Supporting Infobuttons with Terminological Knowledge

James J. Cimino, M.D., Gai Elhanan, M.D., Qing Zeng, M.S.
Department of Medical Informatics
Columbia University, New York, New York

We have developed several prototype applications which integrate clinical systems with on-line information resources by using patient data to drive queries in response to user information needs. We refer to these collectively as infobuttons because they are evoked with a minimum of keyboard entry. We make use of knowledge in our terminology, the Medical Entities Dictionary (MED) to assist with the selection of appropriate queries and resources, as well as the translation of patient data to forms recognized by the resources. This paper describes the kinds of knowledge in the MED, including literal attributes, hierarchical links and other semantic links, and how this knowledge is used in system integration.

INTRODUCTION

There are many ways and degrees by which on-line medical information resources can be integrated with clinical information systems. We believe that one interesting and useful approach is to use information about individual patients to suggest questions which might be asked about those patients, leading to assistance with resource selection and querying. One such project, Hepatopix, was developed in 1989 by Powsner and colleagues to execute Medline searches based on findings in liver biopsy pathology reports.[1] Since then, many other applications have been developed which take a similar approach: terms identified in a clinical record are extracted and shown to the user, the user identifies one or more terms of interest, the application translates these terms to forms suitable for searching (usually to MeSH[2]), and the terms are used for searching (usually in Medline[3]). Eleven such applications were reviewed recently.[4]

A key aspect of all of these applications is the process of translation of clinical terminology from its source form (the clinical record) to a target form (i.e., a form which will be recognized by the desired information resource). Most such applications attempt a direct one-to-one translation; however, this may not be possible in many cases. For example, the Medline Button,[5] which translated patient discharge diagnoses and procedure codes from ICD9-CM[6] to MeSH, found that even using the Unified Medical Language System (UMLS),[7] appropriate translations could only be found for one third of the ICD9-CM terms.

Other difficulties arise when the true concept of interest is not the verbatim term in the clinical record but rather one which is evoked by the clinical term. For example, if a clinician reads a record and finds that a the result of an antinuclear antibody test is abnormal, the question raised may not be about the test per se or even about antinuclear antibody, but instead about systemic lupus erythematosus.

An additional aspect to be considered in such applications is the type of query posed to the information source. The simplest form, "Tell me about..." may be easy to perform but relatively unrewarding. More specific questions may be useful in some cases and inappropriate in others. For example, a query to answer "What's the treatment for...?" will be relevant when the clinical concept of interest is a disease, but not when it is a medication. Applications which tend to ask sensible, useful questions will be more appreciated than those which do not. Yet, knowing what question to ask is not necessarily straightforward.

Considering these and other complexities, bridging the gap from clinical record to information resource may therefore not be as simple as passing patient data to a search engine and hoping for the best. However, when done successfully, the result is an application in which the user need merely point to a few terms and/or questions of interest to have a query performed. We refer to such applications as *information buttons* or *infobuttons* for short. This paper describes some of the strategies we are employing, based on terminological knowledge, to build these buttons.

BACKGROUND

Patient data in the Columbia-Presbyterian Medical Center (CPMC) Clinical Information System (CIS) are coded using the Medial Entities Dictionary (MED).[8] The MED is a concept-oriented controlled medical terminology in which each term corresponds to a node in a semantic network. Terms may have one or more parent terms in the MED's multiple hierarchy, and each term has slots which it inherits from these parents. The slots may have locally assigned values or inherited values. When the values in the slots are literal (string-valued) the slots are referred to as attributes. When the values are controlled terms (i.e., other terms in the

