Ceusters W, Cimino J, Rector A. Medical Language and Terminologies. In: Sosa-Iudicissa M, Oliveri N, Gamboa CA, Roberts J (eds.) Internet, Telematics and Health, IOS Press, Amsterdam, 1997;:197-203.

Medical Language and Terminologies.

W. Ceusters (1), J. Cimino (2), A. Rector (3),

(1) Office Line Engineering NV, Belgium
(2) Columbia-Presbyterian Medical Center ,USA
(3) Medical Informatics Group, University of Manchester, UK

1. Introduction

Currently we are entering the third phase of introducing information technology to the health care delivery system. The first phase was mainly concerned with automating the administrative process. The second phase was about delivering medical applications. The third phase is about integrating a diversity of medical and administrative systems in one coherent compatible environment.

Now we see an emerging need for safe and sensible communication between these applications. The transmission and comparison of heterogeneous data is indeed recognised as a major issue in health care informatics. Only a multi-purpose representation of medical knowledge and information, based on explicitly stated principles within a uniform framework, will allow smooth integration of health care information systems. Due to the diversity of languages and terminologies used within Europe, CEN\TC251 has recognised this need, and has outlined the approach to be taken [1]. Similar developments are conducted on the American continent. Central in this approach is the dream to develop concept systems that are language- and purpose independent, highly expressive, fully formalised, and when implemented on a machine, computationally tractable. Whether or not this is really feasible, is a question on its own [2].

Medical nomenclatures, vocabularies and coding and classification systems are particular (though “simple”) examples of such concept systems. Nevertheless, development and maintenance of them, is problematic. Currently the development/revision cycle for major medical nomenclatures and classification systems is measured in decades. This is totally inadequate to the needs of current medical practice. The need for rapid deployment of specialised nomenclatures for new problems, international trials, international work on improvement of quality, etc. leads to fragmentation and proliferation of special purpose extensions, controlled vocabulary, data sets, definitions, etc. National centres for coding and classification and national centres for translation of standard nomenclatures work in isolation, interacting only at the level of entire sections or chapters, often only at the level of entire systems comprising tens of thousands of items.

There is an urgent need for means to support continuous, fine-grained cooperation and extension of coding and classification systems and nomenclatures. Telematic support through tools for distributed cooperative development of common concept models, terminologies, and linguistic resources offer the best hope for achieving such cooperation. Close cooperation supported by distributed information systems represents a major organisational as well as technological challenge. The non-technological factors and pitfalls will require as careful an analysis as the technical challenges.

In this paper the possibilities of the Internet related to the issues described above, are explored. Development and exploitation go hand in hand which is pointed out by some real life examples.

2. The Need for Internet-based Vocabulary Services

The rapid acceptance of communication standards under the Internet Protocol (such as FTP, Gopher, and HTTP) has given new meaning to the term "distributed application". Now, the components of a complex application may not only reside on different machines but at different institutions in different countries. Communication protocols, however, solve only one of the tasks involved in information exchange. Message format standards are important for structuring the information to be exchanged. When the messages contain coded data, additional standards are desirable for controlled vocabularies used in coding. When a single standard is not readily available (as in medicine), vocabulary translation is required. In a distributed architecture, different remote components typically use their own local terminologies.

Trying to maintain copies of each vocabulary at all sites can be problematic; Internet-based vocabulary services are an alternative means to support communication among components. To illustrate such services, we will consider a prototypical controlled vocabulary. The vocabulary is made up of many terms, each of which is identified by a unique code. Each term is associated with a set of attribute-value pairs. All terms have the attribute "name" which has a unique, literal value. Other literal-valued attributes are "synonym" and "abbreviation". If the term was derived from some coding system, it might have a literal attribute to hold its code from that system. Attributes can also include pointers to other terms in the vocabulary. For example, if the vocabulary includes a hierarchical arrangement, attributes like "parent" and "child" would be associated with values which are term codes. Similarly, non hierarchical attributes can have relationships between terms (such as "opposite" or "has part").

The simplest query a server could support would be "give me the name for code X". Only slightly more complicated would be "give me all the synonyms of code X". Term relationships can be retrieved through a similar mechanism, such as "give me all the children of code X" or "give me all the parts of X". Servers must also be able to perform look-up functions, such as "find me all the terms which have 'yyy' as a synonym" or "find me all the terms which have Y as a part". Some form of lexical look-up should also be included; for example, "find me all the terms which have 'yyy' in their name".

