MEDINFO 2004
M. Fieschi et al. (Eds)
Amsterdam: IOS Press

Using a Terminology Server and Consumer Search Phrases to Help Patients Find Physicians with Particular Expertise

Curtis L. Cole[a], Andrew S. Kanter[b], Michael Cummens[b], Sean Vostinar[a], Frank Naeymi-Rad[b]

[a]*Weill Medical College of Cornell University, New York, New York, USA*
[b]*Intelligent Medical Objects, Inc., Chicago, Illinois, USA*

Abstract

Objectives: *To design and implement a real world application using a terminology server to assist patients and physicians who use common language search terms to find specialist physicians with a particular clinical expertise.* ***Method:*** *Terminology servers have been developed to help users encoding of information using complicated structured vocabulary during data entry tasks, such as recording clinical information. We describe a methodology using Personal Health Terminology™ and a SNOMED® CT-based hierarchical concept server.* ***Results:*** *Construction of a pilot mediated-search engine to assist users who use vernacular speech in querying data which is more technical than vernacular.* ***Conclusion:*** *This approach, which combines theoretical and practical requirements, provides a useful example of concept-based searching for physician referrals.*

Keywords:

Information Storage and Retrieval; Patient Education; Vocabulary, Controlled; Specialties, Medical; Systematized Nomenclature of Medicine

Introduction

Computers require a standard, normalized nomenclature and classification system if they are to manipulate concepts meaningfully. The problem of medical terminology and medical classification long predates computers and medical informatics. A wide variety of clinical, administrative, regulatory, and technical demands have lead to greater emphasis on the creation of ever more specific and systematic terminologies. With the advent of new technology informaticians have had to face the fact that, as Christopher G. Chute modestly points out, "health concept description is actually quite hard." [1]

Comprehensive clinical classifications are now so sophisticated that only a computer can practically represent and manage their complexity. Chute et al [2] argue that an ideal terminology must have "consistency of view", "non-redundancy", and should be "free of context." Encoding and decoding to and from vernacular to the structured terminology is equally necessary if the terminology is to be useful.

The "terminology problem" is therefore three problems: 1) the creation of a standard terminology, 2) the development of techniques for encoding or indexing information as it is stored using the terminology, and 3) translating queries so that real world questions can be answered using the structured data. Most effort to date has focused on the first two problems. Terminology servers have been shown to be relatively reliable methods for converting variable "local" parlance into large-scale standardized terminologies. [3]

In theory, querying a standardized database should be easy. Once the data is codified, it should be relatively straightforward to query because the coding system is knowable. However, in the application of this theory an obvious problem arises; while the coding system is knowable, it typically is not known. Terminology servers leave clinicians free to speak and write as they normally would while querying clinical information using standardized terminologies [4][5].

With a concept-based terminology, computers can access and manipulate data in ways equal to or often far superior to human proxies. For example, the power and utility of conceptual based terminology mapping to web searching was well demonstrated by Elkin et al. [6] who showed how medical content could be more efficiently searched and served using an automated conceptual mapping tool. Kanter et al. [7] demonstrated how clinical concepts could automate searching of electronic content at the point of care.

Now consider the real world problem of a patient searching for a physician to treat "lower back pain". S/he asks a computer to deliver a consistent, non-redundant, answer without entering the context. The pre-technical solution was for the patient to see a primary care physician. This "gate keeper" had context and the implicit permission to use subjectivity to resolve the inherent ambiguities in the problem. In the contemporary world of managed care, physician panel restrictions, direct to consumer advertising, and a general increase in consumer driven health care the need for new methods of matching providers with patients has expanded. Patients and referring providers increasingly look to the Internet to find medical expertise.

However, most sites and search engines lack the lexical sophistication to match patient requests with physicians' areas of interest unless there is a direct match. For example, unless the search is conceptual, a patient searching for an AIDS expert may not find a physician who only lists an interest in HIV. From May to July 2003 we conducted a survey of 93 metropolitan New York City medical center web sites and found that not one had a conceptual based search tool.

