Controlled Vocabulary and Design of Laboratory Results Displays

Gai Elhanan, MD, James J. Cimino, MD
Department of Medical Informatics
Columbia-Presbyterian Medical Center
New York, New York

Traditional data-review displays are driven by the ancillary systems that produced the data. A different paradigm is being used at Columbia-Presbyterian Medical Center (CPMC) where a controlled medical vocabulary - the Medical Entities Dictionary (MED) is the driving force behind laboratory data-review displays. Using hierarchical and semantic networks the authors have constructed a Web-based tool that considerably simplifies the MED-editing task required to create new displays. The tool uses knowledge in the MED to extract contextually relevant hierarchic and semantic sub-nets from the MED. The tool has a sensitivity of 92.2% and a relevance of 94.7% for retrieval of terms from the MED. Based on these results and given sufficient domains' structure within controlled vocabularies, we conclude that similar algorithms will enable applications to design and generate customized displays on-the-fly.

INTRODUCTION

Clinical Information Systems (CIS) are moving from ancillary-oriented to repository-oriented architectures and, as a result, those charged with developing applications such as patient data displays are relatively distant from the systems where the data originate. This distance can lead to a lack of communication between those who might change the source system and those who attempt to display its data.

Consider, for example, the application developer who must create a report showing a particular set of a patient's laboratory results. The developer is likely to determine a list of codes which represent the data of interest and use that list in querying the central repository. If the laboratory personnel add a new test code, and the application developer is unaware of this change, then the report will be incorrect. In order to correct the problem, either the ancillary system must notify the application developer, or the application developer must constantly monitor the ancillary system's vocabulary. Neither of these tasks is easily accomplished.

We refer to the above report program as an example of a *vocabulary-driven* application, since it relies on a predefined list of coded terms. We postulate that we can build a tool to assist the development of vocabulary-driven applications which is itself vocabulary-driven. In this way, developers will better understand the vocabulary that drives their applications during both the creation and maintenance processes. We use a knowledge-based vocabulary tool for supporting a laboratory data display application. This paper describes our approach and results of a preliminary evaluation of the tool's effectiveness.

BACKGROUND

A controlled medical vocabulary - the Medical Entities Dictionary[1] (MED) is an integral part of Columbia-Presbyterian Medical Center's (CPMC) CIS[2] infrastructure. The MED provides translation capabilities to and from coded data and within its own structure. Each MED term has a unique integer code (MEDcode) as well as data about its name, synonyms, units and other information relevant to each term, all contained in unique sets of attribute fields (slots) to create a frame representation of the term. The various frames are linked to create a semantic network that relates terms in a variety of meaningful ways. The MED also supports a classification structure that allows multiple hierarchies.

At CPMC, laboratory results are stored in the central data repository using the MED. Specific tests at CPMC cannot be ordered directly, but rather as components of orderable procedures, unless specifically defined as both a test and a procedure. Tests can be grouped together to form orderable laboratory procedures. The organization of laboratory procedures in the MED preserves this scheme and is expanded even further to conceptually group tests and orderable procedures into multiple level tests- and procedure-classes. Members may be classified into more than one class. Currently the MED holds more than 650 orderable procedures and more than 1866 tests (of which 716 are orderable). One of the slots that is instantiated for tests in the MED is slot 16