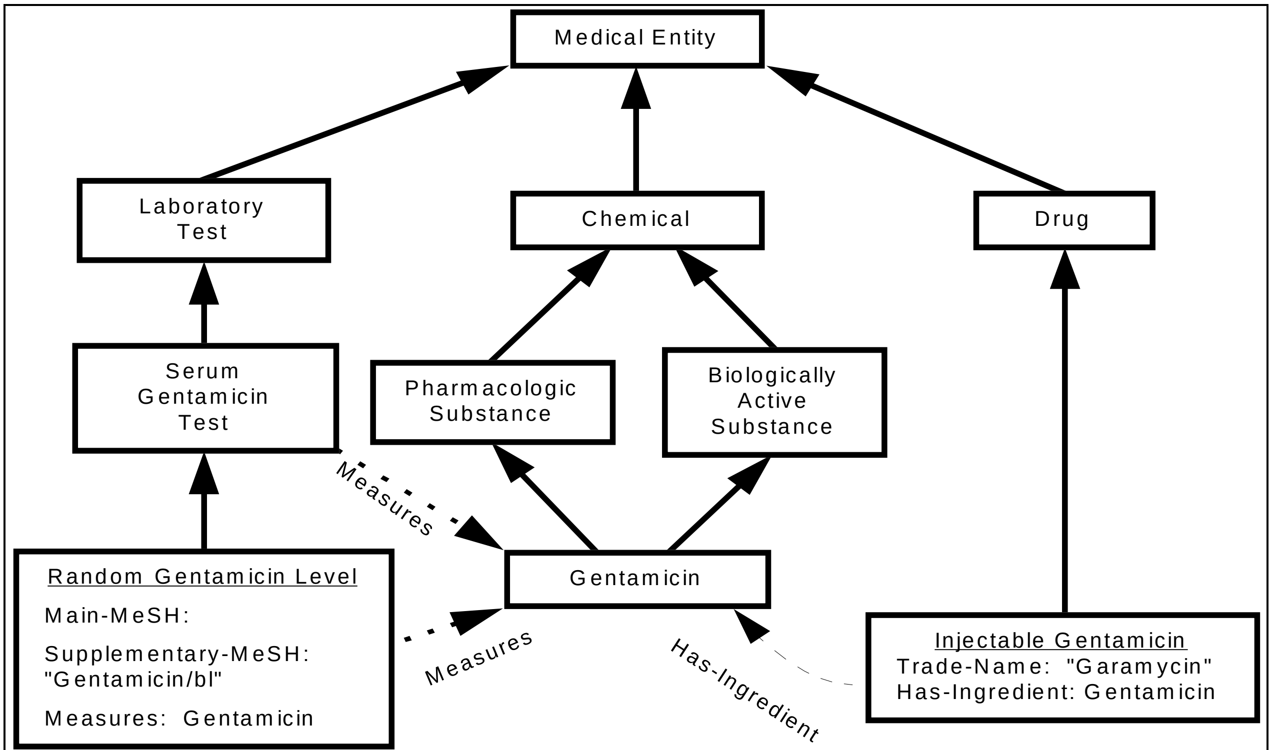

Figure 1: A simplified view of ten terms in the MED, including literal attributes (in quotes), hierarchical links (solid arrows) and nonhierarchical semantic links (broken arrows). An expanded frame is shown for Random Gentamicin Level, including its "Measures" slot with a link to Gentamicin. Note that this link was inherited from its parent term Serum Gentamicin Test.

MED) the slots are referred to as semantic relations, with the values corresponding to links in the MED's semantic network. Figure 1 shows some examples of terms, attributes and relations.

Our infobuttons are generally based on the use of *generic questions*.[9] In its simplest form, a generic question might be "Tell me about <term>" where <term> is filled in at run-time with a term corresponding to some datum the user sees on the screen. However, we can exploit the hierarchical and semantic links in the MED to identify additional terms which can be used in the generic questions. Furthermore, we can use information about the identified terms to allow the infobutton to select, from a larger set, those generic questions which are relevant to the selected terms. We have built several prototype infobuttons which use these techniques to link various parts of our clinical information system to on-line resources, including: patient admission/discharge information linked to Medline,[5] laboratory tests to Medline,[10] medication orders to the *Physicians Desk Reference* (PDR) (unpublished data), laboratory results to a clinical guideline,[11] laboratory results to DXplain,[12,13] and x-ray reports to a variety of resources on the World Wide Web.[14] In each case, we use information from the MED to support the infobutton.