Similarly, patients searching for specific medical information will only find data that is linked by whatever hierarchy or conceptual model the specific engine utilizes. General-purpose search engines have accounted for this in a variety of ways such as link history and keywords [8]. While these strategies work in part when faced with the whole scope of the Internet, they are not specific enough to search an individual medical site with either the necessary breadth or precision demanded of a medical context. It also may fail pedagogically to lead patients where we want them to go. For example, a search engine that uses linkage statistics like Google® is easily diverted by non-medical homonyms. What is needed is a medical vernacular to technical terminology server with applicability to query generation rather than data storage.

Leroy and Chen [9] have approached this problem by providing lay users with a "medical concept mapper." This helps the users learn the correct terms to search against. Solbring et al. [10] have explicitly tested the use of synonyms and synonym clusters as extensions to reference terminologies as a technique for optimizing query recalls despite the use of vernacular terms. We attempt to build on this theoretical foundation in a real world application. A few key differences in our approach are worth noting. First, unlike Leroy and Chen who are facilitating more complex searches than we support, we attempt to shield the user from the terminology as much as possible. Second, we are explicitly building a real world application so certain theoretical purities had to yield to pragmatic design requirements. We present this work as applied informatics rather than as theoretical work. As such, we will focus as much on practical considerations as theoretical concepts. This paper describes our attempts to build and pilot a terminology server-driven search system for a medical web site.

Methods

Weill Cornell has used its public web site as a real world laboratory to develop and test a concept based search engine. The Physician Organization Profile System (POPS) is a web-based application that allows practice administrators to develop and maintain physician profiles for display on the www.CornellPhysicians.com web site.

The driving force behind this project is that consumers searching Internet sites expect to select from possible hits based upon criteria they provide; not based upon limited criteria presented to them. Note, however, that this is not intended to help users search the entire web but simply one medical site. Broader medical searches of the web require a very different approach such as the HelpfulMed project, described by Chen et al. [11] or Health Search described by Kanter et al [12].

Our goal for the pilot was two-fold:

1. To develop a terminology server-driven search engine that allows users, namely patients and referring providers, to find a physician based upon free-text entry, which is parsed and matched to our physician profiles.
2. To develop an administrative interface for maintaining the underlying lexicon, term associations, and display order/precedence logic.

The innovations required to accomplish these goals were two-fold. First, it required mapping of a vernacular medical lexicon with a database of physician expertise and medical content that uses technical concepts. This allows physicians and consumers to each use their own vocabulary and still find one another on the web. Second, due to practical considerations the physician database needed to be as self-describing and self-maintaining as possible. This pragmatic requirement ensures that the site will remain accurate and up to date with minimal overhead costs. This is essential if the tool is to be affordable and applicable to a wide variety of medical sites and institutions.

Definitions of Areas of Interest

Patients seek physicians with expertise in specific topics and may differentiate the level of expertise. For a rare disease or complex condition they may ask for the world's expert or seek participation in a clinical trial. For a more common disorder, they may look for a local expert with strong clinical experience without regard to his/her academic standing. We define all these levels of specialization as "Areas of Interest"; the symptoms (e.g. heartburn), syndromes and diseases (e.g. diabetes), procedures (e.g. endoscopy), and expert services (e.g. child custody) that a provider is known for professionally. We further classify these areas on interest into three overlapping categories:

Personal Sub-Specialization - This may be relatively obscure and highly specific. But it is also what is most likely to differentiate a specific physician from their peers in the community. Examples: Dr. Y. is a general surgeon who is world famous for burns. Dr. D. spends much of his time reading echocardiograms but he is best known for his expertise in Marfan's syndrome. Sources of this information include the physicians' published research, their staff, and the physicians themselves.

Core Specialization - This is the "bread and butter" expertise of any given physician. In the example above, Dr. D.'s expertise in echocardiography would fall here. Operationally, these data can be approximated by looking at billing data though confirmation is needed by the physicians and their staff.

