REPRESENTING CLINICAL INFORMATION IN AN INTERNAL MEDICINE TEACHING IMAGE DATABASE

by

Jason Lyman, M.D.

A MASTER'S THESIS

Presented to the Division of Medical Informatics and Outcomes Research
and the
Oregon Health Sciences University School of Medicine

in partial fulfillment of
the requirements for the degree of
Master of Science

May 2000

School of Medicine
Oregon Health Sciences University

CERTIFICATE OF APPROVAL

This is to certify that the Master's thesis of

Jason Lyman, MD

has been approved.

[redacted]

Committee Chair

Co[redacted]

[redacted]

Committee Member

[redacted]

Committee Member

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

With gracious thanks to my thesis committee chair:

William Hersh, MD

and committee members (in alphabetical order):

Dana Braner, MD
Susan Price, MD, MS
Kent Spackman, MD, PhD

I would also like to acknowledge the assistance of:

Larry Donohoe
Dale Kraemer, PhD
Judith Logan, MD, MS
Mark Mailhot, MD
Daniel Olson
Michael Riben, MD
Melanie Wilson, MLS

This project could not have been completed without the continued love, support and guidance of my future wife,

Kristen Atkins, MD

ABSTRACT

Traditional 35mm slide collections have been used for years to supplement clinical training, particularly in “visually oriented” specialties. As computing power has increased dramatically in recent years, and digital cameras have become affordable and easy to use, many clinicians are building digital image collections for teaching. Digital image libraries offer several advantages to more traditional analog collections, including the potential for distributed simultaneous access by several users, reduced risk of image loss, and more powerful indexing to aid in image retrieval.

Important issues in optimizing flexible retrieval of an image database include the selection of an appropriate data model and the use of a standard terminology to unambiguously represent clinical concepts. No standard approaches to data modeling for a teaching image collection have been identified, and it is unclear which of the many medical terminologies may be most appropriate for classifying clinical content.

A World-Wide-Web based internal medicine teaching image database was designed and developed with particular attention to the data model. A SAPHIRE-based term matching algorithm was used to map terms to the SNOMED 3.5 vocabulary to represent the users' clinical terms. A preliminary evaluation assessed both the semantic model and the feasibility of using SNOMED for this purpose. Study participants, members of the Oregon Health Sciences University Department of Medicine, were generally satisfied with the attributes used to store associated image information. An automated approach to match user terms to SNOMED yielded a successful matching rate of over 70%. Additionally, in 80% of the images indexed, participants indicated that the

SAPHIRE-based algorithm was able to correctly identify SNOMED codes that matched their intended meaning.

This preliminary work suggests a data model that may be useful for others who are developing similar image libraries. Additionally, SNOMED may represent an appropriate vocabulary for representing clinical concepts in such a teaching collection.

INTRODUCTION

The acquisition and analysis of visual information has been central to the practice of medicine for at least as long as history has been recorded. The examination of patients by Egyptian priest-physicians in the third century BC included careful attention to appearance, color, swellings, perspirations, and other signs[1]. The transfer of skills necessary to recognize visual indicators of disease has long relied on bedside teaching methods as well as textual descriptions of physical examination findings. In the Middle Ages, illustrations became a popular approach to representing illness in patients with leprosy, syphilis, etc.[2]. With the availability of photographic techniques to capture visual information circa 1827, it was only a very short time until this approach was used in medicine[3]. In 1845, the first medical publication appeared which utilized these early photographic prototypes, termed daguerrotypes. Twenty years later, surgeon Alexander John Balmanno Squire published the first atlas in medicine which exclusively relied on photographs, with accompanying clinical information about signs and treatments[2].

In more recent years, many clinicians have compiled their own collections of medical photographs, often in 35mm-slide format. These collections serve as source material for clinical lectures as well as an archive for examples of rare or unusual diseases. The difficulties with storage and retrieval in these systems have been acknowledged[4], and some physicians have developed elaborate filing schemes to circumvent these problems[5].

Still more recently, clinicians are taking increasing advantage of the availability of digital cameras and scanners to develop digital image collections. Interest in digital image libraries has spread to many fields within medicine, including radiology[6], pathology[7], dermatology[8], ophthalmology[9], and others[10]. This conversion from traditional analog systems offers significant advantages, including potential for distributed simultaneous access by multiple users, reduced risk of image loss, and more powerful approaches to image retrieval[11]. The benefits of simultaneous access increase dramatically when the World-Wide-Web (WWW) is used as a distribution media, and many institutions have taken advantage of this fact[12-15].

In order to take full advantage of the retrieval benefits of digital image collections, special attention must be paid to two issues related to image indexing: the design of an appropriate data model and the use of a standard terminology.

Data Models

The term "data model" is used in this context to refer to the set of descriptors or attributes associated with a given image. Bidgood et al. have referred to this concept as the *semantic model*, often comprised of administrative data (relating to the image subject), an identifier for the procedure, coding for the procedure type, the date acquired, and the individual or group responsible for acquiring the image[11]. Picture archiving and communication systems (PACS) are often able to automatically extract this "core" data from the DICOM header when images are being transferred among DICOM-compliant devices[16].

It stands to reason that the selection of elements in the data model is primarily determined by the users' needs. Collections that are developed for the purpose of teaching may benefit from more detailed descriptions of the clinical features of the images. A comprehensive data model potentially allows more flexibility in image retrieval, as users gain the ability to search by any of several methods: diagnosis, findings, anatomic location, date, patient demographics, etc. Another important factor in the development of the data model is the means by which the indexing will occur. Thirty-five-mm slide collections used for teaching commonly require manual data entry, while image archive systems have the potential to automatically extract associated information from a number of sources.

Several approaches to the automated indexing of medical images have been described. Lowe et al. have described four approaches to representing image content. Demographic and procedural information can be automatically extracted when DICOM (Digital Imaging and Communication in Medicine)-compliant devices are used. Semantic representation can be performed by parsing the textual information often found in associated reports. Sophisticated algorithms can be used to directly represent the image content by analyzing the colors, textures, and shapes that are present. Finally, knowledge-based approaches integrate procedure information, knowledge sources, and patient-specific information to represent image content in a way that models the approach a clinician may take[17].

Personal or departmental teaching collections are often developed independently of normal patient care processes, and require extra effort for labeling and storage. This requirement for manual data entry necessitates the utilization of indexing strategies that do not place undue burden on busy clinicians for whom time is a precious commodity.

The advantages of a standards-based approach to documenting image and procedural information has been described[18], and proposals have been made towards this end[11]. Digital image collections that arise in the course of patient management may be more amenable towards these efforts that rely on information captured during normal workflow. Formalized approaches towards the development of standardized data models for teaching collections have not been defined, but a review of on-line atlases reveals many common elements. The Dermatology On-Line Atlas (DOIA), for example, lists diagnosis, anatomic location, patient age/gender/skin color, quality for teaching, image source (clinician and university), supervising personnel, and lesion description fields with each image[19]. The Digital Atlas of Ophthalmology, of The New York Eye and Ear Infirmary, stores their images with anatomy, primary area, specific topic (a categorical description), diagnosis, image type, history, clinical features, treatment, differential diagnosis, and links to other web sites with similar information[9].

Standard Terminology

Equally important in the development of a teaching collection is the use of a standard terminology to unambiguously represent clinical concepts and facilitate the retrieval of relevant images. The benefits of incorporating such a terminology are many, including the potential to take advantage of explicit descriptions of the concept hierarchy. This allows the retrieval of concepts at a "higher level" than that at which they were entered, so that an image labeled with an anatomic location of "right arm" could be retrieved by a search of "upper extremity", or an image with a diagnosis of "congestive heart failure" could be retrieved using the term "heart disease". Additionally of value is the ability of such a nomenclature to recognize the many synonyms that exist in medicine. For example, all images with labels including the terms "URI", "upper respiratory infection", or "cold" would potentially be retrieved with a query using just one of these terms.

Many medical vocabularies have evolved over recent years for varying purposes. ICD-9, arguably more properly termed a *classification*, originated to categorize groups of patients for mortality calculations or billing purposes[20]. The Read Codes and SNOMED (the Systematized Nomenclature of Human and Veterinary Medicine) are primarily oriented towards the ability to comprehensively represent a wide variety of clinical concepts in a much more detailed fashion. MeSH (Medical Subject Headings) is used for the indexing and retrieval of biomedical literature in databases such as MEDLINE. The most recent version of the UMLS (Unified Medical Language System) Metathesaurus provides a listing of over 730,000 concepts derived from over fifty disparate medical vocabularies with the goal of allowing "translation" from one terminology to another[21]. Each concept is represented by a concept unique identifier (CUI), organized into a hierarchical semantic network, and is associated with a wide variety of data elements, including its strings and their source terminologies, its data type, etc.[22].

Previous efforts to represent the content of image databases have utilized the UMLS Metathesaurus[13,23], a modified form of the ICD-9[15], and a combination of SNOMED, DICOM, and the SNOMED-DICOM microglossary[11]. Despite these efforts, the most appropriate terminology to represent the image-related information in a collection of general medicine images has yet to be determined.

Several investigators have proposed standards for assessing medical vocabularies. Chute et al have described a comprehensive framework for their evaluation, describing four primary categories of characteristics: general (completeness, comprehensiveness, non-redundancy, etc.), structure (atomicity, compositionality, synonymy, etc.), maintenance (unique, context-free identifiers, responsiveness, definitions, etc.), and administration (coordination, access, funding) [24]. Cimino published a similar comprehensive set of "desiderata" for clinical terminologies which included, for example, broad content, an orientation towards concepts, multiple granularities, and rejection of "not elsewhere classified", to name a few[25].

The requirements for a terminology for use in a teaching image collection have not been well described. In the domain of internal medicine, the list of potential diagnoses, findings, and anatomic locations is a long one, mandating the use of a terminology with broad coverage (comprehensive), but one that also allows representing concepts at varying degrees of granularity (completeness). To optimize retrieval, synonymy and multiple hierarchies (many concepts fit into more than one "category", requiring the ability to have more than one parent) must be well represented.

Comparative evaluations of medical vocabularies with regards to content coverage, completeness, taxonomy, mapping, definitions, and clarity have been conducted in recent years[20,26]. A study described by Chute et al in 1996 was designed to compare the content coverage of several classifications, including ICD-9-CM, ICD-10, CPT, SNOMED 3, Read V2, UMLS 1.3, and NANDA. Inpatient and outpatient clinical records were randomly sampled from four medical centers, and fifty notes were randomly drawn from the initial collection. Textual data was manually parsed in 3,061 unique concepts. Each of the classification schemas was assigned a primary reviewer whose responsibility it was to determine whether or not each concept was represented in that classification. Each concept was given a score dependent upon the extent to which it was adequately represented; 0 was assigned for no match, 1 for an "approximate" match and 2 for a complete match. The concepts were classified into various semantic types: diagnoses, findings, modifiers (stage, severity, acuity, etc.), treatments, and "other" (demographics, administrative, etc.). In each of these semantic types, SNOMED 3 consistently rated the highest scores, receiving a 1.90 for diagnosis, 1.82 for findings, and 1.69 for modifiers. Its overall score, 1.74, compared favorably to the others, with UMLS 1.3 receiving the next highest score of 1.11. Read V2 followed closely behind with a score of 1.05[20].