Depending on the purpose of the vocabulary and the applications which will typically require vocabulary services, the server should also provide some well-chosen complex query facilities. For example, if we wish to know all the classes of some term X, we might ask "give me all the parents of X". The result might be the terms Y and Z. Now we must ask "give me all the parents of Y" and "give me all the parents of Z". we must then take the results of these two queries and formulate new queries. In a distributed, net-based environment, a recursive process such as this is inefficient. If this type of query is common, the server can be set up to respond to a more sophisticated query, such as "get me all the ancestors of X". Now only one query and response needs to be passed over the network. More complex queries for some servers might include "expand the meaning of X", e.g. "expand the meaning of cystitis" expecting an answer something like "an inflammation of the urinary bladder". In other situations, key questions might be to translate a term from one language to another or to find its nearest match, if any, according to a given criterion in a particular coding scheme such as ICD-10 or the SNOMED Disease chapter.

A an architecture for providing vocabulary services, such as those described above, has been described [3]. In the next few sections, we consider several examples of current research to extend the vocabulary server concept to the Internet.

3. Galen

The GALEN and its successor GALEN-IN-USE projects are seeking to produce a compositional 'common reference (CORE) model for medical terminology to be delivered via a 'terminology server' [4]. Currently GALEN makes use of the Internet primarily as a means of developing the model cooperatively and communicating amongst different sites in a distributed development project. In the future it expects to make many aspects of the service available via the Internet as part of a broader programme to encourage cooperative development of information and decision support systems.

The long term goal of GALEN is to allow workers to develop medical records and decision support systems independently and integrate them smoothly. The heart of this effort is to develop a large model cooperatively in a multilingual and multicultural environment. Currently work is focusing on development a model for surgical procedures.

In order to understand GALEN's use of the Internet, it is necessary to describe the GALEN approach in more detail. GALEN's terminology server encapsulates a number of distinct functions in a single applications programming interface:

- The concept module or 'ontology server' which provides a formal description of medical concepts. The concept model holds the Common Reference Model of clinical concepts. The concepts and relations in the model are the building blocks for more complex concepts found in standard terminologies - expressions such as:

  "Fracture due to osteoporosis" =

  Fracture <u>which</u>

*hasLocation AnatomicalNeck *which* isDivisionOf Femur*
hasContributingCause Osteoporosis.

- A Code Conversion module which maps existing coding schemes to and from expressions made up of concepts in the Common Reference Model. The Code conversion module contains both strategies for mapping and mechanisms for dealing with specific details such as the dagger-asterisk notation in ICD.
- A multilingual language module which contains the lexicons and grammars needed to express concepts in the CORE model in various natural languages - currently English and French, soon in Dutch and German, and later in other European languages.

Once developed, the GALEN Terminology Server will most commonly be used inside other tools invisibly to users. For example, in the PRESTIGE project it is being used as part of the authoring tools for a series of protocol management systems, and in the PEN&PAD GP system, the model is being used as part of a data entry system for a commercial system for general practitioners. For these tasks, the model is typically delivered intact to a local server by ftp and queried locally, since at the present time the number of requests required to use the model routinely would be unacceptably cumbersome over via the usual Internet HTML protocols.

However, the primary task in realising the model is to achieve consensus amongst a wide variety of disparate users. For this the Internet is an key medium. The advent of the world-wide-web and HTML browsers makes it possible to produce a much higher level of cooperation than previously.

However, simple email and list-server discussions remain the backbone of a process which requires successive refinement, discussion and compromise. However, the use of HTML mediated services is increasing rapidly and expected to become the backbone of the project over the next few years.

All those collaborating with the project are routinely given access to a private Web page which includes the preliminary and draft information on the work of all groups as well as a running question and answer and documentation service. Separate private and public sections are essential for giving participants the confidence to post preliminary and draft documents and to make a maximum of material available. The plague of collaborative projects is reluctance to share early results.

The Web pages also contain summaries, explanations and documentation of the material in the server itself as well as files in readily accessible formats such as Microsoft's Rich Text Format and Word outliner which make it easy for newcomers to gain a quick overview of the project.

Whether the use of the Web within GALEN will grow to a primary means of distribution remains to be seen. For the time being it is being use where its strengths are critical - communication, ease of updating, persistence, and wide accessibility -- in order to achieve consensus. Where its weaknesses -- poor reliability, slow interaction, limitations in the interface -- are serious liabilities, bulk transfer via ftp is used. Where immediacy is critical, email and list-servers remain the main-stay of the project, though archives may appear on a web page. All are, of course, manifestations of the Internet, without which a multi-national cooperative project of this nature would be, literally, impossible.