TECHNIQUES

We employ three types of MED information for use in transferring patient data to on-line information resources: literal attributes, hierarchical links, and semantic links. We also make use of hierarchical links to assist with selection of appropriate generic questions. We describe each of these techniques below, with examples drawn from actual working infobuttons.

Literal Attributes

Conversion of clinical terms for use in information resources is not always as complicated as depicted in the Introduction above. There are many situations where simple translation is possible. Suppose, for example, that a user is reviewing laboratory results and encounters a Random Gentamicin Level for a patient. The laboratory results infobutton can proffer the question "What can cause a false-positive result of a Random Gentamicin Level?". If the user selects this question, the infobutton may decide to query Medline and will therefore attempt to translate the test term to MeSH. In Figure 1 we see that the Random Gentamicin

Level has no value for its "Main-MeSH" attribute, since there is no synonymous term in MeSH. However, the "Supplementary-MeSH" attribute has the value "gentamicin/bl" which is a combination of a MeSH Main Heading (the chemical Gentamicin) and a subheading ("Blood"). The infobutton can use this supplementary term in its Medline search strategy.

Now suppose that the user decides to check pharmacy orders on the patient and finds that the patient has an order for "Injectable Gentamicin". The pharmacy infobutton can proffer the question "What are the contraindications for Injectable Gentamicin?". If the user selects this question, the infobutton may decide to query the PDR. In Figure 1 we see that Injectable Gentamicin has the value "Garamycin" in its "Trade-Name" attribute. The application can then use this term as an entry point into our Web-based PDR.

Additional translation information is available from the UMLS, with its much richer synonymy and lexical matching capabilities. For example, a chest x-ray infobutton produces questions relevant to findings in chest x-ray reports and links to Web-based resources outside CPMC. We have identified the local terminologies of these resources and used lexical mapping routines (some from the NLM and some of our own) to create a translation table for converting our radiology findings into equivalent terms in these other resources.[15]

Semantic Links

Infobuttons can also make use of the semantic links in the MED to obtain additional terms of interest which, in turn, can evoke additional questions of interest. For example, the user reviewing the Random Gentamicin Level result will be offered questions such as "Tell me about Gentamicin " since, as can be seen in Figure 1, the MED includes the knowledge that this test measures the chemical Gentamicin. In this case, the question is about a chemical, not the test itself, and will return a different set of Medline citations.

Depending on the application, these semantic traversals can be quite complex. For example, many of the test terms are linked in the MED to terms which refer to abnormalities of the tests (not shown in Figure 1). The Serum Sodium Test, for example, is linked to the terms Hypernatremia and Hyponatremia through the semantic relation "Defines-Abnormal-Finding". If the user is reviewing a serum sodium test result which is below normal, the laboratory infobutton can use this information to generate the question "What is the differential diagnosis of Hyponatremia?". If the user selects this question, a query containing the term "Hyponatremia" can be sent to the DXplain server at the Massachusetts General Hospital and a list of diseases will be returned.[16]

The UMLS is also a source of semantic links which can be exploited by our infobuttons. For example, we have created an extract from the UMLS containing all instances of links between chemicals and diseases, including a link between the terms Sodium and Natriuresis. The laboratory infobutton uses this information to generate the question "What causes Natriuresis?" if the user indicates an interest in any tests which measure sodium. In this case, the system is making two semantic traversals - one from the test to the substance measured by the test (in the MED) and one from the measured substance to a disease term (in the UMLS).