In 1997, Campbell et al published a follow-up investigation of SNOMED 3.5, UMLS 1.6, and Read 3.1, with an emphasis on *completeness* (the ability of the terminology to represent concepts at sufficient granularity), *clinical taxonomy* (the appropriateness of intra-concept relationships), *administrative mapping* (the ability to map to ICD-9, e.g.), *term definitions* (association of concepts with their definitions), and *clarity* (non-redundant codes). SNOMED 3.5 was found to be "significantly more complete" with 70% of the unique concepts being acceptably represented in the terminology, as compared to 57% for the Read codes and 50% for the UMLS. Similarly, SNOMED was judged as having a richer taxonomy compared to the others. SNOMED lagged behind when it came to clarity, with a duplication rate of 13.9% as compared to 0% for the Read codes and 4.2% for the UMLS. Both Read (40.6%) and the UMLS (36.1%) had better mapping to either ICD-9-CM or ICD-10 than SNOMED (20.7%)[27].

As Campbell et al report, and others have stated[28], comparing the UMLS Metathesaurus with other clinical terminologies may not be appropriate. The stated goal of the UMLS project is to "improve the ability of computers to 'understand' the biomedical meaning in user inquiries and to use this understanding to retrieve and integrate relevant machine-readable information for users"[29], with the implication that the sources of this information may use disparate clinical classifications. It serves not to provide a comprehensive listing of codes to directly describe concepts, but to serve as a link between similar or equivalent concepts in different source terminologies. Despite this, the Metathesaurus has become a popular target vocabulary for many image databases[13,15,30].

SNOMED 3.5, a.k.a. SNOMED International, includes more than 150,000 records organized into twelve different axes: topography, morphology, diagnoses, findings, procedures, living organisms, chemicals/drugs/biological products, physical agents/forces/activities, social context, occupation, and general (modifiers, relationships, etc.). SNOMED concepts are encoded either in a pre-coordinated approach (D7-11050 is "end stage renal disease") or a compositional (post-coordinated) approach. An example of the latter would be the representation of the expression "poisoning by death cap mushrooms", shown in Table 1[31].

SNOMED 3.5 uses codes to represent preferred terms, as described above, but also includes synonyms of those terms with an extension to the code. This allows for the representation of multiple synonyms for a given concept.

The most recent version, SNOMED-RT (Reference Terminology), is in final beta testing, and offers several advantages to its predecessor[32]. By incorporating description logic and the use of multiple explicit hierarchies, the opportunities for redundancy are

Table 1. Compositional concept representation using SNOMED 3.5

SNOMED Code	Concept
DD-84820	Toxic effect from eating mushrooms
G-C001	Due To
L-42221	Amanita phylloides
↓	
SNOMED Expression:	DD-84820 & (G-C001 L-42221)
Concept:	Toxic effect from eating mushrooms, due to Amanita phylloides

significantly reduced, and intra-concept relationships are better represented. Separate tables describe all of the relationships (IS-A, PART-OF, ASSOC-TOPO) of a given concept with other concepts, making the hierarchical organization explicit. The use of description logic allows the system to determine the equivalence of concepts that may, when represented compositionally, use several different codes in combination.

There are several approaches to retrieving SNOMED codes of interest to match a given phrase[33]. Structured data entry allows pre-coding of the SNOMED concepts and guarantees that the appropriate code is retrieved. Natural language processing represent another possibility, by automatically extracting clinical concepts from a textual source and mapping them to the target vocabulary. Finally, commercially available coding tools offer a more direct approach to identifying appropriate codes. SNOCODE by Medsight[34], and Metaphrase by Lexical Technologies[35] are tools that may be used to encode user-entered terms.

While structured data entry may be appropriate for encoding concepts in restricted domains, it is most likely not feasible for an image database for internal medicine patients. Natural language processing tools for SNOMED mapping have yet to be fully developed, though early work by Chute et al appears promising[36]. Coding tools on the market today may play a role in applications such as the image database, but cost and integration issues led to the adoption of another strategy.

SAPHIRE (Semantic and Probabilistic Heuristic for Information Retrieval Environment), first described by Hersh and Greenes in 1990, provides an automated approach for processing textual data and mapping to coded concepts, designed for both indexing and retrieval[37]. By analyzing text for the presence of *concepts* rather than just *terms*, there is increased potential for identifying terminology codes that represent the user's intended meanings. SAPHIRE parses textual data for matching to concepts in the UMLS Metathesaurus. A recent investigation of the use of SAPHIRE to perform automated indexing of radiology reports revealed identification (and appropriate mapping) of almost two-thirds of the important concepts within the documents[30]. SAPHIRE also operates at the command-line level to process individual terms or phrases, returning a list of potential matches that includes the Metathesaurus CUI (concept unique identifier), weighted ranking assigned by the algorithm, and the canonical form of the concept (preferred term). Figure 1 displays the results for the query *heart attack*. Since the UMLS Metathesaurus includes terms from the SNOMED 3.5 terminology, SAPHIRE offers a way to identify SNOMED codes/terms for user-entered terms. There are several approaches to doing this using SAPHIRE. One of the options at the command-line level is to display the source vocabularies along with the CUI, canonical form, and weighted ranking. By limiting the source vocabularies to SNOMED, a way of representing the concept in that vocabulary is guaranteed. This approach is shown in Figure 2 with the same query used previously.

Figure 1. Saphire at the Command Line

```
medir:Saphire4.3 {113}> saphire4.3 heart attack
C0027051 1.693147 Myocardial Infarction
C0155668 1.399075 Old myocardial infarction
C0497237 1.193147 Fear of heart attack
C0018787 1.000000 Heart
C0153500 1.000000 malignant neoplasm of heart
C0153957 1.000000 benign neoplasm of heart
C0699795 1.000000 attack
```

Figure 2. Saphire at the Command Line, With Source Vocabularies

```
medir:Saphire4.3 {129}> saphire4.3 -K heart attack
C0027051 1.693147 Myocardial Infarction MSH99
C0155668 1.399075 Old myocardial infarction ICD10   ICD99   RCD98   SNMI98
C0497237 1.193147 Fear of heart attack ICPC93  SNMI98
C0018787 1.000000 Heart AIR93   ICD10   LCH90   MSH99   MTH    PSY94
RCD98
C0153500 1.000000 malignant neoplasm of heart MTH
C0153957 1.000000 benign neoplasm of heart MTH
C0699795 1.000000 attack AOD95
```

The fourth column of Figure 2 contains the source vocabularies for a given concept. SNMI98 stands for SNOMED3.5. Since some concepts are also represented in SNOMED 2 (SNM2), and labeled as such, both labels must be included in the retrieval (SNMI98 and SNM2). Using the "grep" command in Unix, only those concepts labelled with SNMI98 or SNM2 are returned.

Another option is to target *just* the source vocabularies of interest. Figure 3 reveals the results of the same query with the vocabulary of interest specified at the command-line.

Figure 3. Saphire at the Command Line, Only SNOMED 3.5

```
medir:Saphire4.3 {133}> saphire4.3 -v SNMI98 heart attack
C0027051 1.399075 Infarction of heart, NOS
C0497237 1.193147 Fear of heart attack
C0018787 1.000000 32 HEART
```

Figure 4. Saphire at the Command Line, Identifying SNOMED Terms for a CUI

```
medir:Saphire4.3 {135}> saphire4.3 -S C0027051 | grep SNM
Myocardial infarction AIR93   RCD98   SNM2
Myocardial infarction, NOS SNMI98
MI CSP98   SNMI98
Heart attack, NOS SNMI98
Infarction of heart, NOS SNMI98
Myocardial infarct RCD98   SNM2
Cardiac infarction, NOS SNMI98
Myocardial infarction syndrome SNM2
Myocardial necrosis syndrome SNM2
```

In this instance, no new SNOMED terms are returned, but this is not consistently the case.

A third approach is to take a given CUI returned by a query, and attempt to identify its corresponding SNOMED terms (in the Metathesaurus). This two step process requires that the user chooses the CUI for the concept that best represents their term, and then identifying the matching SNOMED terms for that CUI, using the "grep" command described above. Figure 4 demonstrates this approach, using the top-ranking CUI in Figure 1. Notice that several synonyms are returned as well.

Figure 5. A Portion of the MRSO Table of the UMLS

```
C0000039|L0012507|S0033298|SNMI98|PT|F-63675|3|
C0000052|L0000052|S0575717|SNMI98|PT|F-687A0|3|
C0000163|L0463826|S0576421|SNMI98|PT|F-B2410|3|
C0000163|L0463844|S0576431|SNMI98|PT|F-B2420|3|
C0000167|L0000167|S0576432|SNMI98|PT|F-B2430|3|
```

The most direct approach for retrieving SNOMED *codes*, given a CUI, however, is the MRSO table included in the UMLS. This table links the Metathesaurus CUI's with the codes from the source vocabularies. A portion of this table, limited to rows with the SNOMED 3.5 codes is shown below in Figure 5. Most relevant to this discussion are the CUI identifiers in the first column, and the SNOMED codes in the sixth column.

One may notice that the mapping is not one-to-one. The CUI C0000163 (17-hydroxycorticosteroid), maps to SNOMED's F-B2410 (17-Hydroxycorticoids) and F-B2420 (17-Ketogenic steroids).

The Digital Image Project in the OHSU Department of Medicine

Over the past three years, faculty and housestaff in the Department of Medicine at Oregon Health Sciences University have collected over one thousand digital clinical images. These images typically document interesting physical examination findings on patients seen in the wards or clinics, and are used to support clinical teaching. Senior residents or faculty who give clinical lectures or case-presentations are able to supplement their presentations with these digital images.

Difficulties arose with the Digital Image Project (DIP), however, with technical aspects of image acquisition, storage, and retrieval, as well as limited economic and personnel resources of the department. These digital images were initially stored on non-networked Macintosh computers in a multimedia laboratory. No standard size, file format, or indexing schema was being used in storing the images. Many of the images were not labeled, and the findings and/or diagnosis were unclear. As a result of the lack of consistent image indexing and organized storage, the process of retrieving relevant images had become a time-consuming, often frustrating and unsuccessful process.

Specific Aims

The original goals of the project included the design and development of a database application to facilitate the indexing, storage, and retrieval of these images. Specifically, three primary objectives were targeted:

- developing a data model that allows incorporation of the most relevant information associated with the images without being cumbersome or time-consuming
- choosing an appropriate terminology to represent the clinical concepts.
- developing or adapting a tool to allow mapping the users' terms to the terminology

Two main hypotheses were developed:

- **The data model developed allows for adequate representation of relevant image-related concepts without being overly cumbersome or time-consuming.**
- **SNOMED 3.5 represents a feasible approach to representing the clinical concepts in the application.**

METHODS

Needs Assessment

Several initial meetings were held with members of the Department of Medicine to assess the history of the DIP, including the original goals, the barriers to its implementation, and the current needs of the users. The workflow involved in image acquisition follows several steps. Once a patient is identified with interesting physical examination finding, signed consent is obtained. Standardized forms are used in which patients can select the acceptable uses of the image, including local publications/video only, commercial and non-commercial publications, and commercial and non-commercial broadcast media. After the digital image has been captured, the clinician uploads the image onto a dedicated workstation. Not uncommonly, many images from one patient are obtained and stored. If time permits, they are renamed appropriately, and stored in folders on the desktop.