4. Columbia University's Medical Entities Dictionary

The clinical information system at Columbia-Presbyterian Medical Center (CPMC), relies on a Medical Entities Dictionary (MED) for coding purposes [6]. The vocabulary is used by a variety of distributed systems, including ancillary data-upload systems, data display programs, and a clinical event monitor. Vocabulary services for these applications are provided through remote procedure calls (RPC's) to a vocabulary server running on a Unix platform. Besides local CPMC applications, these RPC's can be called by World Wide Web applications through Common Gateway Interface (CGI) programs. The reader can examine how such services are used by exploring some of CPMC's Web-based applications.

An interactive browser is available at http://www.cpmc.columbia.edu/intermed/MCST.html. The interface provides the user with a focused view of one term in the MED, with a screen consisting of three sections. The top section shows the current position in the MED hierarchy - either by displaying a graphic image of part of the MED tree, or with lists of the parents and children of the current term. Users may traverse the hierarchy by clicking on a term in the tree display or by selecting one of the terms in the parent/child lists. The middle section of the screen shows a frame representation of the term of the current term. Each slot in the frame corresponds to an attribute of the term. If the attribute has a value, the value is displayed. When a value is a pointer to another term the name of the term is displayed; users can traverse the MED's semantic network by clicking on terms which appear as attribute values. All of the information displayed in these two sections of the display are obtained from the vocabulary server via RPC's; queries include "get all parents", "get all children" and "get all attributes", as well as (for each attribute) "get attribute name" and "get values for attribute in term". For parent terms, child terms and term-pointer attribute values, the browser application must also query the server for the name of each term.

The bottom section of the display provides a term look-up facility. Terms can be found by term code or string value. String values are matched by as word matches, partial word matches or exact matches, depending on the option selected by the user. For example, if a user requests a partial word match for "electro", the system will find all terms with "electrocardiography" and "electroencephalography" in their names or as any slot value (for example, in the "synonym" slot of "12 Lead EKG").

An example of an clinical application which uses vocabulary services can be found at http://www.cpmc.columbia.edu/cisdemo which has been described in a recent publication [7]. This application queries an Internet-based clinical information server for patient data. The data are provided in an HL7 (Health Level 7) format which provides coded data. The application must then query the MED server with queries such as "what is the test name for term X" and "what are the 'units' for term X".

The application can also use the vocabulary server to help with additional clinical queries. For example, when the user clicks on a "summary" button, the application takes the MED code for the selected test (e.g., "Chem-7 Glucose Test") and queries the vocabulary server for the test's parent class (e.g., "Intravascular Glucose Test"). It can then query the clinical data server for all test results in that class, allowing the application to produce a graph of all similar test results (e.g., all serum, plasma and whole blood glucose tests).

The same clinical application can use the vocabulary server for providing access to other Internet-based information resources. Starting with the MED code for a laboratory test (being displayed to the user), a function queries the vocabulary server to "get the MED code of the substance measured by the test" and then uses the results to query "get The MeSH name for the substance". With this result, the system can provide a "Medline Button" [8] by generating queries to the medical literature about substances measured by tests. In similar approach, using medication data from the clinical information system, we query the Vocabulary server to "get the trade name for the drug"; the trade name, in turn, is used to provide entry in to CPMC's on-line Physician's Desk Reference. In a more complex approach, the vocabulary server is used to convert laboratory test values into clinical findings (for example, a serum sodium of 120 mEq/l is converted to "hyponatremia"). Clinical findings derived in this way are then used as input to the Massachusetts General Hospital' Internet-accessible DXplain diagnostic system.

5. InterMed

The researchers at Columbia are working with other groups (at Stanford University, Harvard University and the University of Utah) on an Internet-based collaboratory project called InterMed [9]. Part of this project is an effort to develop shared model of vocabulary and vocabulary services such that multiple sites can contribute vocabulary content to a single central server, to their mutual benefit.

