Standardized Problem List Generation, Utilizing the Mayo Canonical Vocabulary Embedded within the Unified Medical Language System

Peter L. Elkin, MD[1], David N. Mohr, MD[1], Mark S. Tuttle[2], William G. Cole, PhD[2], Geoffrey E. Atkin[3], Kevin Keck[2], Thomas B. Fisk[4], Brian H. Kaihoi[4], Karen E Lee[4], Michael C Higgins[5], H. Jaap Suermondt[5], Nels Olson[2], Paul L. Claus, MD[4], Paul C. Carpenter, MD[4], Christopher G. Chute, MD[3]

Departments of Area Internal Medicine[1] and Health Sciences Research[3], Mayo Clinic[4], Rochester, MN
Lexical Technology, Inc.[2], Alameda, CA, and Hewlett Packard Laboratories[5], Palo Alto, CA

Vocabulary: The Mayo problem list vocabulary is a clinically derived lexicon created from the entries made to the Mayo Clinic's Master Sheet Index and the problem list entries made to the Impression/ Report/ Plan section of the Clinical Notes System over the last three years. The vocabulary was reduced by eliminating repetition including lexical variants, spelling errors, and qualifiers (Administrative or Operational terms) [1]. Qualifiers are re-coordinated with other terms, at run-time, which greatly increased the number of input strings which our system is capable of recognizing.

Implementation: The Problem Manager is implemented using standard windows tools in a Windows NT™ environment. The interface is designed using Object Pascal. HTTP calls are passed over the World Wide Web to a UNIX based vocabulary server. The server returns a document, which is read into Object Pascal structures, parsed, filtered and displayed.

Study: This paper reports the results of a recent Usability Trial focused on assessing the viability of this mechanism for standardized problem entry. Eight clinicians engaged in eleven scenarios and responded as to their satisfaction with the systems performance. These responses were observed, videotaped and tabulated. Clinicians in this study were able to find acceptable diagnoses in 91.1% of the scenarios. The response time was acceptable in 92.5% of the scenarios. The presentation of related terms was stated to be useful in at least one scenario by seven of the eight participants. All clinicians wanted to make use of shortcuts which would minimize the amount of typing necessary to encode the concept they were searching for (e.g. Abbreviations, Word Completion).

Conclusions: Clinicians are willing to choose a canonical term from a suggested list (as opposed to their own wording). Clinicians want an "intelligent" system, which would suggest terms within a category (e.g. Types of "Migraine"). They are able to make functional use of our system, in its current state of development. Finally, all clinicians appreciate the value of encoding their problems in a standardized vocabulary, toward improved research, education and practice.

INTRODUCTION

In this study, we report the evaluation at the Mayo Clinic of our clinically derived lexicon, and a problem list entry tool, which utilizes a remote vocabulary server to serve up this vocabulary to clinicians. Clinically derived lexicons are particularly useful when one wishes to provide a clinically oriented retrieval set for any given query [2]. This tool was evaluated by eight clinicians that were the subjects of a protocol executed within the Mayo Clinic's Usability Laboratory. The results of this study are presented in this manuscript.

The Mayo Clinic's Electronic Medical Records (EMR) project is now in its third year of development. As part of this effort, the Vocabulary Committee has had the charge of deciding what to use for its canonical vocabulary. The Mayo Clinic has a long history of exacting manual record keeping. This is exemplified in the indices or vocabulary lists that it maintains. Our vocabulary was chosen from the compilation (the Logical Union) of the Mayo Clinic's Master Sheet Index, and Clinical Notes System, which has been used clinically for three years at the Mayo Clinic.

The Master Sheet Index relates to Mayo's view of a clinical episode. At Mayo, an episode of care is defined at the divisional level and is related to the notion that there is a point in the time course of caring for a particular patient when the case is stable and is therefore no longer in flux. Often this is after the diagnoses have been established, or after the major interventions have been accomplished. At this point in time the Primary Physician caring for the patient is required to list the "Final" Diagnoses for this episode of care. This requires the physician to perform the actions of Filtering, Subsumption, and

1091—8280/97/$5.00 © 1997 AMIA, Inc.

Prioritization. Filtering would include eliminating diagnoses originally contained in one's differential diagnosis, that were ruled out during this episode. Subsumption allows the clinician to state that the patient's presenting problem of Chest Pain was in fact due to Atherosclerotic Coronary Artery Disease (CAD)[3]. Prioritization relates to the notion that although the patient's problem list may in fact be very long, the clinician may have only dealt with a subset of the patient's problems during this episode. For example, our patient with Chest Pain may also have Seborrheic Keratosis, which we may not have had occasion to discuss during the work up and treatment of this patient's CAD.