As an initial step towards identifying potential solutions, a set of system requirements was developed. Functional requirements, which represent *testable attributes*, are displayed in boldface type, while non-functional requirements are presented in normal type.

General

- the software should be expandable (to perhaps include sound and/or video data in the future)
- the software should facilitate collaboration, as chief residents at other institutions have expressed interest in contributing and retrieving images

- the software should be easily accessible from multiple on-campus locations
- the software should require little maintenance
- the interface should be GUI-based and easy to understand
- the response time for viewing/transferring images should be as minimal as possible
- the software should be cross-platform (Mac/Windows)
- a low-cost solution is desired

Image Acquisition from Digital Camera

- **There needs to be a standardized way to download the image onto the local workstation for viewing, editing, and formatting.**

Image Transfer to Database

- **The software needs to support/allow transfer of the image from a local workstation to the server.**
- Image transfer should be as simplified as possible, minimizing the requirements of the user to choose things like directory location, etc.

Image Indexing / Data Model

- **Users desire the following information (metadata) to be stored along with the image:**
  - **patient name (hidden from non-departmental users) or a non-unique patient "code"**
  - **consent number and consent type**
  - **date**
  - **photographer**
  - **finding(s) present**
  - **diagnosis**
  - **patient age, gender**
  - **file type (sound, video, image)**
  - **certainty rating**
  - **quality rating**
- The indexing process should be thorough/complete enough to facilitate retrieval, but not be overly time-consuming

Image Storage

- **a networked solution should be developed, allowing image storage on a local server with access from multiple on-campus workstations**

- **images should be formatted in such a way as to support quality printing (120 pixel per inch resolution), and of adequate size for presentation (800X600)**
- sufficient storage space should be available on the server
- backup should be performed regularly to minimize possibility of loss of data

Image Retrieval

- **users should have the option of browsing a list of diagnoses and/or findings in order to retrieve images**
- **users should also have the option to search specific associated data (as described above, for example, by patient(consent number), date, photographer, diagnosis, finding, or procedure)**
- **users should be presented with the option to see a collection of thumbnail images retrieved by their query**
- **users should be able to view a full-size image when the thumbnail is clicked, and transfer that image to their local workstation**
- this process should require as few screens as possible, and be as simple as possible while maintaining flexible search options.

A review of "off-the-shelf" media asset management systems prepared for the Department of Pathology at Oregon Health Sciences University by Dr. Michael Riben was examined to investigate possible solutions that would meet the user requirements. Comparing the functional requirements with the functionality of these third-party systems suggested that a "home-grown" approach to the DIP might be more appropriate. The benefits of a WWW-based solution have been shown by other investigators, both in terms of cost-effectiveness and the increased opportunities for collaboration[12,13,15]. Work done by these authors also suggests that mapping user queries to standard terminology codes for representation in a database is technically feasible.

Application Specifications

After review with the users, the requirements were translated into a set of application specifications, providing the details for how those requirements would be fulfilled. These specifications are shown in Table 2.

Data Model

Three general categories were considered in creating the descriptor set. Clinical information includes diagnoses, findings, anatomic location, and diagnostic certainty. Demographic information includes patient age and gender and a non-unique patient code. Finally, procedural information contains data about the process, including original filename, photographer name, the clinical department, image resolution, image quality for teaching, and date fields for image production and addition to the database. The decision was made to rename files with a date/time stamp to use as a unique identifier.

There was significant debate within the department about the need for patient codes, but it became apparent that non-unique codes were desired by all, and if developed properly, would not compromise patient confidentiality. Many users reported often remembering a patient's name, and wanting to use that information to aid in retrieval. The benefit, in the users' perspective, was the ability to quickly narrow down a search for an image. A system for creating the codes was developed using the last letters of the patients' last and first name, respectively, combined with the two-digit representation of the month they were seen. As an example, a patient by the name of Janet Doe, seen in May of 1999, would have an associated code of "et05". Since there are potentially numerous patients

Table 2. Application Specifications

Requirement	Specification
General	To allow full indexing potential on an expandable, cross-platform system, a WWW interface will be used as a front-end for an Oracle database residing on a local server
Security and Confidentiality	The database will reside on a local server behind the OHSU firewall. Patient identifiers will not be stored in the database. Consent numbers will be stored with the images – each patient will have a consent form that will be stored locally within the department and provide a link between patient name and consent number. Only faculty/housestaff will be allowed to query using the consent number, or view it with the image. These individuals will have a group ID and password assigned.
Image Acquisition	Acquiring the image off the camera directly from within the application will not be supported. A document will be developed to explain the steps of acquisition, editing, and compression. Both the VA and OHSU hospitals will have one workstation each that will serve as dedicated image acquisition machines.
Image Transfer	Image uploading to the server will be carried out via HTTP into a predefined location on the server. The application will support image uploading from any WWW client.
Image Indexing	Users will be presented with forms on HTML scripted pages. Some may have drop down menus (date, consent type, etc.). Underlying CGI scripts will translate the non-clinical information into SQL statements for database modification. Clinical data (diagnosis and findings) will be mapped to the terminology using an adapted tool, and the user will be presented with the terms for confirmation/modification. The terminology codes will then be represented in the database, along with the original user terms for diagnosis and findings. (See Figure 2)
Image Storage	Oracle 7.3.2, which resides on at least one server in the DMIOR, will be used to store the images and associated metadata.
Image Retrieval	Again, HTML forms with underlying PERL CGI scripts that translate user entries to SQL will be used to query the database.

with names that would yield similar codes, it was felt that this code would not permit the identification of a specific patient.

To facilitate the submission process, drop-down menus, radio buttons, and check lists were used whenever possible. Free text is allowed in the fields for the clinical concepts, photographer name, and department name. All other fields allow the selection of the appropriate data from one of the lists described above.

Standard Terminology

SNOMED 3.5 was chosen as the target clinical terminology for this image database due to its richer taxonomy, greater completeness and greater content coverage, as described previously.

A SAPHIRE-based approach for matching user terms to the SNOMED vocabulary was developed. Since SAPHIRE was not developed for targeting *specific* terminologies of the Metathesaurus, in particular SNOMED, a combination of multiple approaches was required. Two different SAPHIRE-based matching algorithms were developed for the two phases of the investigation. In the first, which was intended to identify matching SNOMED terms in an automated fashion, SAPHIRE was used to return the top-ranking CUI for a given term. This CUI was then matched to SNOMED using the approach exemplified in Figure 4. This allowed the comparison of the original term with a potential matching SNOMED term. For future reference, this algorithm will be referred to Algorithm A. The second approach (Algorithm B), used in the application to match user-entered clinical terms to SNOMED codes, works in a multi-step fashion, building a list of unique CUI's using a combination of the approaches described above. The next step iterates through the array of CUI's, matching each to all of its SNOMED codes in the MRSO table. Finally, a SNOMED "concepts" table is used to generate the term that goes with that particular code. An array of potential SNOMED codes and their terms is generated by this last step. The application returns this list to the user, providing the opportunity to select the most appropriate code(s). The PERL script used for Algorithm B is included in Appendix I.

Evaluation

Two independent, sequential investigations were conducted. The goal of the Phase I study was to characterize the distribution of the semantic types (diagnosis, finding, treatment, etc.) of the terms currently used to describe the images, and measure the ability of the automated SAPHIRE-based algorithm to match those terms to the SNOMED 3.5 vocabulary. Phase II focused on a preliminary evaluation of the data model used by the application, as well as a further examination of the technical feasibility of incorporating a SNOMED term-matching algorithm into the indexing process. At the time of these evaluations, the database was populated only with "test" images, not actual images from the Department of Medicine.

Phase I

At present, the vast majority of the clinical images exist on local folders on a dedicated workstation. No associated information is stored with the images; the only indication of the image contents is found in the filename. These images were all loaded into Extensis Portfolio™, an image management system. A list of the filenames was extracted and loaded into a spreadsheet application (Microsoft Excel™), resulting in 1008 names. Duplicate filenames were discarded. Unique concepts were manually extracted from the filenames. As an example, a filename "hyperpigmentation melanin.jpg" was separated into two unique concepts, "hyperpigmentation" and "melanin". Understandable abbreviations (eg. CLD, RA) were expanded into the full terms (chronic liver disease and rheumatoid arthritis). While these judgments are somewhat arbitrary, an effort was made to keep intact words in combination that represented specific concepts. "CMV retinitis", for example, was judged to be a specific concept, and was not separated into its two terms. These concepts were manually classified into various semantic types (diagnosis, findings, anatomic locations, etc.) by the author.

In order to determine how well an automated SAPHIRE-based algorithm could map to the SNOMED 3.5 vocabulary, the terms were batch processed through Algorithm A, described previously. The result of the PERL script that performed this task was a list of user concepts and SNOMED terms, associated in pairs. Each pair was examined and the appropriateness of the matching was judged as either acceptable or unacceptable by the author. For each unacceptable (or unsuccessful) match, a SNOMED code for the term was manually searched for, and the reasons for the failure were examined.

Phase II

The second phase of the investigation involved an evaluation of the data model used in the database, and an examination of the feasibility of using SNOMED to represent clinical terms in the application during the indexing process. Images were given to eight faculty and housestaff within the Department of Medicine. They were given instructions on the submission process and asked to fill out a survey. The survey consisted of an electronic format, with questions on the web pages of the application, as well as a traditional paper format. The survey was piloted with several colleagues within the department for clarity, length, etc..

Eighty-four images were randomly selected from the current collection in the Department of Medicine of approximately 1100 images. Eight sets of 6-7 purposefully selected images were chosen, with an effort made to include the following: at least two images from one patient, images that were relatively simple (1-2 "straightforward" diagnoses), and more complex images(2-3 complex, unclear diagnoses). These were fairly arbitrary judgements, but it was performed with the goal of providing everyone with a balanced set of images.

Each study participant received either a floppy disk or an e-mail with their set of images, an instruction sheet, and the three page general survey. They were asked to submit the images one at a time, with image-specific information included on the instruction sheet (see Appendix) containing falsified patient names, age, dates, etc. After the SNOMED matching algorithm returned a list of possible codes for the user to review and select, the user was presented with questions on the web page asking about their perception of how well their concepts had been matched to SNOMED. They were also given the opportunity to enter terms that did not match correctly. At the end of the submission process for that image, they were asked to answer more on-line questions about the process and appropriateness of the various fields for that particular image. Screen shots of these questions are displayed below in Figures 6 and 7. In Figure 6, the user-entered term is shown in parentheses next to the header indicating the originating field for the term. Below each header is a list of potential matching codes/terms that the user may select from.