1091—8280/97/$5.00

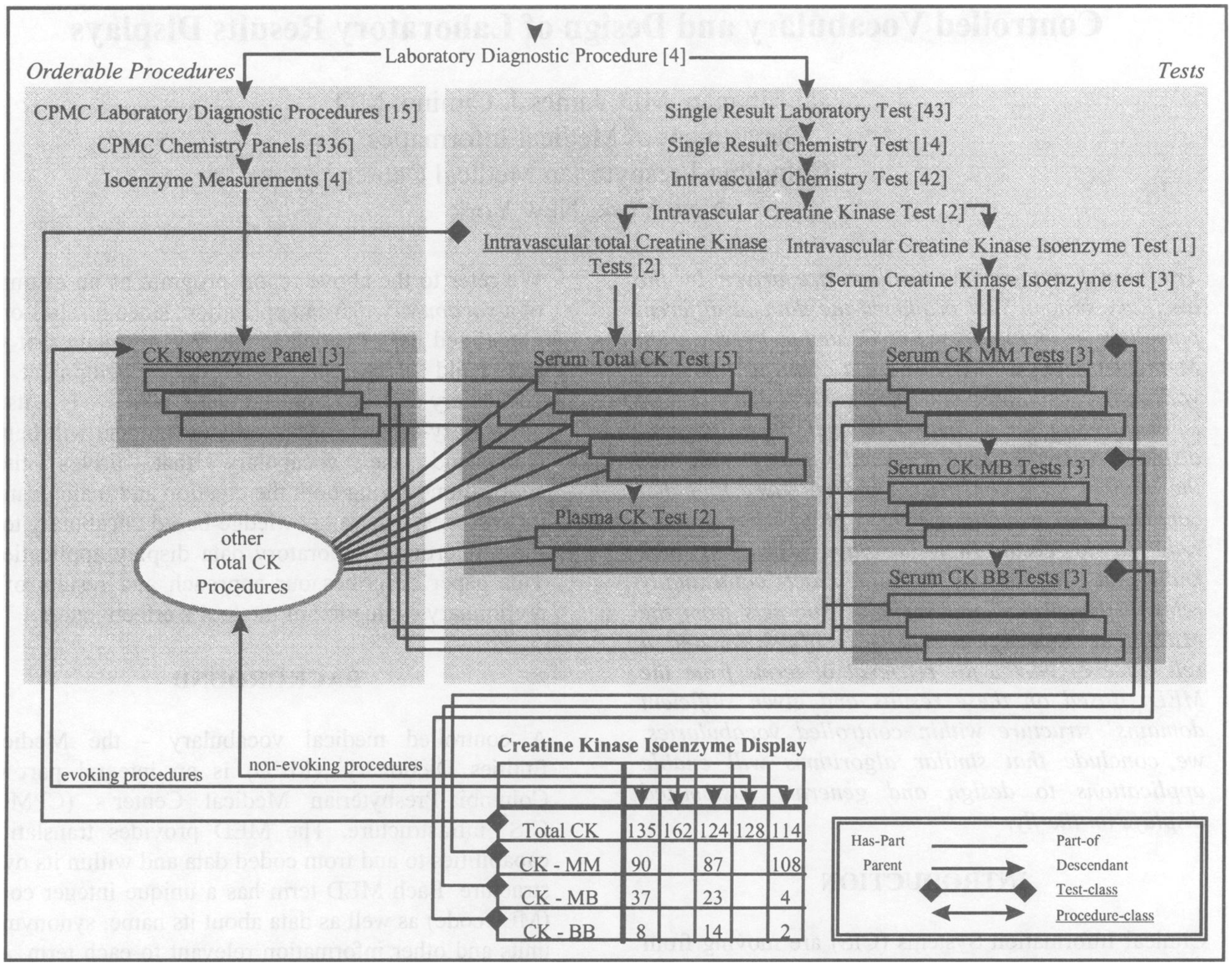

Figure 1: Simplified schematic diagram of a vocabulary-driven display. The Creatine Kinase Isoenzyme Display will be invoked by any of the CK Isoenzyme panels. Once invoked, it will also display any available results for Total CK procedures, given available data for a specific patient. For example, members of the "20-Test Chemistry Panel" procedure-class (see Figure 2) will not evoke a CK Isoenzyme Display but the result of their CK component will be included if another panel invoked the display (Numbers in brackets signify number of direct descendants).

which holds the value for the {SUBSTANCE-MEASURED} by that test. The values in that slot are MEDcodes which form their own MED hierarchy. Test classes can also have slot 16 instantiated, while pure orderable procedures and their classes do not have that slot as part of their frame.