Basic Competency - This is the general domain of the physician's specialty and sub-specialty. Any internist can treat diabetes. Any general surgeon can perform an appendectomy. For the purpose of the search tool, these data can be derived from a standard ontology or medical knowledgebase.

This classification was necessary due to the lack of context inherent in real world queries. The search engine could prioritize these levels and returns results accordingly. Without this classification a conceptual based search could still work but some other arrangement would still be required to determine the display order of results. This particular system has the advantage of being relatively easy to implement due to the use of easily acquired data (such as billing records).

Use Case

The primary use case is that a consumer searches the web to find a physician with specific expertise. The search returns the links to the web page of a faculty member with a special interest in the topic requested.

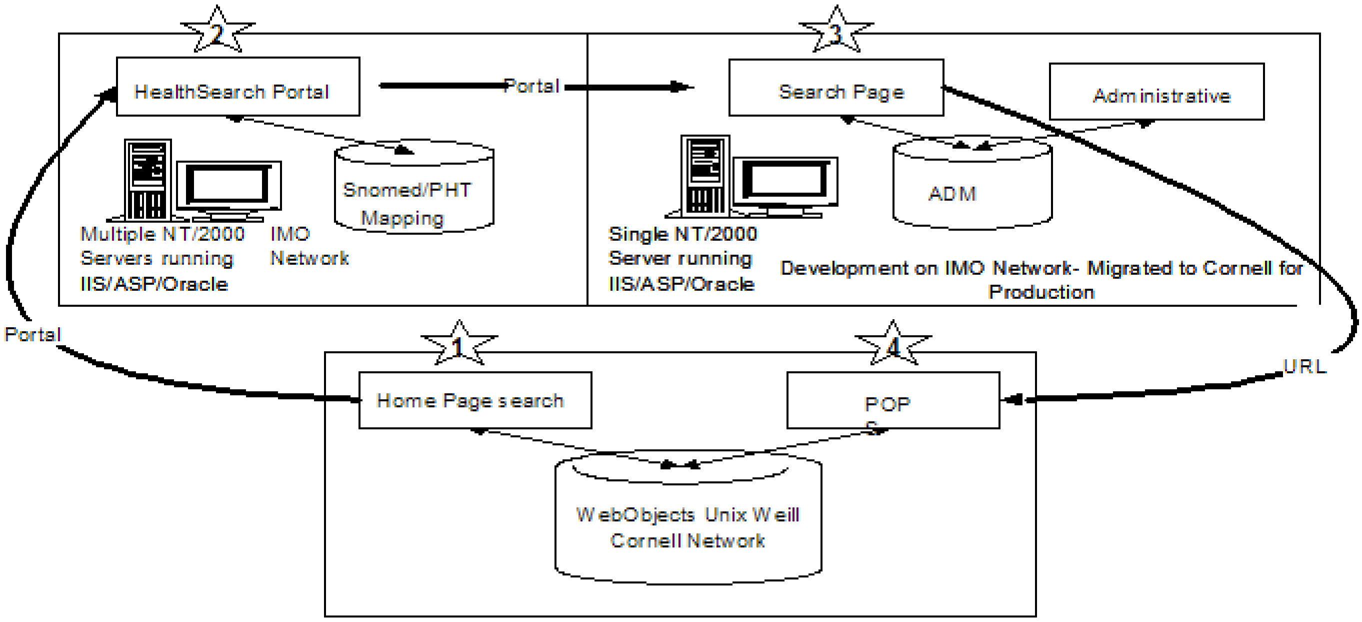

Figure 1 - Schematic of Vernacular Concept-based Search

There are five main tasks in accomplishing this:

1. Parsing user input to determine the area of interest,
2. Lexical expansion of area of interest,
3. Concept expansion of area of interest,
4. Linking of expanded search term to Cornell faculty member's area of interest,
5. Refining the resulting search based on user input.

Example

A patient is moving to the area and has a relatively common genetic blood disease, hemochromatosis. He knows that he will need a specialist to treat his condition so he visits our web site and enters "hemochrometosis" into our search. The search engine should first recognize that the patient mistakenly misspelled the condition.