Figure 6. SNOMED-Matching Questions

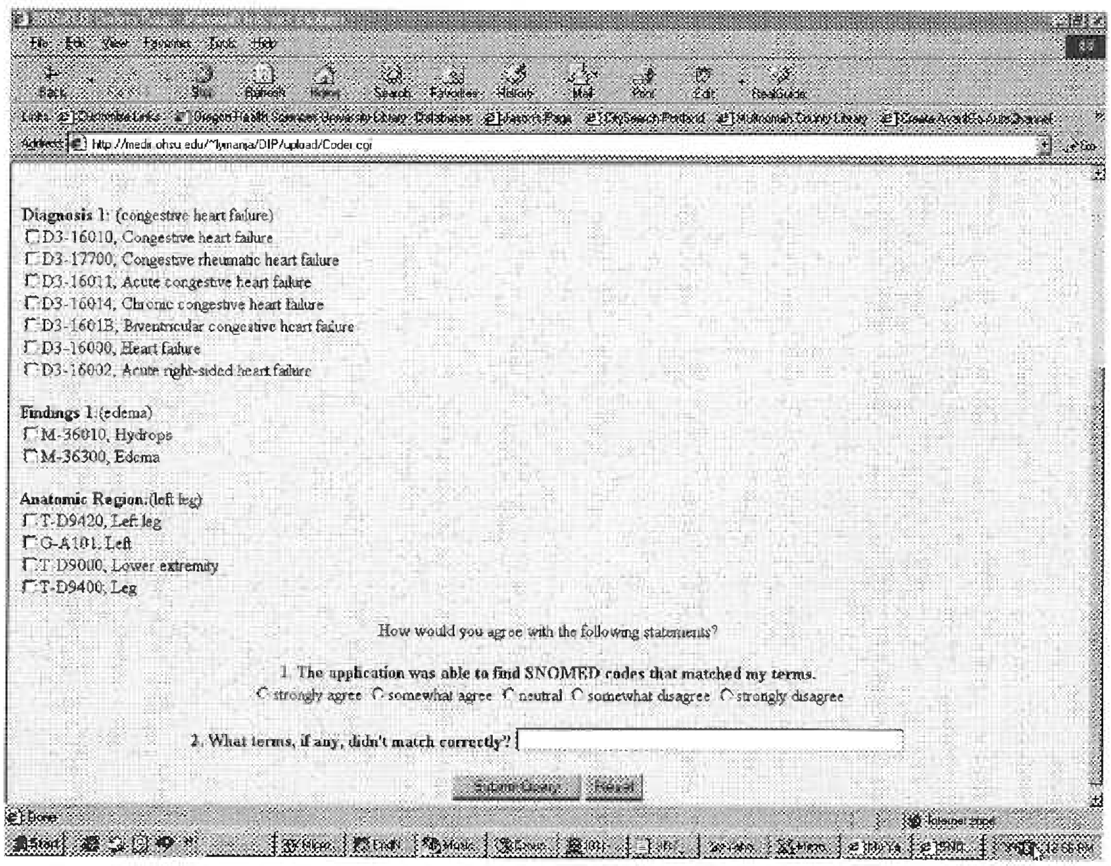
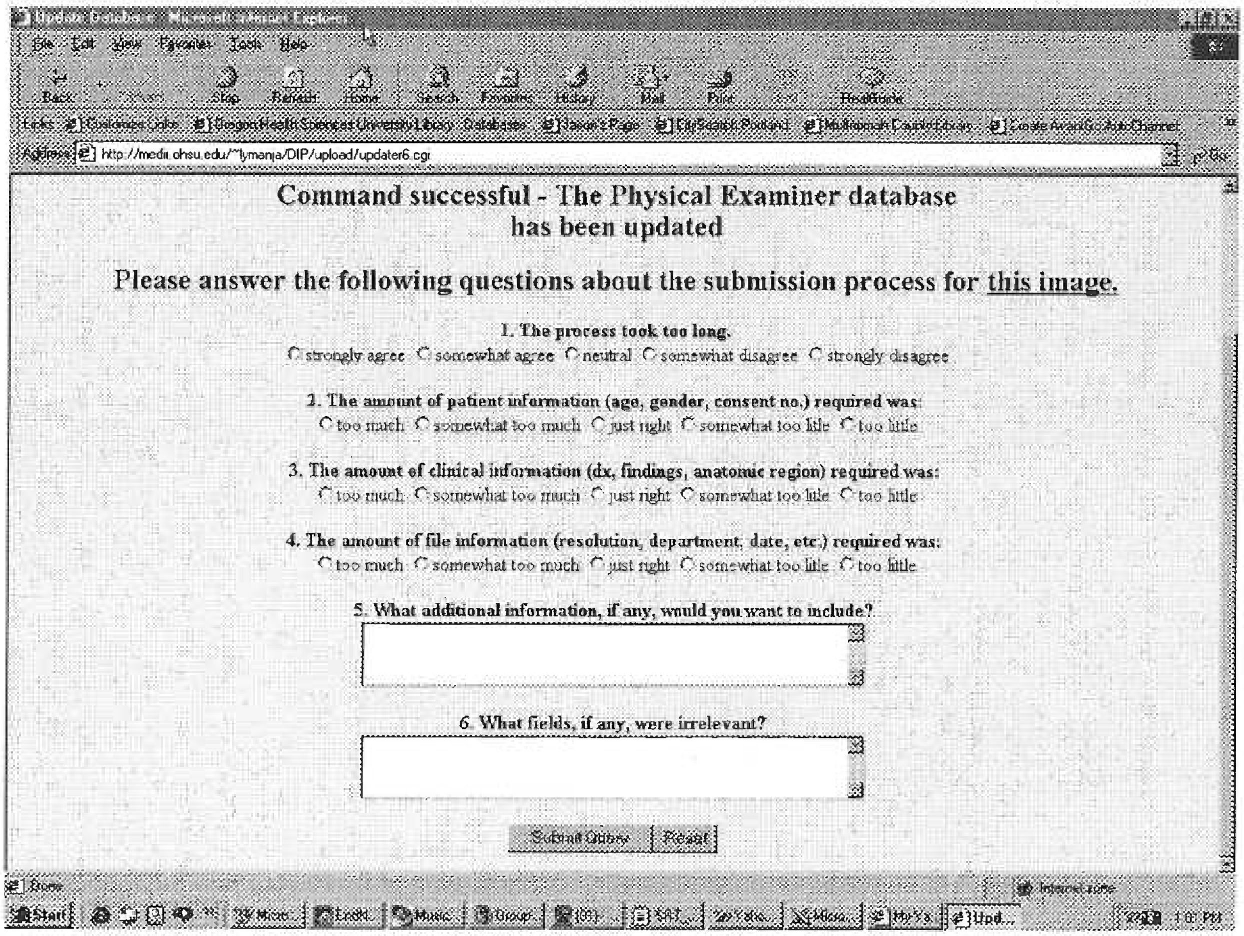


Figure 7. Indexing Questions



In Figure 7, the user is presented with confirmation that the information has been entered into the database, followed with the image-specific questions described above.

After all of the images were submitted, the participants were asked to fill out a paper-based questionnaire (see Appendix II) with more general queries about the data model, the SNOMED matching process, and given an opportunity to rate the importance of each field in the model. Subjects were given numeric identifiers to ensure confidentiality and link their paper-based data to the electronic data.

RESULTS

Application Status

The current functionality of the image database offers the ability for users to upload, index, and retrieve images at the project web site: http://medir.ohsu.edu/~lymanja/DIP/Welcome3.html. Currently, the site is password-protected to allow access only to departmental members. Since there are often many images for one patient, each image is described using a consent number which applies to each image in that session and is specific to a particular patient at a particular time. The image submission process begins with the opportunity to retrieve a new consent number (if the first image from a session is being uploaded), enter the consent number, or bypass the consent process. After retrieving a new consent number, the user is instructed to document that number on the patient's consent form, which is routinely filled out before photographs are taken. Additionally, the user is prompted to enter in the patient code, age, and gender. If the user *enters* a consent number, that patient information is retrieved from the database for confirmation from the user.

After completing the consent process for an image, the user is prompted to browse their accessible drives to locate the image of interest and upload it. After confirming the image, users are prompted to enter in the indexing information described previously (clinical and "file" information). The next script in the process attempts to match the user's clinical terms to SNOMED, allowing selection of the most appropriate code(s) as shown in Figure 6. Finally, the user is presented with a data confirmation page, and

Table 3: Database Tables and Fields

*terms mapped to SNOMED codes

EXAMINER table	CONSENT_INFO table	SNM_Codes table
image identifier (KEY)	consent number (KEY)	image identifier (KEY)
original filename	patient code	SNOMED code (KEY)
*diagnosis (3 fields)	patient age	
diagnostic certainty	patient gender	
*findings (3 fields)		
*anatomic location		
photographer		
department		
image resolution		
image quality for teaching		
date image taken		
date image added to database		
consent number		

prompted to submit the information into the database if it is correct. The current organization and composition of the tables in the database are described in Table 3.

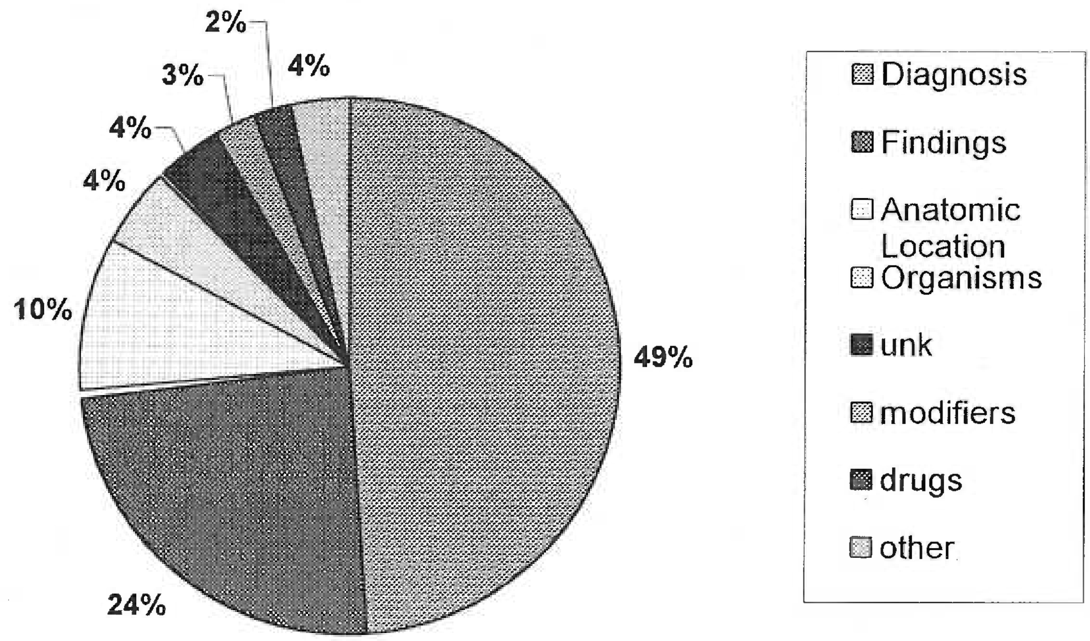
To retrieve images, users have the option of browsing a list of images organized alphabetically, or searching by any combination of queries on specific fields. Currently, searching can be performed on diagnoses, findings, anatomic location, photographer, department, image date (both date added to the database as well as image acquisition date are searched), consent number, original filename, or patient code. When multiple fields are filled out, they are combined using "AND" in the ultimate SQL statement. Some fields can be searched using a wild card, "*" to designate any combination of characters. For example, a search for a diagnosis using the term "*acne*" will retrieve images indexed with terms such as "cystic acne" or "acne vulgaris".

Query results are returned as rows of image thumbnails with all the metadata. Users can modify portions of the associated data by clicking a "modify" button at the end of the row, which returns a larger picture of the image, as well as a form with the original data supplied as default values. Currently, since the site is password-protected, any user may modify the data.