Some of CPMC's user interfaces are utilizing a novel paradigm to display patients' laboratory data[3]. According to that paradigm (see Figure 1), laboratory data displays are driven by the controlled vocabulary, the MED. Laboratory results data are being displayed in groups of interest (e.g. Therapeutic Drug Levels, Tumor Markers, Viral Serologies...) by different displays. Display definitions (Ddef) are stored in the MED as frames. Each frame is a MEDcode that defines what procedures are part of that display, what tests should be displayed, and their order. These values are entered as MEDcodes into the appropriate slots of the Ddef and have reciprocal values entered in appropriate slots of the procedures and tests included in that display. One additional advantage of the paradigm is that orderable procedures can be defined as display-invoking (e.g. the presence of a result for a procedure from that procedure-class will cause the application to generate that display) or as also_displayed_by (e.g., if the display had been invoked by another procedure it will also display results from that procedure-class in that display). The dynamic list of available displays is presented to the user according to the available patient data, and once a display is selected it will be generated on-the-fly according to its Ddef in the MED. The above scheme allows for multiple views of the same data within one application (e.g. show me serum_enzymes_display vs. show me cardiac_profile_display) or different views across applications. This scheme simplifies updates; if a new test or procedure is added to the MED, as long as it is appropriately linked to the correct test-class and/or procedure-class, the relevant Ddef will be updated automatically with no further need for intervention.

While the process generating the actual display is fully automated, the process by which Ddefs are created in the MED is manual and labor-intensive. The MED currently holds almost 50,000 terms in a multitude of hierarchies. Finding all appropriate instances of procedures and tests and deducing whether an appropriate class exists or need be created is the weak link in this whole chain. Even with the MED's highly hierarchical structure and the available (AccessMED) editing tools[4], this task has proven to be quite difficult when comprehensiveness and completeness are the required end-points for each display[5]. We used the MED's intricate structure to extract semantic and hierarchic sub-nets that enabled us to create a semi-automated process by which display frames are generated with minimal human intervention.

METHODS

Using {IS-A} relationships and semantic links in the MED (slot 16), an algorithm was devised by which test-classes appropriate for display rows are elucidated. The algorithm recursively checks for the {SUBSTANCE-MEASURED} of parents while verifying the consistency of the same slot for all descendants. Once the test-classes are deduced the algorithm uses the {PART-OF} on the tests side and the {HAS-PART} on the procedures side to collect all the relevant procedures associated with the test-classes. A Web-based interface allows the user to review and evaluate the results (see Figure 2). The application input is any combination of tests, test-classes, orderable procedures, procedure-classes or pre-defined display MEDcodes. For each input element its test components' leaf-node descendants are abstracted and used as input MEDcodes for the algorithm.

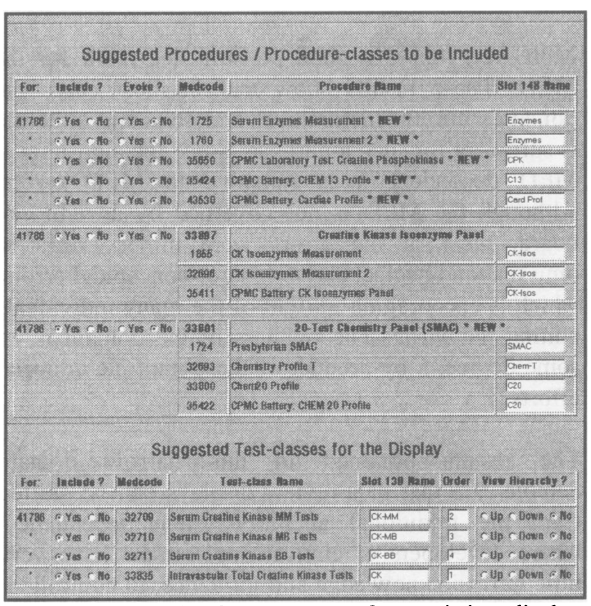

Figure 2 : Results for an entry of an existing display: 41786 - Creatine Kinase Isoenzyme Display. Four test-classes and 12 procedures were correctly identified. Nine procedures are correct new additions. All procedures contained within procedure-class 33897 will invoke the display while the other are also_displayed_by only.