Hierarchical Links

We are also experimenting with the hierarchical links in the MED to see where they may be relevant to user information needs. For example, when reviewing x-ray reports, users can choose a finding of interest, such as Pleural Effusion. The chest x-ray infobutton recognizes this as a child of the MED term Pleural Disease and offers to search for other x-ray reports for the same patient which show other findings (such as Pleural Thickening, Calcification of Pleura, or Pneumothorax) which are also children of Pleural Disease.

Selecting Appropriate Questions

Most of the questions generated by the various infobuttons are context-specific and invariant. For example, the questions generated about laboratory test terms will always include a question about the test itself as well as the substance measured by the test. However, depending on the classification of *specific* tests or substances, infobuttons may generate additional questions. For example, Figure 1 shows that the term Gentamicin is classified as both a Biologically Active Substance and as Pharmacologic Substance. Many of the substances linked to test terms are classified under Biologically Active Substance, but only some of these are classified under Pharmacologic Substance. This latter classification allows the laboratory infobutton to suggest the question "What is the dose for Gentamicin?" which would otherwise be generated only by the pharmacy infobutton when the user is reviewing the pharmacy orders for the patient.

In another example, the chest x-ray infobutton will recognize when a finding of interest (such as Cardiomegaly) is in the class Heart Disease. When this occurs, the infobutton adds the question: "Because the X-ray shows Cardiomegaly, would you like to see electrocardiograms from the same time period?" (If the user selects this question, the infobutton use the CIS itself as an information resource and query for electrocardiogram reports.)

DISCUSSION

The infobuttons described in this paper are more than simple links to relevant information resources. They carry out true integration by transferring patient information to the resource for use in automated information retrieval. The emergence of the Word Wide Web has removed many barriers to this task by providing a computing environment in which multiple, heterogeneous applications can communicate easily with each other. This capability does not, however, relieve us of the problems associated with choosing appropriate target applications and performing accurate translation of information from source to target.

Much work is still needed in determining how resources should be selected to answer questions and even what the questions should be. The work described in this paper does not purport to have solved that problem. We do believe, however, that as we learn more about the information needs of clinical system users, knowledge-based terminological information will be invaluable. For example, it seems reasonable to expect that the information needs arising from the review of an electrolyte test result will differ from those arising from review of a culture result. If an infobutton is to respond differently in these two situations, it will require precisely the kind of semantic information provided by the MED.

The translation task remains challenging. We have accomplished some of our translations through "hard-wiring" synonymy in the MED and by using a variety of lexical techniques. The UMLS has played a role in some cases but has proved inadequate in others. A number of researchers have experimented with knowledge-based translation techniques (a recent paper by Rocha and colleagues describes one such experiment and surveys several others[17]). However, these methods require a great deal more knowledge than is currently available in the MED or the UMLS.

Fortunately, the availability of formal terminological knowledge promises to increase. For example, LOINC, a terminology for naming laboratory observations, uses a knowledge-based design which is quite similar to the MED.[18] The maintainers of the more general-purpose terminology SNOMED are also exploring ways to include this kind of knowledge in their offering.[19] We believe that controlled medical terminologies will inevitably evolve in this direction and, as they do, the methods described in this paper will be applicable for use with a wide variety of information system integration efforts.

CONCLUSION

If we expect our systems to work together in intelligent ways, they will require more than simple information to be passed from one system to another; they will require knowledge about that information. Our infobuttons make use of rudimentary knowledge, including literal attributes, hierarchical links, and other semantic links to demonstrate the kind of sophisticated integration which is possible. The infobuttons can be found at: http://www.cpmc.columbia.edu/cisdemo/infobutton/

Acknowledgments

This work was supported in part by the National Library of Medicine through Unified Medical Language System, High Performance Computing and Communication, and Electronic Medical Record projects. Dr. Elhanan is supported in part by National Library of Medicine training grant LM07079. The authors thank Socrates Socratous for programming support.