In this example we have no specialists who have listed hemochromatosis as their sub-specialty. However, via ontological association, the search engine identifies hemochromatosis as a genetic or metabolic disease of the liver. A specialist who deals with diseases such as Wilson's Disease or hemochromatosis might have flagged metabolic disease of the liver as one of his/her core specialization. The search engine would then identify this specialist and provide a link to that personalized web page.

Design Requirements

Chute et al.[4] have defined the desiderata of a clinical terminology server, as distinct from a terminology development server. He focuses on the necessary and desired functions that facilitate term entry by clinical users. These "data entry" requirements include word normalization and spelling correction as well as more complex lexical functions such as term completion and conceptual decomposition. The use case of operational clinicians entering data is similar to, but also somewhat the converse of, a patient (or physician) asking a question of that data. We divide these "desiderata" into two classes; User Input Parsing and Concept Mapping.

User Input Parsing

Free-text entry - The user should be allowed to enter free-text search criteria. Users may be required to select from a list of other criteria where we can assume that s/he has "domain expertise" (such as insurance plan).

Misspellings - The most common problems among health-related web sites are misspellings; the search engine should correct or provide suggestions for common misspellings.

Conceptual relationships - The search engine should match the entered criteria with conceptually similar terms. For example, a search for HIV should additionally return physicians with AIDS-related expertise.

Logical Operators - The search engine should recognize and accept English Boolean operators such as "and" and "or." For refinement of the search "not" is desirable.

Word Normalization - The search engine should account for lexical variants, dual terms, and other foreseeable term variations. Examples include "MRI" vs. "M.R.I."; "Dentistry, General" vs. "General Dentistry"; "Birth/Pregnancy" vs. "Birth and Pregnancy".

Conceptual Mapping

Lexical Expansion - Iterative lexical expansion of expanded concepts is necessary to enhance mapping to standard terms.

Conceptual Expansion - There are several possibilities for the concept expansion. SNOMED® CT is becoming a standard for reference terminology and was used in our pilot.

BREAST CANCER (Diagnosis)
May also be referred to as Breast Neoplasms.
Tumors or cancer of the breast.

The following physicians have specific expertise in breast cancer:

Department	Name	Specialties	Top Expertise
Medicine	Vahdat, Linda	Medical Oncology	Breast Cancer, Oncology, Hematology
Medicine	Chuang, Ellen	Medical Oncology	Breast Cancer, Hematology, Oncology
Surgery	Swistel, Alexander	Surgery	Breast Cancer, Skin-sparing Mastectomy
Surgery	Pressman, Peter I.	Surgery	Breast Surgery, Breast Cancer
Radiology	Vallejo, Alvaro	Radiology	Gynecologic Tumors, Breast Cancer
…			
Psychiatry	Abrams, Robert C.	Psychiatry	Geriatric Psychiatry
…			

.. results

The following physicians are qualified in the treatment of breast cancer:

Department	Name	Specialties	Top Expertise
Medicine	Bergsagel, Peter	Medical Oncology	None available
…			
Medicine	Adelman, Robert A.	Geriatric Medicine	Geriatric Medicine, Internal Medicine
…			

.. results

Figure 2 - POPS Provider Search Results

For example the area of interest "Diagnostic Radiology" could be lexically expanded to "Diagnostic Imaging" which would subsume concepts for MRI, CT, X-ray etc.

Computer science and medical lexicons were used to respond to these problems. IMO Personal Health Terminology™ (PHT) is a medical lexicon containing both provider and patient descriptions for medical concepts. IMO PHT is mapped to several standardized vocabularies including ICD-9, UMLS, MeSH and SNOMED® CT. IMO Health Search is a web portal which does parsing, stemming and phonetic search string matching [7]. In this project, Health Search takes the query string and maps it to PHT which then provides links to SNOMED® concepts.