Evaluation – Phase I

An analysis of the current image collection reveals that there are approximately 1,100 digital images. Typical diagnoses represented include gouty tophus, rheumatoid arthritis, Stevens-Johnson syndrome, Herpes zoster, etc.. The vast majority were taken with a digital camera and stored on a local workstation. The only textual information associated with the images is found in the filenames, and may include a diagnosis, finding or anatomic location, or some combination of these. Typical filenames are "erythema nodosum2.jpg", "finger lesion.jpg", "endocarditis/petechia.jpg", etc..

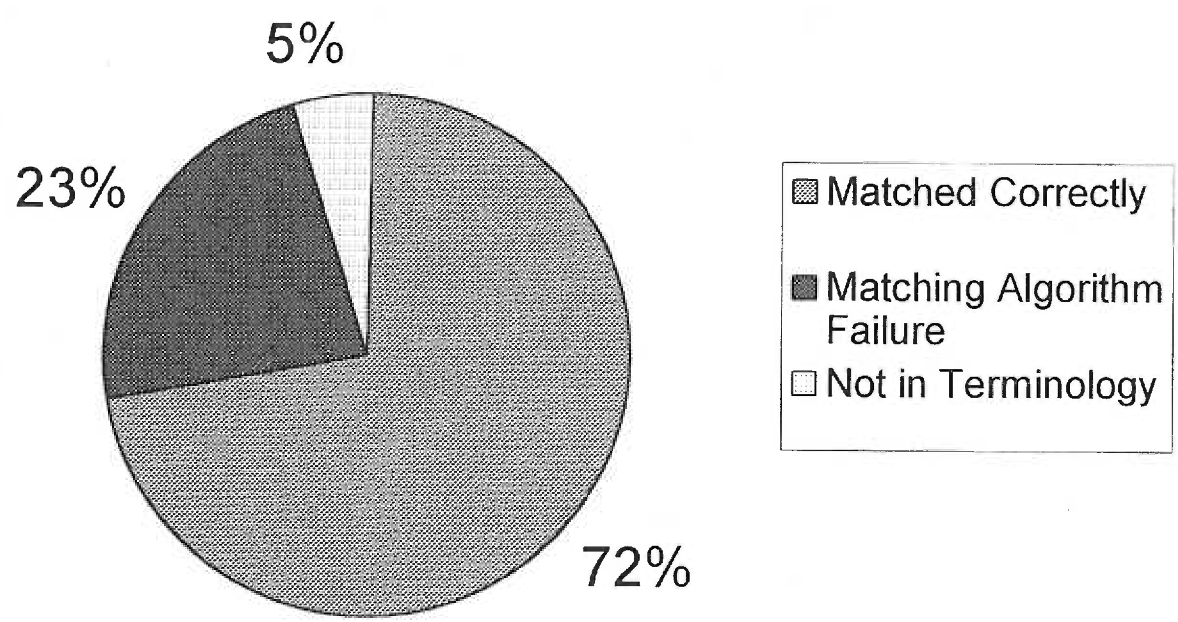
Figure 8: Distribution of Semantic Types



There were 273 unique concepts identified in the original user-created filenames. The vast majority of these terms represented diagnoses (49.0%), findings (23.8%), and anatomic locations (9.8%). Figure 8 displays the distribution of data types

Of the 273 concepts, 197 (72.2%) automatically matched correctly to concepts in the SNOMED vocabulary. There were two primary reasons that 76 terms were not correctly matched. Either the algorithm was unable to identify the correct SNOMED concept from the user's term, or the concept did not exist in that part of the SNOMED vocabulary represented in the UMLS. Matching failures included both term misspellings (e.g., erysipelis) and term phrasing (e.g., "Bells" rather than "Bells palsy"). It also included matching errors for reasons more difficult to pin down. For example, the term "lazy eye" mapped correctly to the UMLS Metathesaurus, but not to its synonym, "strabismus," in SNOMED. It is a matter of opinion whether this error is due to the user's choice of the lay term, a failure of UMLS to associate the term with its more scientific synonym, or a deficit in the SNOMED vocabulary. Figure 9 shows the results of the SAPHIRE matching process.

Figure 9: SNOMED Term Matching



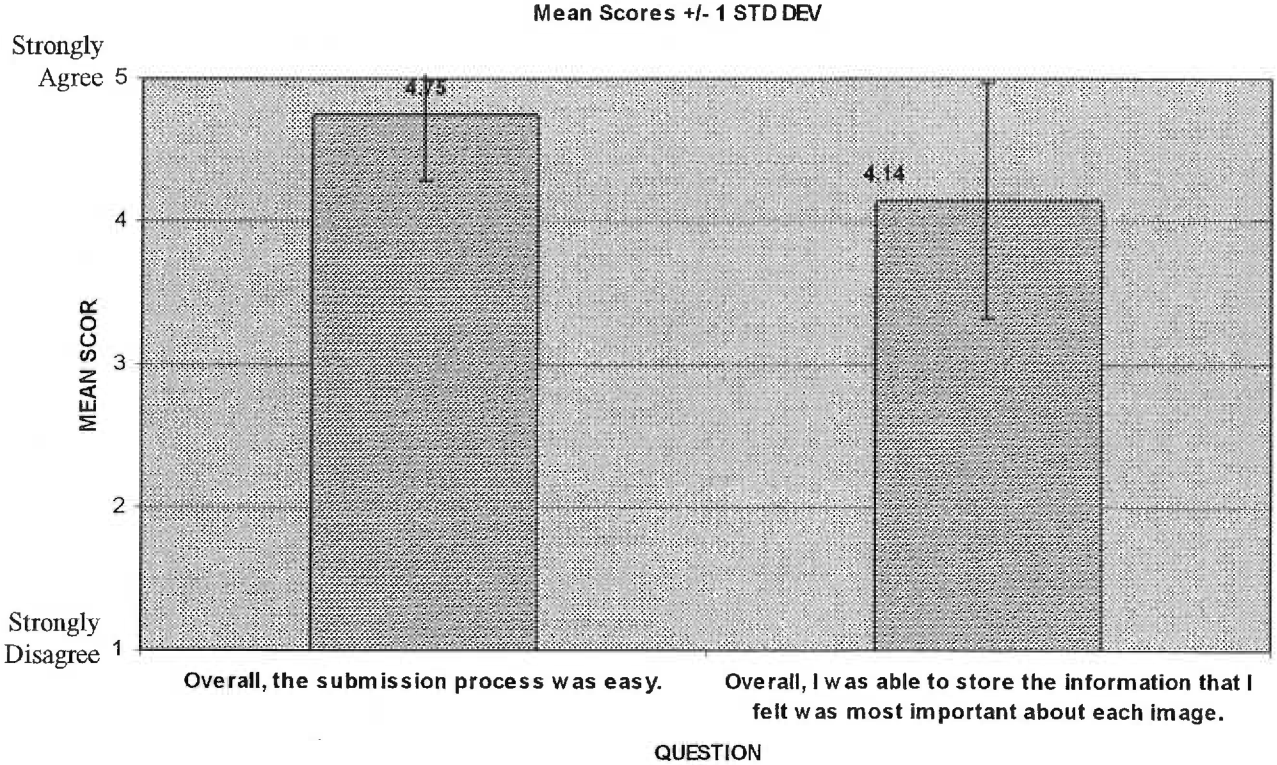
Evaluation – Phase II

The goals of the Phase II evaluation were to evaluate the data model, both from a usability perspective as well as how well it allowed study participants to represent meaningful information in the database. In addition, the feasibility of the SAPHIRE-based SNOMED matching algorithm was examined, with respect to its integration into the indexing process as well as its ability to identify appropriate codes based on users' terms.

The Data Model

Figure 10A displays the mean responses to the questions "Overall, the submission process was easy." and "Overall, I was able to store the information that I felt was most important about each image." Using a Likert scale with a score of 1 representing *strongly disagree* and 5 representing *strongly agree*, users had favorable responses.

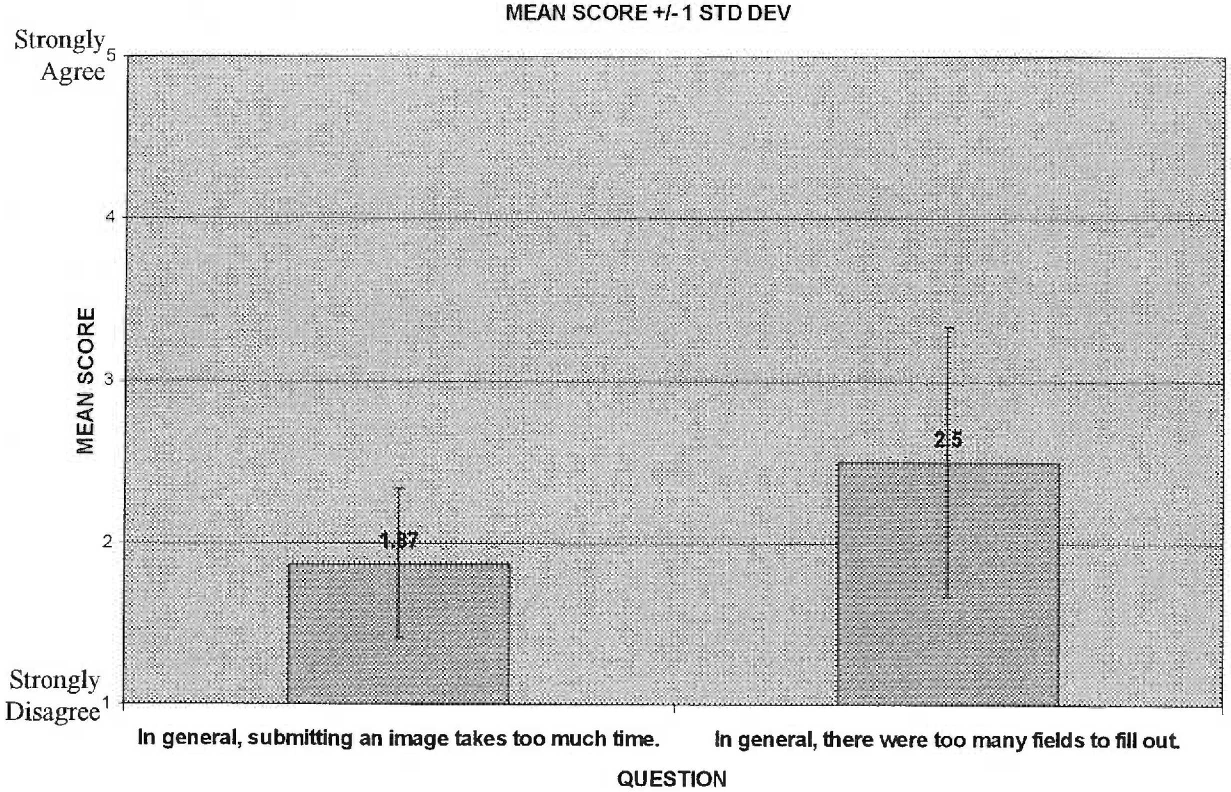
Figure 10A. The Data Model
Mean Scores +/- 1 STD DEV



Similarly, participants were asked questions designed to examine how cumbersome the indexing process was, such as "In general, submitting an image takes too much time", and "In general, there were too many fields to fill out." The same Likert scale was used, with responses shown in Figure 10B. Users generally did *not* think that too much time was required. The response to the *number of fields* was more equivocal; there was, however, no strong evidence that they felt there were too many.