References

1. Powsner SM, Riely CA, Barwick KW, Morrow JS, Miller PL. Automated bibliographic retrieval based on current topics in hepatology: Hepatopix. Computers and Biomedical Research. 1989;22:552-64.
2. National Library of Medicine, Library Operations. Medical Subject Headings. Bethesda, Maryland: National Library of Medicine, Updated annually.
3. McCarn DB. MEDLINE: an introduction to on-line searching. Journal of the American Society for Information Science. 1980; May:181-192.
4. Cimino JJ. Linking patient information systems to bibliographic resources. Methods of Information in Medicine; 1996;35(2):122-126.
5. Cimino JJ, Johnson SB, Aguirre A, Roderer N, Clayton PD. The Medline Button. In Frisse ME, ed. Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care; 1992 Nov 8-10, Baltimore (MD). New York: McGraw-Hill, 1992:81-85.
6. Commission on Professional and Hospital Activities. International Classification of Diseases, Ninth Revision, with Clinical Modifications (ICD-9-CM), Ann Arbor, 1978.
7. National Library of Medicine. UMLS Knowledge Sources - 3rd Experimental Edition. Bethesda (MD): The Library, 1992.

8. Cimino JJ, Clayton PD, Hripcsak G, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical terminology. Journal of the American Medical Informatics Association. 1994;1(1):35-50.
9. Cimino JJ, Aguirre A, Johnson SB, Peng P. Generic queries for meeting clinical information needs. Bulletin of the Medical Library Association. 1993;81(4):195-206.
10. Cimino JJ, Socratous SA, Clayton PD. Internet as clinical information system: Application development using the World Wide Web. JAMIA, 1995;2:273-84.
11. Cimino JJ, Socratous SA, Clayton PD. Automated Guidelines Implemented via the World Wide Web (Poster). In Gardner RM, ed. Proceedings of the Nineteenth Annual Symposium on Computer Applications in Medical Care; New Orleans, LA; October-November, Hanley & Belfus, Philadelphia, 1995:941.
12. Barnett GO, Cimino JJ, Hupp JA, Hoffer EP: DXplain. An evolving diagnostic decision-support system, Journal of the American Medical Association. July 3, 1987;258(1):67-74.
13. Elhanan G, Socratous SA, Cimino JJ. Integrating DXplain into a Clinical Information System using the World Wide Web. In Cimino JJ, ed. Proceedings of the American Medical Informatics Association Annual Fall Symposium (formerly SCAMC); Washington, DC; October, Hanley & Belfus, Philadelphia, 1996:348-352.
14. Zeng Q, Cimino JJ. Linking a clinical information system to a heterogeneous resource. Masys DR, ed. Proceedings of the American Medical Informatics Association Annual Fall Symposium, 1997 (in press).
15. Zeng Q, Cimino JJ. Mapping Medical Vocabularies to the Unified Medical Language System. In Cimino JJ, ed. Proceedings of the American Medical Informatics Association Annual Fall Symposium (formerly SCAMC); Washington, DC; October, Hanley & Belfus, Philadelphia, 1996:105-109.
16. Barnett GO, Hoffer E, Kim RJ, Famiglietti KT. DXplain on the World Wide Web. In Cimino JJ, ed. Proceedings of the American Medical Informatics Association Annual Fall Symposium (formerly SCAMC); Washington, DC; October, Hanley & Belfus, Philadelphia, 1996:988.
17. Rocha RA, Rocha BH, Huff SM. Automated translation between medical vocabularies using a frame-based interlingua. In Clayton PD, ed. Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care; Washington, D.C.; November, 1991:690-694.
18. Forrey AW, McDonald CJ, DeMoor G, Huff SM, Leaville D, Leland D, Fiers T, Charles L, Griffin B, Stalling F, Tullis A, Hutchins K, Baenziger J. 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. 1996:42:81-90.
19. Campbell KE, Musen MA. Representation of clinical data using SNOMED III and conceptual graphs. In Safran, C, ed. Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care; Washington, DC; November, McGraw-Hill, New York, 1993:354-8.