The Clinical Notes module requires clinicians to enter diagnoses into an Impression/Report/Plan section of the dictated clinical note. The information is stored electronically and is available as a resource at the Mayo Clinic. This is available for research purposes, but perhaps has its greatest utility as a clinically derived lexicon, which directly represents the strings which clinicians in a busy multispeciality group practice are currently employing in their day to day patient care. The Clinical Notes module is one of the step projects in our journey toward an integrated electronic medical record.

The first five-thousand terms by frequency of occurrence recorded in the Master Sheet Index over the last three years were combined with the first five-thousand terms, again by frequency of occurrence of the diagnoses, recorded in the Impression/Report/ Plan section of the clinical notes database. This result formed the core dataset of terminology for the Mayo Problem List Manager Lexicon. Our goal in creating this vocabulary is to provide our clinicians with a focused input to a controlled vocabulary for the purpose of encoding diagnoses. Since this list of diagnoses also includes the clinician's working list, many manifestations must also be included within our working environment. Currently this is being utilized by focused prioritization as a contributing vocabulary to the UMLS[4]. The YATN (Yet Another Terminology Navigator) tool created by Lexical Technology, Inc. in collaboration with the Mayo Clinic and Harvard's Beth Israel Hospital, is the retrieval engine being utilized for this project [5]. Our view is that our clinically oriented lexicon will give the YATN retrievals a much greater specificity or precision, without sacrificing much sensitivity or recall.

The utility of a vocabulary server is best tested by its use in capturing clinical information at the point of care. In our Usability Trial, we have attempted to simulate a broad number of scenarios, which represent the variety in a clinician's day. These included diagnostic dilemmas, straight forward cases, and encounters with many diagnoses. We were able to observe the physician's behavior, monitor their successes and frustrations, and to tabulate their responses to specific inquiries.

SPECIFIC AIMS

1. To determine if our vocabulary server can provide acceptable encoded terms for a clinician's problem list.
2. To determine if a vocabulary server can respond rapidly enough to meet the demands of a busy clinical environment.
3. To determine if the clinicians find the additional effort of entering an encoded problem list acceptable.
4. To determine what is the appropriate number of suggested encoded terms to be displayed to the clinician.
5. To determine if for any given encoded term, is it helpful to display a list of more specific or other related terms.
6. To determine if clinicians will take advantage of shortcuts in problem name entry (e.g. Word Completion or Abbreviations).

METHODS

A Usability study evaluates how a particular process or product works for individuals [6,7]. Optimally one would test a population of individuals who are a sample of typical users of the type of process or product being tested. It should be stated clearly to participants that the purpose of the study is to evaluate the process or product and not the individual participant. Usability sessions are videotaped from multiple angles (including the computer's screen image) and participants are encouraged to share their thoughts verbally as they progress through the scenarios provided ("think aloud"). This helps to define the participants' behavior in terms of both their intentions and their actions. For example, in our study, we had the user identify which terms they were looking for before they initiated their search. We could monitor what was entered into the program and we were able to view the retrieval list, the term finally selected. Then we compared both the degree to which the clinician-user, felt that they were satisfied with the choice that they had made, and how often what they eventually chose was the same as what they stated in advance was the optimal description of their patient's problem.

The Usability Laboratory at the Mayo Clinic is a suite of rooms, which provides space for study planning, execution and review. There is a conference room with white board space for planning

and evaluation. The facility utilized for executing the study includes the study room, a control room and a developer's viewing booth. The study room is a space, which in our study included a desk and chair with a computer and screen, keyboard and mouse on the desk. There are cameras on each of three corners of the room and the back wall is a one way mirror. The user sits in this space and works on the scenarios provided by the study team, after a short introduction to the facility and purpose of the study by the study director (who is not part of the development team). Behind the one way mirror is a soundproof room with multiple monitors and video recording equipment. The control person directs the videotaping from the available source input (including a video input from the screen). The study director has a microphone with which they can communicate with the study participant. The development team, if present, sits in a third space separated by a soundproof division and is located behind the control room. In this space, the development team has no contact with the participant but can easily observe the study and gain direct experience with the user's interaction with their tool.

In our study, we involved eight clinicians from various specialties, each engaged in eleven scenarios using the Mayo Clinic's problem list entry tool. They were all active clinicians. There was one woman and seven men. Their ages ranged from the late thirties to the late fifties. One participant had been at Mayo for only one and a half years, two for six to ten years, two for eleven to fifteen years, two for twenty-one to twenty-five years and one participant had been with Mayo for over twenty-five years. All participants had used a computer before. Five of the eight participants used a computer outside of their employment. Their specialties were a Pediatrician, two Hematologists, an Obstetrician, three General Internists, and a Vascular specialist. The study was authored by two General Internists and an administrator and consisted of two introductory scenarios followed by nine more free-form scenarios, of which five were free-form case presentations where no terms were suggested by the scenario's author.