Users were also given the opportunity to indicate how appropriate they felt the *amount* of information requested in the indexing process was in each of the three main categories: clinical information, patient information, and file information (image resolution, photographer, department, etc.). This question was asked in an "image-specific" manner, after each image was submitted, but also as a separate general question after they were

Figure 10B. The Data Model
MEAN SCORE +/- 1 STD DEV



finished indexing all of the images they were given. The responses to this latter question are shown in Figure 10C. As in the previous data, these are mean responses averaged over the eight study participants. Generally, the responses are close to the midline value, labeled "Just Right" on the questionnaire. The amount of clinical information, however, is somewhat oriented towards the "too little" end of the spectrum. This question was also asked after each image, and the mean responses *for each subject* are shown in Figure 10D. The desire for the ability to add more clinical information isn't as apparent here, with six out of the eight subjects indicating satisfaction with the amount of clinical information requested by the application, with no variability in those responses. Also interesting is that six of eight users were satisfied with the amount of *file* information, one indicated a "2" with no variability across his or her images, and one had a mean of 2.5 with a confidence interval that included the midpoint.

Figure 10C. The Data Model
MEAN SCORE +/- 1 STD DEV

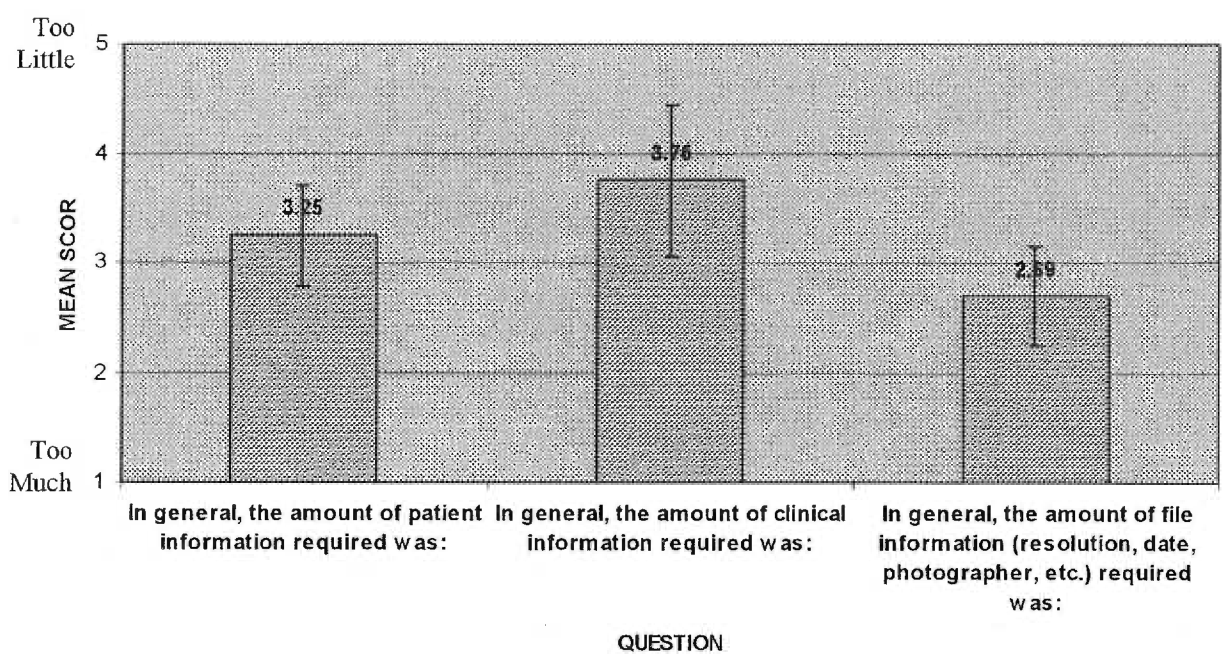
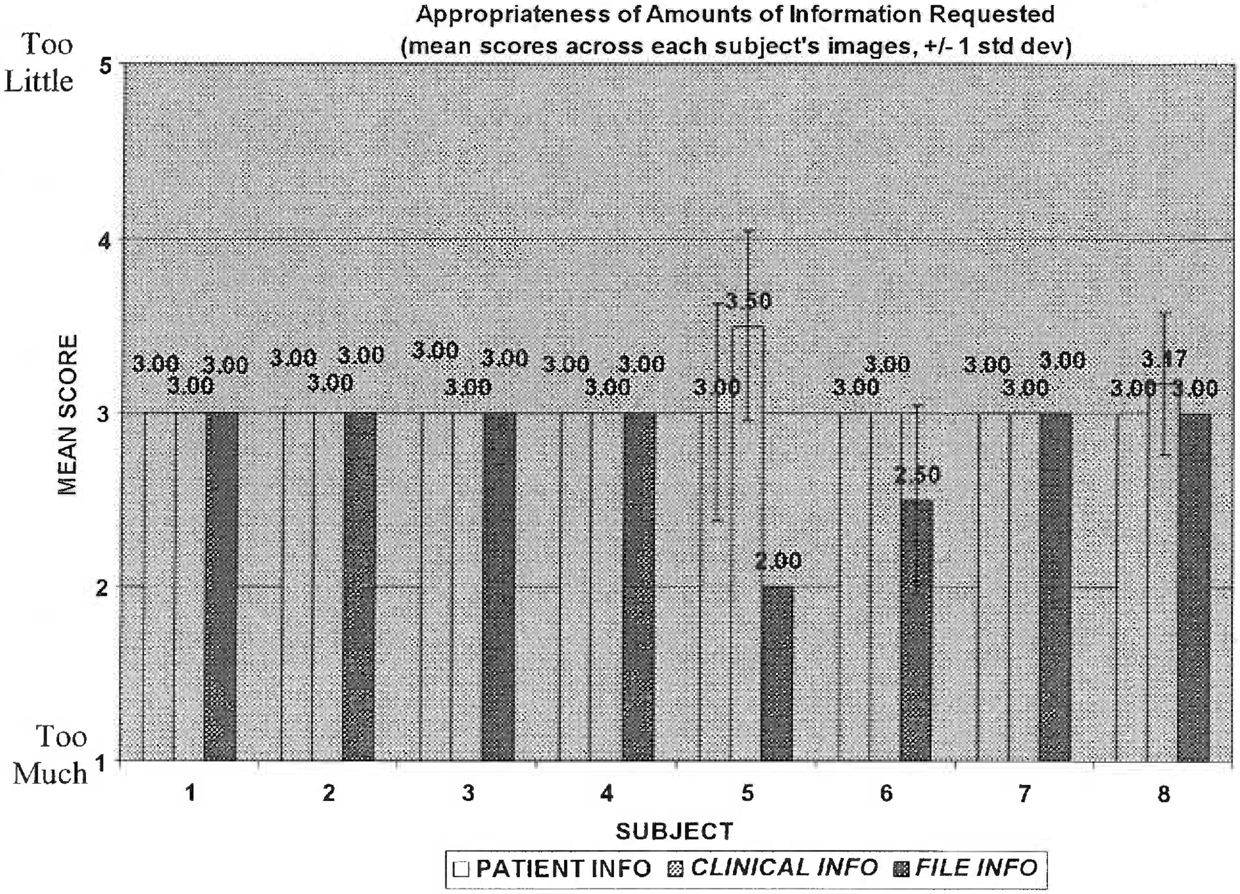


Figure 10D. The Data Model
Appropriateness of Amounts of Information Requested
(mean scores across each subject's images, +/- 1 std dev)



In order to determine the perceived importance of each field in the database, participants were asked to rate each field on a scale of 1 ("very unimportant") to 5 ("very important"). Results are shown in the next figures, divided by category for clarity. In Figure 11A, the clinical fields show a noticeable decline in perceived importance with each added diagnosis and finding field, though significant variability is noted. Table 4 is added to display the frequencies of responses to each Likert score; the variability in the second and third "findings" field is partially explained by one participant who rated them both as "very unimportant."

Figure 11A. Importance of Fields - Clinical Information
(mean score +/- 1 std dev, averaged across all users)

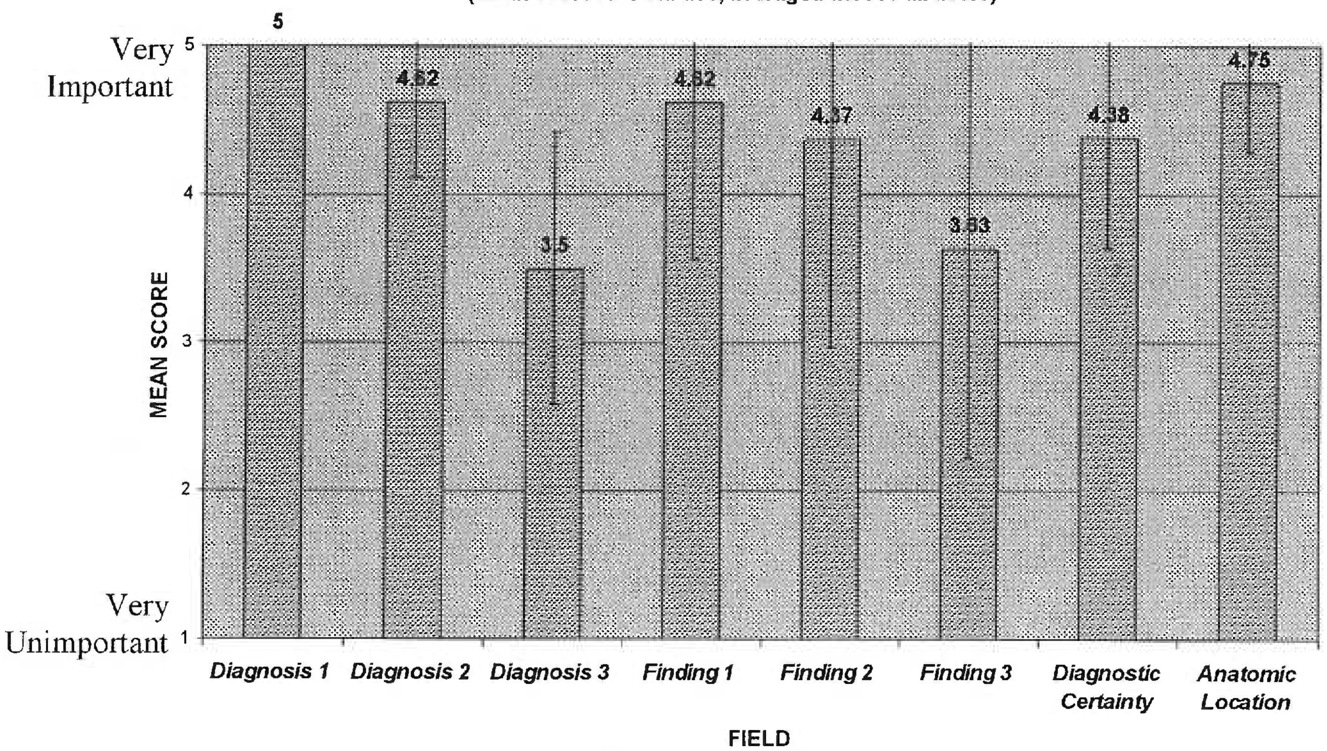


Table 4: Distribution of Responses – Clinical Fields (DX = Diagnosis, FDG = Finding)

LIKERT SCORE	DX 1	DX 2	DX 3	FDG 1	FDG 2	FDG 3	Diagnostic Certainty	Anatomic Location
1 (Very unimportant)					1	1		
2				1				
3			6			3	1	
4		3			1	1	3	2
5 (Very important)	8	5	2	7	6	3	4	6

The responses for the *patient information* fields are presented next in Figure 11B. There is general agreement among all users that these fields are more important than not, though Frequencies are shown in Table 5 for the "patient code" and "consent number" fields which had greater variability than the other two.