The application output is an ASCII file that is used to generate a new MED version containing the new and modified Ddefs. No intervention is required in any of the applications that use the Ddefs; once a new Ddef is in the MED, it will be displayed given that that patient has data for that display.

Evaluation

General procedure retrieval performance:
Utilizing randomly selected single tests, procedures or panels (15 each), the algorithm's aggregate leaf-node procedure retrieval results were compared to human MED search results using the MED editor search tools (a total of 166 leaf-node procedures).

Algorithm vs. existing displays:
Fourteen existing Ddefs that were manually refined over time but not previously modified by the algorithm were used as input. The results obtained by the algorithm were compared to each input display that thus served as its own historical control. The displays represent 20% of the currently available displays and contain 352 procedures in 77 procedure-classes and 1031 tests in 199 test-classes. The least complicated display had 3 procedures in one procedure-class and 10 tests in 3 test-classes, while the most complex had 84 procedures in 22 procedure-classes and 188 tests in 38 test-classes.

Procedures retrieval - as measured by the aggregate number of leaf-node procedures retrieved. Newly added procedures were assessed as either true (missed by controls) or false positive errors that were divided into a) Major - conceptually irrelevant or b) Minor - not relevant in the context of the input MEDcode.

The following definitions were used for procedure retrieval:

- *Sensitivity* - percent of correct procedures retrieved compared to the relevant standard in each case (total number of leaf-node procedures involved).
- *Relevance* - percent of retrieved procedures that should have been included in a display.

Elucidation of test-classes - the algorithm's ability to arrive to a correct test-class. Test-class errors were divided into a) Major - requires opening a MED-browsing mode to correct (i.e., false negative) or b) Minor - needs only marking out with no browsing of the MED (i.e., false positive).

RESULTS

General procedure retrieval performance - Overall, the algorithm identified 153 out of 166 procedures (92.2% sensitivity) of which 145 (94.7%) were relevant. No significant differences existed between the different types of input laboratory procedures.

Algorithm vs. existing displays -

Procedures retrieval - The algorithm did not miss a single procedure that should have been included in any of the control displays. 72 procedures and procedure-classes were found to have been correctly added to 8 of the control displays (20.4% of the original 352 procedures). Of the added procedures, 12 were significant concepts while the rest were categorized as also_displayed_by procedures. 23 (6.5%) procedures and procedure-classes were found to be false positives in 5 displays leading to a 93.5% relevance. Of these false positives, only one was a major error.

Elucidation of test-classes - As a result of added procedures 3 test-classes were identified as necessary additions to two of the control displays. Twenty four major test-class errors (12.1% false negative) were made in eight displays, and nine minor errors (4.5% false positive) were made in 5 displays. Overall, test-class errors involved 46 tests (4.4% out of 1031) in nine displays.

Estimated time savings - for small-sized displays containing up to 5 test-classes it is estimated that the algorithm could have saved between 1 to 5 hours of work. For larger displays containing above 10 test-classes it is estimated that time savings range from 7 to 24 hours.

Overall, four out of the 14 Ddefs generated by the application were without a single error. Analysis of the cases with erroneous procedures and test classes from both sets of evaluations showed that the errors originated mainly due to errors in slot 16 ({SUBSTANCE-MEASURED}) declarations in the MED; either missing values, lack of granularity or as a result of use of a wrong substance MEDcode for a specific test.

DISCUSSION

This paper shows how a controlled vocabulary, such as the MED, can be utilized for purposes other than its translational and coding capabilities. The hierarchical structure of the MED, coupled with semantic links, enabled us to devise our algorithm and extract the appropriate semantic and hierarchic sub-nets and deduce the correct test- and procedure-classes. Overall, the algorithm performed with more than acceptable sensitivity and relevance, and for procedures performed better than the refined human effort as expressed by the control displays. It should be noted that these results are far superior to first-pass results obtained by moderately-experienced users searching the same domain using AccessMED (18% completeness)[5]. We believe that such a wide discrepancy resulted from applying a customized tool (the algorithm) as compared to a general tool (AccessMED).