They are now experimenting with a platform at Stanford which provides an Web browser to vocabulary knowledge bases in the Ontolingua environment (a language developed for representing sharable ontologies, using the Knowledge Interchange Format [10]). The desired capabilities for this

platform include "tools for building and maintaining the vocabulary, tools that allow users to browser and search through the vocabulary, and a protocol for direct programmatic interface into the vocabulary for use by medical applications" [11]. Thus far, several ontologies have been constructed, including a shared version of the CPMC MED, allowing a direct comparison of browsers and servers at the Columbia and Stanford platforms. These ontologies can be accessed via http://www-ksl-svc.stanford.edu:5915/.

One application of an Internet-based vocabulary server can be found in the clinical information demonstration at Columbia (http://www.cpmc.columbia.edu/cisdemo). In this application, a "guideline" application is called (available in the demo whenever the user views a cholesterol test result; also available at http://www.cpmc.columbia.edu/chguide.html) which accepts patient information and returns a recommendation for patient management. In order for the guideline to recognize the patient data, it must be converted to a controlled form. The vocabulary server provides the necessary functions for this task. The guideline application can be called by any Web browser and can therefore be integrated into clinical applications at other sites. The vocabulary server provides the necessary link for converting local patient data into a form recognizable by the application.[12]

6. UMLS

The US National Library of Medicine is developing vocabulary services for its Unified Medical Language System (UMLS) [13]. The UMLS brings terms from multiple controlled vocabularies together into a single set of knowledge sources which provide information about the names, codes, meanings and interrelationships of terms from these vocabularies. The server is being established initially to support users who wish to query the UMLS to find if the terms they are using in their own systems can be found in the UMLS, thus providing a way to link to information sources which are coded in UMLS source vocabularies. The server is accessible via the World Wide Web at http://wwwkss.nlm.nih.gov. The interaction may be in either a batch mode or through an interactive interface.

7. References

1. CEN/TC251. Directory of the European Standardisation Requirements for Health Care Informatics and Programme for the Development of Standards.
2. Rocha RA, Huff SM, Haug PJ, Warner HR. Designing a Controlled Medical Vocabulary Server: The VOSER Project. Computers in Biomedical Research,1994;27(6):472-507.
3. Ceusters W, Deville G, Buekens Ph. *The Chimera of Purpose- and Language Independent Concept Systems in Health Care*. In Barahona P, Veloso M, Bryant J (eds.) Proceedings of the XIIth International Congress of EFMI, 208-212, 1994.
4. Nowlan WA, Rector AL, Rush TW, Solomon WD. From terminology to terminology services. In Ozbolt JG, ed. Proceedings of the Eighteenth Annual Symposium on Computer Applications in Medical Care; Washington, DC; November, Hanley & Belfus, Philadelphia, 1994:150-154.
5. Rector AL, Solmon WD Nowlan WA and Rush TW A terminology server for medical language and medical information systems. Methods of Information in Medicine 34:147-157
6. 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.
7. Cimino JJ, Socratous SA, Clayton PD. Internet as clinical information system: application development using the World Wide Web. Journal of the American Medical Informatics Association; 1995: 1995;2(5):273-284.
8. 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; Baltimore, MD; November, McGraw- Hill, New York, 1992: 81-85.
9. Oliver DE, Barnes MR, Barnett GO, Chueh HC, Cimino JJ, Clayton PD, Detmer WM, Gennari JH, Greenes RA, Huff SM, Musen MA, Pattison-Gordon E, Shortliffe EH, Socratous SA, Tu SW. InterMed: an Internet-based medical collaboratory. In, Gardner RM, ed. Proceedings of the Nineteenth Annual Symposium on Computer Applications in Medical Care; New Orleans, LA; October, Hanley & Belfus, Philadelphia, 1995:1023.
10. Genesereth MR, Fikes RE. Knowledge Interchange Format, Version 3.0 Reference Manual. Computer Science Department Technical Report Logic-92-1, Stanford University, CA.
11. Gennari JH, Oliver DE, Pratt W, Rice J, Musen MA. A Web-based architecture for a medical vocabulary server. In, Gardner RM, ed. Proceedings of the Nineteenth Annual Symposium on Computer Applications in Medical Care; New Orleans, LA; October, Hanley & Belfus, Philadelphia, 1995:275-279.

12. Cimino JJ, Socratous SA, Clayton PD. Automated guidelines implemented via the World Wide Web. In, Gardner RM, ed. Proceedings of the Nineteenth Annual Symposium on Computer Applications in Medical Care; New Orleans, LA; October, Hanley & Belfus, Philadelphia, 1995:941.

13. Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. Methods of Information in Medicine. 1993; 32(4):281-291.