Figure 11B. Importance of Fields - Patient Information
(mean score +/- 1 std dev, averaged over all users)

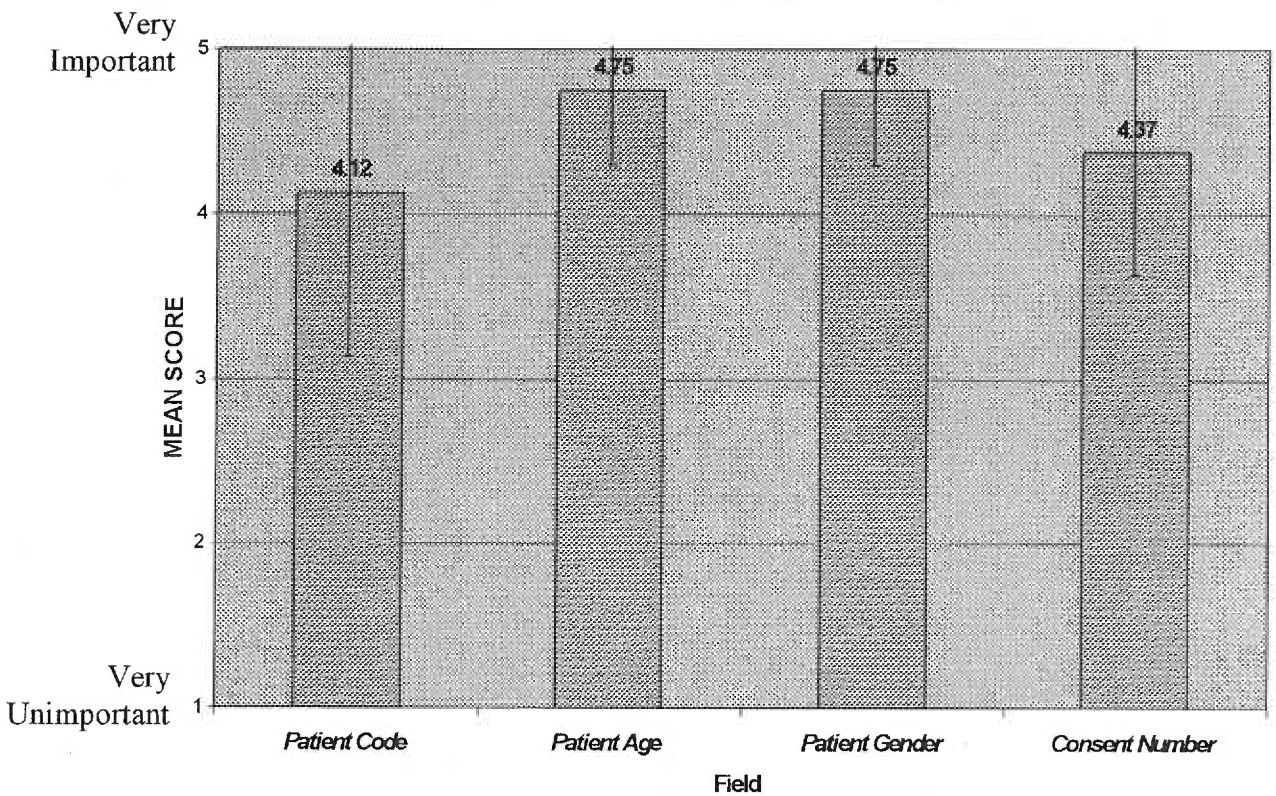


Table 5: Distribution of Responses - Patient Fields

RATING	Patient Code	Consent Number
1 (Very unimportant)		
2		
3	3	1
4	1	3
5 (Very important)	4	4

Figure 11C shows the responses for the perceived importance of the *file information* fields. Mean scores are noticeably less than the fields in the previous two categories, with significant variability, making conclusions difficult to draw. Table 6 shows the response frequencies for these fields to better describe the variability.

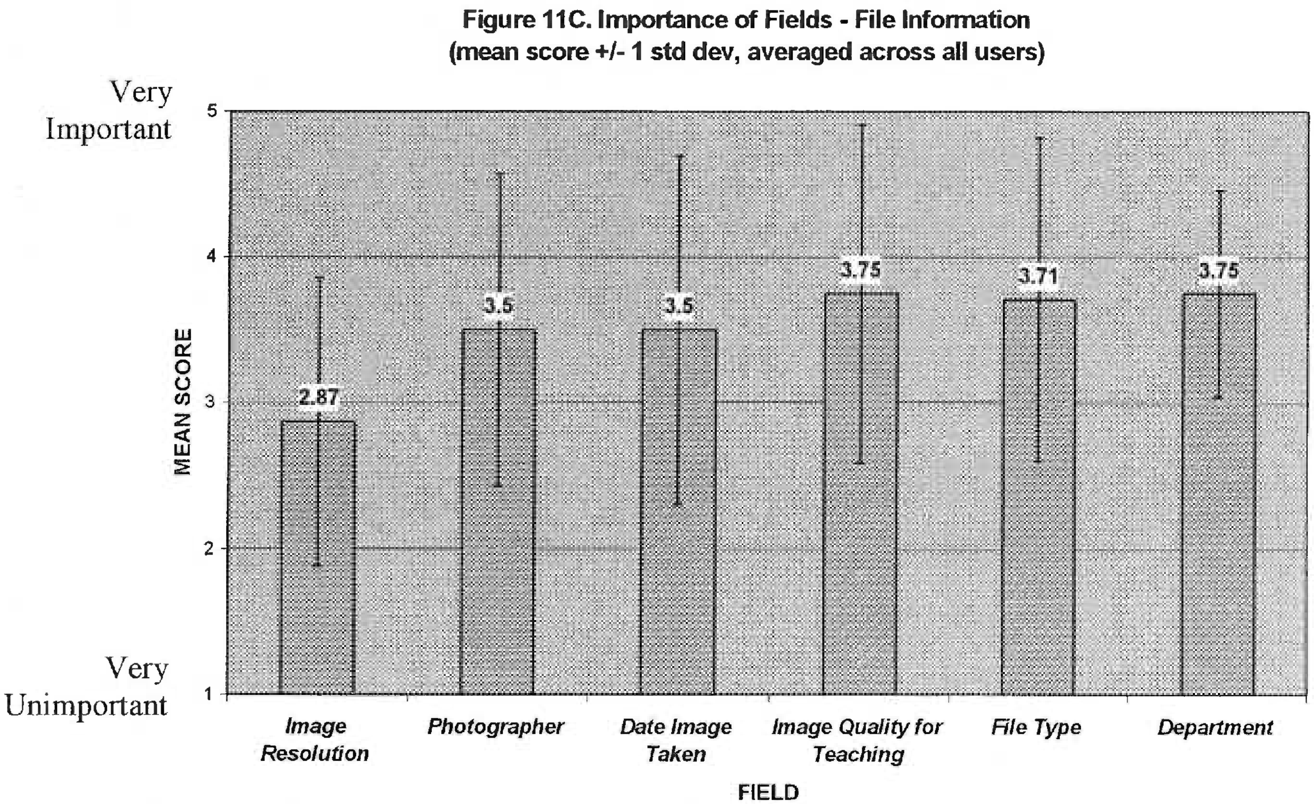


Table 6. Distributon of Responses - File Information

RATING	Image Resol-ution	Photo-grapher	Date Image Taken	Image Quality for Teaching	File Type	Dept.
1 (Very unimportant)	1					
2	1	1	2	2	1	
3	4	4	2		2	3
4	2	1	2	4	2	4
5 (Very important)		2	2	2	2	1

Study participants were given the opportunity to indicate what additional kinds of information they would have like to store with the images, or what additional fields they would like to see added. Samples of the responses are shown in Tables 7 and 8. Study participants were also asked to indicate which fields, if any, were unnecessary or irrelevant, as shown in Table 9.

Table 7. Image-Specific Additional Desired Information

What additional information would you have liked to be able to store with the image?
"mode of injury"
"appeared to spare upper arms"
"hospital?"
"A field to give a narrative description of the patients presentation Need more ability to use descriptive terminology to explain what we are seeing - narrative field. An ability to be descriptive in explaining finding"
"descriptions of lesions - size, color, etc"
"tests"

Table 8. Additional Desired Fields

Are there additional fields that you would like to see added? If so, what fields?
"Perhaps something for a 1-2 sentence free text description of symptoms? i.e. 45 y.o. ♀ with abd pain, nausea, + renal dysfunction"
"a free text field to give short narrative of patient presentation & findings"
"ways to describe lesions better. (ie color, size, texture)"
"Perhaps Hospital/Clinic Location, but I'm not sure it'll always be worth the extra time"

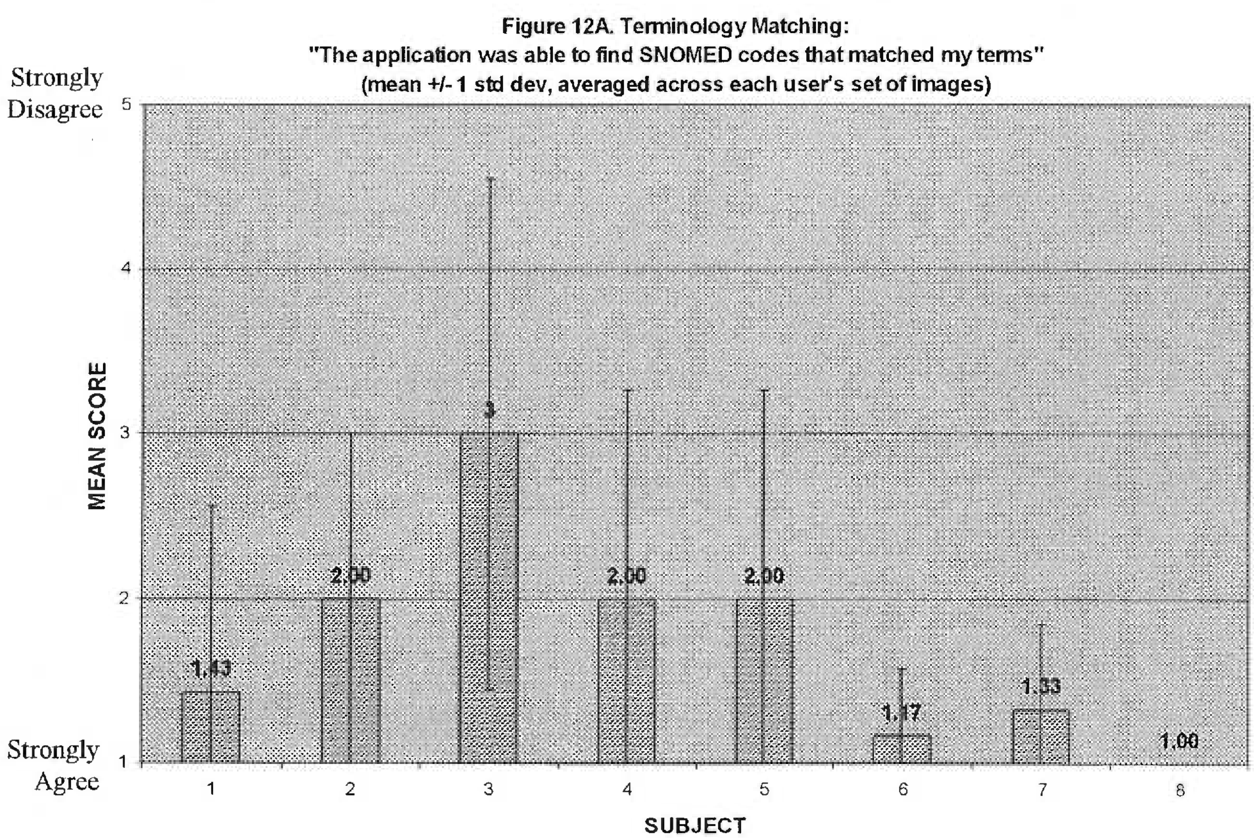
Table 9. Unnecessary or Irrelevant Information