Moreover, this is a good performance mark for the MED. These results show that in most cases the domain's current hierarchical and semantic structure is sufficient for such a purpose. Since the algorithm is MED dependent it is sensitive to MED errors. Although the MED is not governed by an editorial board, most of the errors were minor errors of omission and neglect and not major conceptual errors, which is encouraging in light of the many individuals editing the MED. The richness of the domain is, to some degree, a by-product of these multiple editorial efforts.

The results obtained in this narrow domain demonstrate that data-review displays can be shifted from the traditional ancillary-system dependency, where developers of clinical applications have very little intrinsic knowledge and influence, to controlled-vocabulary driven systems, in which the controlled vocabulary contains all the pertinent conceptual data and is independent of the ancillary system. Such a paradigm is already in place in some of the applications at CPMC, where pre-defined MEDcodes are used to generate data-review displays on-the-fly.

Moreover, the current effort demonstrates that in well organized and semantically rich domains the latter step can be by-passed. Applications can use similarly constructed algorithms to conceptually design displays on-the-fly; applications will scan available patient data and use the conceptual structure embedded in the controlled vocabulary to design and generate the display with no pre-defined display concepts. Moreover, vocabulary-driven displays have significant advantages over the traditional ancillary-system dependency whenever maintenance is an issue. As long as new procedures and tests are added into their respective appropriate structures within the vocabulary, they will be displayed by the appropriate display and within the right concept, with no further updating. Although our current algorithm is only semi-automated, the estimated time-savings incurred by it are significant and serve to demonstrate the possible maintenance advantage.

As controlled vocabularies will be expanded to answer for similar needs in a comprehensive fashion for more and more domains, they will gradually become more of a knowledge-base than a controlled vocabulary by its narrow definition. Such structures are much like metadatabases[6,7] and can perform many of their roles. We believe that this approach of vocabulary-driven application development will extend to work with emerging standard vocabularies, especially where those vocabularies are knowledge-based; for example, the LOINC vocabulary contains definitional information similar to the MED for describing laboratory test terms[8] which could be used to drive our tool.

CONCLUSIONS

Vocabulary-driven displays require much work on the vocabulary side in order to achieve an appropriate level of conceptual representation of the knowledge contained within a domain, but this appears to be time well spent. Given appropriate hierarchical and semantic representation, such vocabularies can be used to automatically design and generate data-review displays.

Access to site

http://www.cpmc.columbia.edu/homepages/elhanan/Display/webdisplay.cgi

Acknowledgment

Dr. Elhanan is supported by a postdoctoral training grant number LM07079 from the National Library of Medicine.

References

1. Cimino JJ, Clayton PD, Hripcsak G, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical terminology. *JAMIA*, 1994; 1:35-50.
2. Clayton PD, Sideli RV, Sengupta S. Open architecture and integrated information at Columbia-Presbyterian Medical Center. *MD Comput*, 1992; 9(5):297-303.
3. Barrows RC, Allen BA, Smith KC, Arni VV, Sherman E. A decision-supported outpatient practice system. In: Cimino JJ, editor. *Proceedings of the 1996 AMIA Fall Symposium*; 792-796.
4. Allen BA, Barrows RC, Desai N. AccessMED, a tool for browsing, searching and maintaining a large controlled medical vocabulary. In: Cimino JJ, editor. *Proceedings of the 1996 AMIA Fall Symposium*; 940.
5. Hripcsak G, Allen B, Cimino JJ, Lee R. Access to data: comparing AccessMED with Query by Review. *JAMIA*, 1996; 3:288-299.
6. Eccles JR, Saldanha JW. Metadata-based generation and management of knowledgebases from molecular biological databases. *Comput Methods Programs Biomed*, 1990; 32(2):115-23.
7. Long JM. The POSCH data processing experience. The problem of metadata. *J Medical Systems*, 1986; 10(2):173-183.
8. Forrey AW, McDonald CJ, DeMoor G et al. Logical observation identifier names and codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results. *Clinical Chemistry*, 1996; 42(1):81-90.