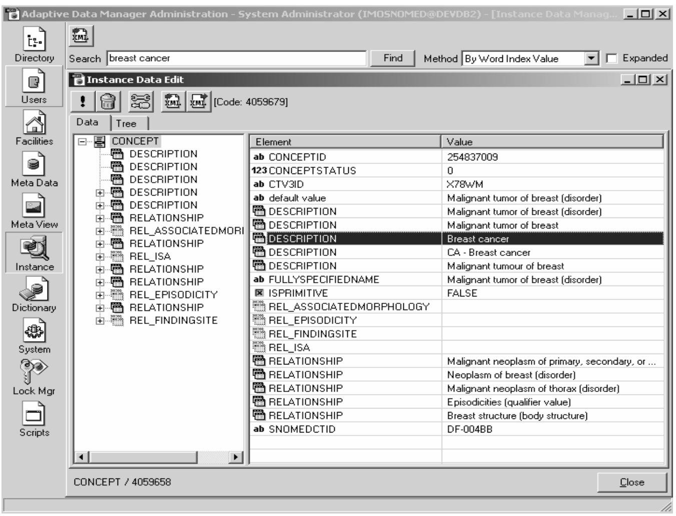

Figure 3: IMO ADM showing SNOMED Relationships

For the Weill Cornell physicians, a similar mapping was used to assign SNOMED® concepts to their areas of interest. An existing flat, non-heirarchical, non-conceptual database of POPS "areas of interest" was manually mapped to SNOMED®. An interface for physicians to select their areas of interest was developed to allow physicians to identify their expertise prospectively using these areas of interest. They could also manually assign SNOMED® concepts if the mapping table was incomplete.

Results

The pilot application meets the requirements outlined above by integrating IMO Health Search, Personal Health Terminology, a SNOMED-based vocabulary server, and the POPS database. It works in the following manner (shown in Figure 1):

- A user enters a search term (1) into a search box or advanced search page on the Cornell/Weill website.
- The search term is sent to the IMO Health Search Portal (2), where the search is performed against the IMO Personal Health Terminology (PHT).
- If the search matches a PHT term, its SNOMED® codes (acquired through a PHT-SNOMED map) are sent via a portal string to an ADM database search page (3).
- If the search term is not found in the IMO PHT vocabulary, the entire search string is passed to the ADM database search engine where a Boolean search will be performed. This search goes against IMO's enhanced SNOMED® CT word index, which includes all synonym and lexical variants for each SNOMED® CT concept.
- The database maintains the link between POPS physicians and their SNOMED-coded areas of interest.
- The ADM terminology server will recursively query the SNOMED® concept hierarchy/graph until it finds physicians with appropriate specified areas of interest.
- Search results (Physician name, specialty, unique identifier) are displayed on the ADM Search page. The results are ordered based on what specialization classification the match came from with personal sub-specialization higher than basic competency. The results will contain links to existing POPS physician pages hosted on the Cornell Network (4) (See Figure 2).

Discussion

SNOMED® CT "provides a "common language that enables a consistent way of capturing, sharing and aggregating health data across specialties and sites of care." The SNOMED® CT data structure is highly granular and contains 344,000 unique medical concepts. The content also includes 913,000 descriptions and synonyms.

All of the concepts are arranged in a hierarchical structure through a table of IS-A relationships [13,14]. Concepts may have and often do have multiple IS-A relationships. For example 'breast neoplasm is a disease of breast" and "breast neoplasm is a neoplastic disease." Navigating the hierarchical tree dynamically can pose problems for developers or users wishing to query a SNOMED-encoded database.

IMO® Adaptive Data Manager™ was developed to support the storage and retrieval of complex relationships within an RDBMS. The model creates a conceptual layer which allows users to model directed graph relationships while maintaining basic hierarchical relationships understood and required in any medical data application. The directed graph allows for constructing complex interrelationships among database elements while maintaining source data in their original and intended form. Using ADM as a vocabulary server allows for the management of terminologies from multiple sources.

