingly, this did not always correlate with acceptance of the suggested term as the final entry on their list. In 85% of the case scenarios (26/48 = 54% of concepts entered), the participant modified the suggested term in part. The kernel concept was modified in 11/48 (23%) of entries, while a modifier or qualifier was changed in the other 15/48 (31%). The usability of the interface influenced the modification of some terms, as many users accidentally added
extra terms to their working list rather than replacing the term as they intended. One participant noted the problem, but did not know how to fix it, and stated the intention of leaving the extra term on the list even when it was inaccurate. Another participant noticed the extra terms but not until they were added to the working list.

Automated Term Dissection (ATD). Four out of the five participants accepted the hierarchical list format without question. One participant stated a concern that the hierarchical modifiers confused the use of problem subsumption and wanted a line composition format as an alternative format. In 12/48 (25%) of the entered problems, participants mentioned that the composition, while acceptable, was not the way they would have represented the concept. These comments surrounded compositional expressions that may have been too complex, such as “induced (sub) steroids” instead of the preferred phrase “steroid induced” which was not returned from the UMLS. Many users also had confusion about the hierarchical relationship of composed terms and placed the modifiers or qualifiers in the problem list, failing to note that the term modified the wrong concept in the hierarchy. For example, a user would enter the term “fracture” and the term “left foot” in the same problem at the same hierarchical level without noting that “left foot” actually should have modified the term “fracture.”

The Automated Term Dissection (ATD) feature was difficult to evaluate because most participants entered only short search strings of 1-3 words. Participants generally felt that this feature was helpful (8/13 = 62% of case scenarios), but many also indicated that they were not really sure what the feature did. This uncertainty may have stemmed from confusion about the hierarchical relationship of terms as noted above.

Problem Entry. Participants used a variety of search strategies in their problem list entry. Most entered 1-3 word phrases on their initial search, however if they could not find an adequate response they would attempt to enter a more specific phrase. All were willing to choose a term different than original entry, however the threshold for term acceptance varied from participant to participant. Some users spent little time looking for more accurate terms, while others seemed compelled to look at every term on the screen even when they had an exact match automatically returned by the system. Often, the participants expected the related term list to give them a narrowed definition, and would use a general search term to find a close concept. For example, most participants wanted to enter the term “diabetes mellitus” to find “steroid induced glucose intolerance.”

Usability. Although in 85% of the case scenarios participants stated that they found acceptable terms, 3 of the 5 participants spontaneously mentioned that the overall interface was too cumbersome for complex problem entry. 3 of the 5 stated they would use the program reluctantly, and 1 stated that he would refuse to use the program for clinical practice. All agreed that they would use the program only if it were faster at entering complex cases, although most felt that they would become faster with more familiarity of the program. There was also significant confusion about the terminology “modifiers,” “qualifiers,” and “kernel concepts.”

Many different methods were expected to accomplish tasks. For example, some users clearly felt comfortable using dragging or right mouse menu's and used these methods preferentially, while others stated clearly that they would not have expected to be able to accomplish such tasks. Most users commented that they wanted more relevant, limited and prioritized search returns. Several participants also mentioned the need for improved spelling correction and the ability to type abbreviations for search entries.

DISCUSSION

Compositionality has been widely hailed as a likely solution in the development of a meaningful, useful controlled vocabulary yet to date there are no widely accepted tools which allow physicians to actively compose terms for problem entry. The usability tests of the Mayo Problem List Manager represent a step toward this goal, however highlight the need for further work in this area.

The limited nature of this trial should be noted. Although attempts were made to constrain the amount of bias introduced by using realistic test scenarios and environmental time limits, as with other usability studies, this study suffers because of the controlled and therefore artificial nature of the testing environment. With the complex nature of the test problem, more extensive teaching using case examples was required to introduce the concept of the
program quickly. This may have influenced the way a participant uses the program without education. On the other hand, the limited time allotment restricted the response of users who had sufficient training to be fully familiar with the application. The program was written to be an incomplete prototype, and as such general usability may have suffered unnecessarily as routine bugs and interface issues are ironed out. The small sample size gave a wide range of responses indicating that future complex studies should include more participants. The study also may have had significant differences if tested on sub-specialists rather than primary care physicians.

Nonetheless, much useful information was obtained. Automated term composition (ATC) was generally well accepted. Physicians indicated that they generally liked the suggested terms, even while they often modified the suggested composition. Work needs to be done with user directed composition to automatically recognize which terms are being replaced, and which concepts they may modify. Physicians indicated that they were willing to compose terms under some circumstances, however stated a strong desire to do as little typing as possible.

Automated term dissection (ATD) was thought to be helpful, though poor physician understanding of the concept limited our ability to study this feature. The concept of a hierarchical list was acceptable for most users, though one participant did indicate a desire for viewing this information in a line format because it confused the subsumption of problems with the modification of a problem.

Many users noted the causal relationships between the patient’s problems in the scenarios, and 2 of the 5 participants subsumed problems in the problem list. One of the five participants made an effort to put the concept “Coccidiodes immitis pneumonia” under “HIV,” and also attempted to list “diabetic retinopathy” under “diabetes mellitus.”

The usability study highlighted the need for quick and easy problem entry with minimal effort required to search for the correct term. Many participants spontaneously commented on the need for abbreviations such as “CHF,” and the lack of commonly used phrases such as “steroid induced.” Many also commented on the need for more systemization of related terms. These deficiencies are related to problems with the semantic content of the UMLS itself. Many users noted the impracticality of searching long lists of related terms that require the use of scrolling. The variety of methods that our participants used to enter terms highlights the need for multiple ways to accomplish the data entry task by using such methods as a mouse or keyboard, or perhaps other methods such as voice recognition.

The authors believe there is a need and practical answer to user directed composition of controlled health terminologies, but this can only be accomplished with further improvement of the user interface and the underlying terminology.

References

1 Cimino JJ. Distributed Cognition and Knowledge-Based Controlled Medical Terminologies. Artificial Intelligence in Medicine 1998; (2):153-68.