RESULTS

Cue Card

Question #1: Did you find acceptable clinical diagnoses using the Mayo vocabulary navigation system?

91.1% of the time they did

Question #2: Was the response time of the system acceptable?

92.5% of the time it was

Question #3: Given that there are both clinical and research benefits to having a standardized representation of our patients' problems. Was the effort required to make use of this system acceptable?

88.9% of the time it was
Range 60% to 100%

Question #4: Were the number of terms presented, Just Right, Too Many, or Too Few?

Just Right 71.4%
Too Few15.7%
Too Many 12.9%

Question #5: Did you find the presentation of Related Terms to be helpful?

87.5% Found them useful

Question #6: Did you find the ability to enter abbreviations or word fragments (i.e. "p.n.* pn*") helpful?

100% Found them useful
70.7% of the time

Traditional Usability methodology was abandoned for portions of this study to answer questions about specific functions and methods. In traditional usability trials, the user is left to make use of the system with little or no guidance. In scenarios 1 through 5, and 11 we instructed the user and then tried to answer a specific question. In scenario 1, we wanted to see if the clinician would pick a term off a list or the free text that they had entered, or will they resist the imposition of controlled terminology (i.e. "you can't put words in my mouth"). All clinicians studied were willing to choose the canonical term *Congestive Heart Failure* over the entry *chf.* In scenario 2, we wanted to see if a clinician could reliably find an entry on a list if it was not the first one and was of similar string length to the other terms. Uniformly they could. In scenario 3, we had a term, which was close, but not identical in meaning in the Past History (Coronary Artery Disease) for Ischemic Heart Disease. Most clinicians were satisfied with the term Coronary Artery Disease. In scenario 4, we had them enter a problem, which had a long retrieval time (average one to two second delays). None of the participants thought that the response time was inordinately long. In scenario 5, we looked at the behavior of clinicians when entering very long lists of diagnoses. Here we found that clinicians did grow tired of the task and looked for and used shortcuts for finding terms to speed up their task. One of the eight participants thought that this might be too much work in a busy clinical situation. In scenario 11, we were looking to see if clinicians

would recognize the benefits at a follow-up visit of taking the time to encode the diagnoses at the initial visit. Specifically that once encoded, these diagnosis could be attached to a visit from the Historical Problem list, without the clinician needing to reenter them. All participants immediately appreciated the value of the Historical Encoded Problem List.

In scenario 6, the user was presented with a case of a patient with "Acute Cystitis". We wanted to determine if there would be variability in the way clinicians encode this data. The encoding varied from UTI to Cystitis to Urinary Infection to Lower Urinary Tract infection. This variability speaks for the need for rich hierarchies within the Mayo vocabulary and also for the recognition of lexically differing words with the same meaning (synonymy), in order to be able to retrieve this case regardless of how a clinician would choose to frame the query (problem). Scenarios 7 and 8, tested the behavior of clinicians when entering rare and common problems. What we learned, was that their behavior differed not by whether the problem was rare but by whether they knew the unifying diagnosis or not. Scenarios 9 and 10, looked specifically at whether the related terms (most were narrower than terms) were useful. We found utility both in that this function limited the typing required to enter a problem and also it served as a reminder system for clinicians, by suggesting what some of the diagnoses were within a particular category of pathology (e.g. Types of "Fractures").

The system provided acceptable clinical diagnoses in 91.1% of the cases. This indicates that the users of this system would fail to encode only approximately 9% of their diagnoses. The system provided adequate response time in 92.5% of the scenarios. All clinicians felt that the effort required to encode problems was worth it. In 11.1% of the scenarios they felt the effort to encode all of the problems was onerous. There were enough terms (Recall or Sensitivity of the system) in 84.3% of the scenarios. The precision of the system was between 71.4% and 87.1%. Only one of the eight clinicians did not find the presentation of related terms to be helpful in at least one scenario. All of the clinicians made use of short cuts to save them from having to type out their entire entry. The three shortcuts provided by the system were the use of abbreviations, word fragment completion (i.e. "pn*" would return all matches starting with the letters "pn"), and related terms (e.g. entering "dm" would return "Diabetes Mellitus" and all the types of Diabetes Mellitus in the related terms box [e.g. "Adult Onset Diabetes Mellitus"]). They actually applied these techniques in over two-thirds of the scenarios.