ADM allows for navigation "up" and "down" the tree. Examples of hierarchical query methodology using SNOMED have been published previously [15]. In this project, each area of physician interest is linked to one or more SNOMED® CT concept(s). Taking advantage of a metadata architecture running on Oracle®, the entire SNOMED® CT structure has been represented within the database as a navigable graph. (See Figure 3) If a user queries for a concept which has no link to an appropriate physician, a recursive database query navigates the SNOMED "tree" to find the closest ancestor SNOMED® CT concept which has a linked physician. In the case of SNOMED® CT concepts with multiple parents, each parent's ancestor is evaluated for linked physicians and all concepts with an appropriate physician are returned.

This pilot project has demonstrated how a concept-based terminology server combined with a standardized vocabulary with consumer terms can help patients find physicians with particular clinical expertise. This appears to be the first published use of the SNOMED® CT vocabulary as an interactive terminology server for the purpose of concept-based searching. Further elaboration of the administration methodology and customizations to the search engine is expected.

References

[1] Chute CG. Clinical Classification and Terminology: Some History and Current Observations. *J Am Med Inform Assoc*. 2000 May; 7 (3): 298–303

[2] Chute CG, Cohn SP, Campbell JR, and for the ANSI Healthcare Informatics Standards Board Vocabulary Working Group and the Computer-based Patient Records Institute Working Group on Codes and Structures. A Framework for Comprehensive Health Terminology Systems in the United States: Development Guidelines, Criteria for Selection, and Public Policy Implications. *J Am Med Inform Assoc*. 1998 November; 5 (6): 503–510.

[3] McDonald FS, Chute CG, Ogren PV, Wahner-Roedler D, and Elkin PL. A Large-Scale Evaluation of Terminology Integration Characteristics. *Proc AMIA Symp1999*: 864-7.

[4] CG Chute , PL Elkin, DD Sherertz, and MS Tuttle. Desiderata for a Clinical Terminology Server. *Proc AMIA Symp 1999*: 42-6.

[5] Burgun A, Patrick D, Bodenreider O, Botti G, Delamarre D, Poulinquen B, Oberlin P, Leveque JM, Lukacs B, Kohler F, Fieschi M, LeBeux P. A Web Terminology Server Using UMLS for the Description of Medical Procedures. *J Am Med Inform Assoc.* 1997;4:356–363.

[6] Elkin PL, Ruggieri A, Bergstrom l, and Bauer BA. A randomized controlled trial of concept based indexing of Web page content. *Proc AMIA Symp 2000*: 220-4.

[7] Kanter A, Naeymi-Rad F, Buchan IE. Right information, right patient, right time: intelligent content searching supporting point-of-care applications. *Proc AMIA Symp 2000*; 403-7

[8] Brin S and Page L. The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer Science Department, Stanford University, Stanford, CA, USA http://www7.scu.edu.au/programme/fullpapers/1921/com1921.htm accessed August 22, 2003

[9] Leroy G and Chen H. Meeting Medical Terminology Needs – The Ontology-Enhanced Medical Concept Mapper. *IEEE Transaction on Information Technology in Biomedicine*, Vol 5(4), Dec 2001.

[10]Solbrig HR, Elkin PL, Ogren PV, and Chute CG. A Formal Approach to Integrating Synonyms with a Reference Terminology. *Proc AMIA Symp 2000*: 814-8.

[11]Chen H, Lally AM, Zhu B, and Chau M. HelpfulMed: Intelligent Searching for Medical Information over the Internet. *Journal Of The American Society For Information Science And Technology* 2003,54(7): 683 –694.

[12]Kanter AS, Maldonado J, Varshney A, Naeymi-Rad, F A Mediated Health Search Portal for Trusted Medical Content. Medinfo 2004 (submitted).

[13] www.snomed.org

[14]American College of Pathologists. SNOMED Clinical Term Guide: Technical Implementation Guide. 2002. 172-174

[15] Dolin RH, Spackman KA, Markwell D. Selective Retrieval of pre- and post-coordinated SNOMED concepts. *Proc AMIA Symp* 2002; 210-4.

Address for correspondence

Dr. Curtis L. Cole, Weill Cornell Physicians, 1300 York Ave. A123, New York, NY 10021