Limitations

Our study had several recognized limitations that impact on our ability to generalize our results. First, all participants voluntarily agreed to participate. This may imply that their interest level and motivation are high. Therefore, these results may not be generalizable to reluctant users. Although participants were encouraged to act as if they had normal clinical time pressures, in fact they did not. Consequently, the results might vary if the study had included normal workflow interruptions and time pressures. All participants were new to the system. This eliminates bias, but limits us in generalizing about the systems suitability for experienced users. All participants were practicing staff clinicians, which was our target audience.

DISCUSSION

We have described our vocabulary effort at the Mayo Clinic, where we took terms from a clinically derived lexicon and processed the input to filter redundancy, abbreviations, and qualifiers. By utilizing Mayo's clinically focused lexicon as a structured entry point into the UMLS, we provide a powerful tool for focusing the YATN retrievals to provide a highly relevant retrieval list for a problem list entry tool. Further research should include work on the ontology of modifiers and the relationships, which can exist between modifiers and canonical terms and between modifiers themselves. We are also engaged in identifying the relationships, which exist within the Mayo problem list vocabulary. We plan to use these relationships to facilitate a browsing function, which would allow users to view multiple layers of related terms from within the UMLS and Mayo lexicons.

The Usability Trial verified the necessity of a robust and broad-based approach to the entry of structured vocabulary. Clinicians required a short but relevant retrieval list, generated with the fewest key strokes possible and without the need to enter redundant terms (i.e. If the term was already on the Historical Problem List the clinician did not want to have to reenter it.). It was clear that the clinicians were able to find acceptable diagnoses using the system. Overall they felt that the response time afforded by this approach was acceptable. Although only further testing will ensure that this result is scalable (i.e. that the response time will remain adequate as we increase the number of simultaneous users). We were further able to define many strengths and some weaknesses of our Graphical User Interface (GUI) design. By analyzing the weaknesses we will be able to create a stronger implementation system, thereby validating the methodology.

After analyzing the utility of the study design, the investigators unanimously felt that free form cases were more useful than directed entry of named problems. Universally the clinician's found the exercise more clinically relevant and were observed to take more care with the entry of the problem names. Future usability study designs will involve a higher concentration of free form case based scenarios.

This interface design allows the flexibility available within the Windows NT™ environment combined with the flexibility of the World Wide Web. This paradigm provides for powerful vocabulary processing on a remote server to be combined with a GUI without a decrement in response time. This design supports the intense processing necessary to provide clinicians with a flexible mode of entry of a structured vocabulary for patient care, education, and research. This mode of entry via a common standardized vocabulary engine will set the stage for multicenter collaborative trials, meaningful outcomes studies and utilization reviews. We must all work toward commonality in our coding systems. As we create these systems, we must be certain to plan on capturing what is actually happening in our practices to a clinically relevant level of granularity. This requires a strong, clinically oriented vocabulary, with an excellent structure, which reliably discourages redundancy, and error, and provides the backbone and methods necessary to maintain the data and structure over the months and years to come.

Acknowledgements

This work was funded in part by NIH grants HS/LM08751, LM05416, and AR30582.

References

1. Chute CG, Elkin PL. Reduction of Qualifiers from a Clinically Derived Lexicon. Submitted to Masys (ed), Proceedings of the Twenty-first Annual Symposium of Computer Applications in Medical Care, 1997.
2. Rassinoux AM, Miller RA, Baud R, Scherrer JR. Modeling just the important and relevant concepts in medicine for medical language understanding: A survey of the issues. In Chute C (ed) Proceedings of the IMIA WG-6, 1997, 53-68.
3. Tuttle MS, Sheretz DD, Erlbaum MS, Sperzel WD, Fuller LF, Olson NE, Nelson SJ, Cimino JJ, Chute CG. Adding your terms and relationships to the UMLS Metathesaurus. Proceedings of the Annual Symposium on Computer Applications in Medical Care. 1991, 219-23.
4. Baud R, Lovis C, Rassinoux AM, Scherrer JR. Alternate Ways of Knowledge Collection, Indexing and Robust Language Retrieval. . In Chute C(ed) Proceedings of the IMIA WG-6, 1997, 81-93.
5. Yang Y, Chute CG. A schematic analysis of the Unified Medical Language System. Proceedings of the Annual Symposium on Computer Applications in Medical Care. 1991, 204-8.
6. Weir C, Lincoln MJ, Green J. Usability testing as evaluation: Development of a tool. In Cimino JJ (ed), Proceedings of the Twentieth Annual Symposium of Computer Applications in Medical Care, 1996, 870.
7. Kushniruk A, Patel V, Cimino JJ, Barrows R. Cognitive evaluation of the user interface and vocabulary of an outpatient information system. In Cimino JJ (ed), Proceedings of the Twentieth Annual Symposium of Computer Applications in Medical Care, 1996, 22-26.