Are there fields that you think are unnecessary and would like to see removed? If so, what fields?
"'findings' are almost redundant"
"image resolution->hard for me to judge"
"image resolution, actual date picture taken (month, year ok) date irrelevant"
"Image Resolution - cant it tell automatically?"
"image quality?"
"image resolution not necessary"
"findings field is again irrelevant as the picture is worth a thousand words"
"for the generic person, not sure we need the 'image resolution' category needs to be there when the 'image quality' field is there (like the 2nd better)"

Terminology Matching – SNOMED feasibility

Questions targeting the feasibility of SNOMED in this phase of this investigation were split into "image-specific" questions that users were prompted to answer while indexing the image, and general questions to be answered after all images were submitted.

Figure 12A shows the mean responses (per participant) to the question "the application was able to identify SNOMED codes that matched my terms" that was asked for each image indexed. While the means were generally below the midpoint value "3", the variation was often significant enough to result in the midpoint being included in the confidence interval 50% of the time.



The response frequencies are displayed in Table 10. When the data from Participant 3 was reviewed, the image that resulted in a score of 5 for the SNOMED matching actually was stored with codes that were very appropriate for all the terms, raising the possibility that the individual had intended to respond "strongly agree" rather than "strongly disagree."

Users were prompted to enter terms that the matching algorithm was unable to find SNOMED codes for. These terms included *onychonychia, prayerhands, Henoch Schonlein Purpura, varicose veins, drug erruption, ankle excoriations, infected emboli, erythematous papular lesions, squamous cell lung cancer, psoriatic plaque, erythematous patch, hirsuitism, hairline, bifrontal, facial,* and *Rash.*

In addition to the measuring the success of the algorithm in matching terms to SNOMED codes, the extent to which the matching process was integrated with the rest of the indexing process was evaluated.

Table 10. "The application was able to find SNOMED codes that matched my terms"

	Participant (with 5-7 images each)								
LIKERT SCORE	1	2	3	4	5	6	7	8	Total Count
1 (Strongly Agree)	6	1	1	3	3	5	4	6	29
2		1	2	1	1	1	2		8
3		1		1	1				3
4	1		2	1	1				5
5 (Strongly Disagree)			1(?)						1
Total Images:	7	3	6	6	6	6	6	6	46

Figures 12B and 12C display the mean responses to questions addressing this aspect of the evaluation. Participants felt strongly that the matching process was well integrated into the indexing process, and also indicated that they felt the potential benefits of using SNOMED outweighed any additional time required, though there was more variation to this response. Similarly, there was strong disagreement, with little variation, with statements suggesting the matching process took too much time.

Figure 12B. Terminology Matching
(mean +/- 1 std dev, averaged across all users)

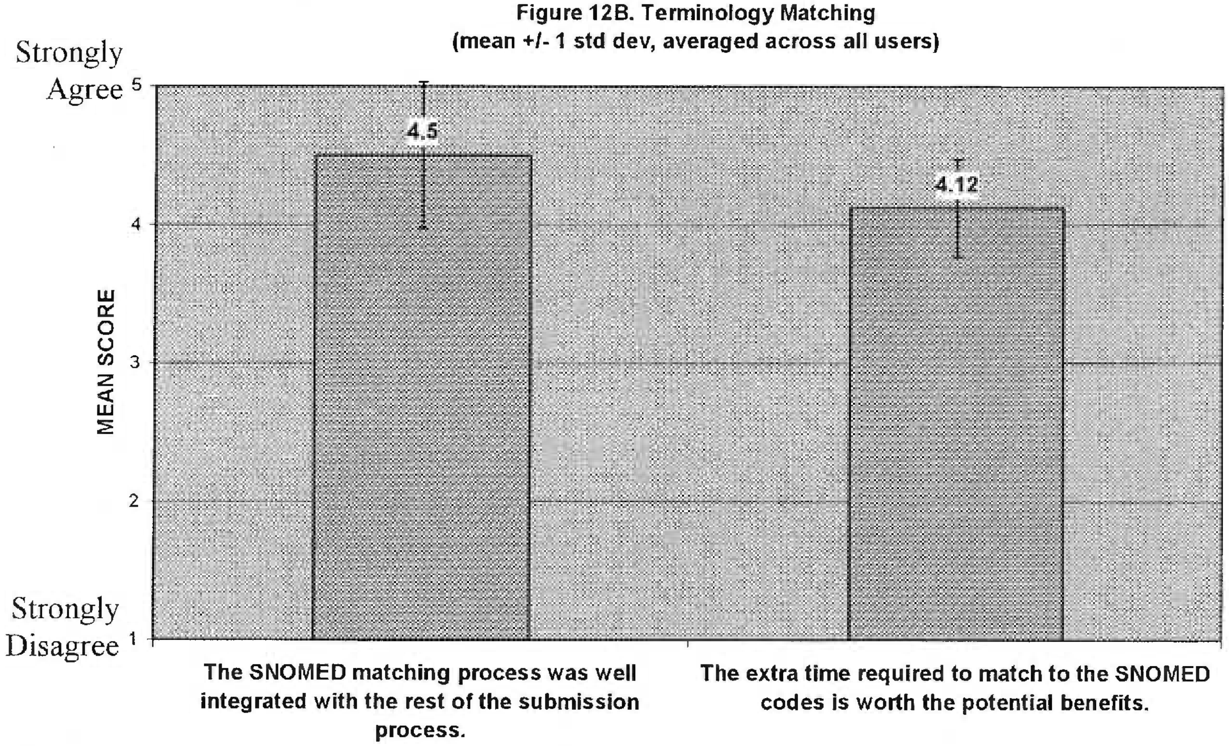
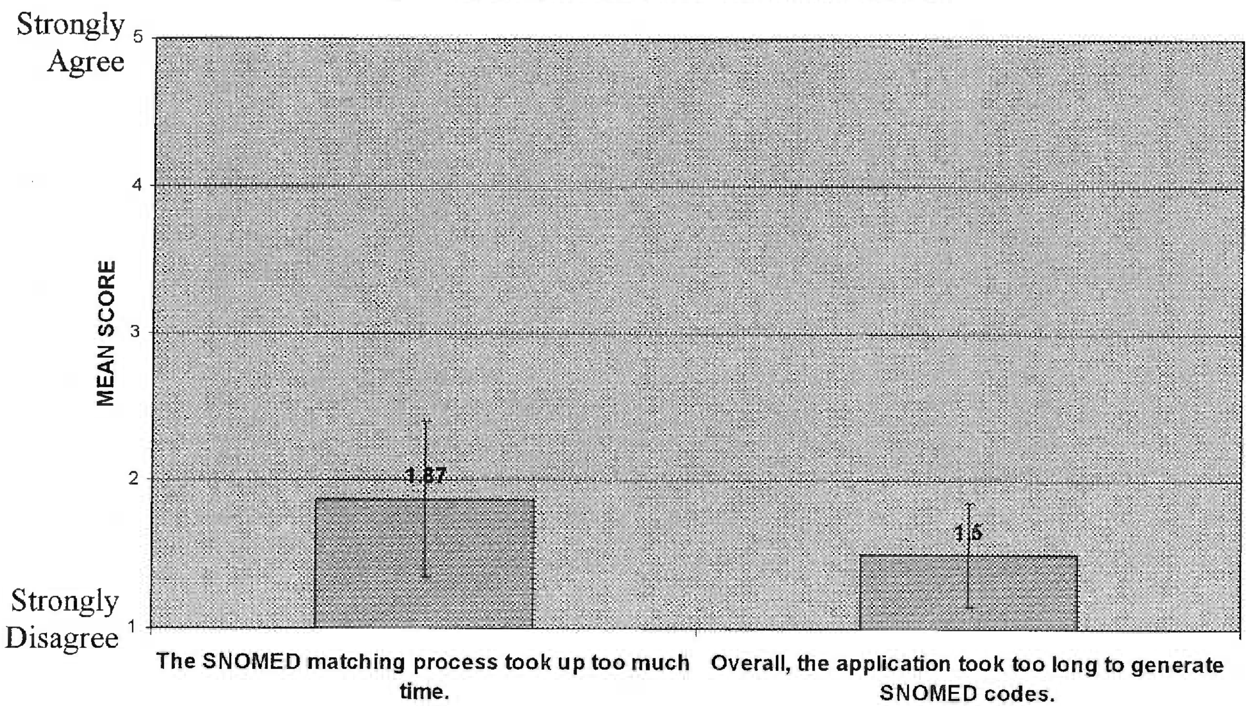


Figure 12C. Terminology Matching
(mean +/- 1 std dev, averaged across all users)



DISCUSSION

Semantic Types

A review of the semantic types currently used in describing image content is instructive. Drawing significant conclusions is somewhat limited by the recognition that when forced to limit information to the contents of a filename, users are considerably constrained in the amount of information they can include. It does, however, force users to prioritize the most critical types of information, and it is interesting, but perhaps not particularly surprising, that almost 85% of the information is either a diagnosis, a finding, or an anatomic location. Eleven concepts were abbreviations that were unclear and therefore could not be classified. Additionally, the classification process can be somewhat arbitrary; the distinction between diagnosis and finding is not always clear. This difficulty has been recognized and described by Greenes et al.[38]. Similarly, it is difficult to know whether to categorize the term "Varicella" as an organism or as a diagnosis without access to more information.

The Data Model

Overall, users appeared quite satisfied with the data model used in the image database. The submission/indexing process was generally felt to be easy, and users felt it allowed the storage of the most important image-related information. When asked in general terms about the appropriateness of the amounts of patient, clinical, and file information requested in the indexing process, mean scores indicated a tendency toward *too little* clinical information, but no evidence of *too much* or *too little* file or patient information.

Free text responses about the desire for storage of additional information suggests a need for the ability to represent the clinical history and perhaps more descriptive data about findings.

Most of the clinical fields were judged to be important, with the questionable exceptions of the third diagnosis and findings fields, which had fairly large standard deviations. All of the patient information fields were considered important, with the greatest variation in the *patient code* field. The scores for this field were somewhat bimodal in distribution, with three participants scoring a "3", one a "4", and four individuals a "5". One possible explanation for this is the fact that not all study participants were involved in the initial needs assessment meeting in which this attribute was discussed. Due to the confidentiality of the survey, however, this cannot be known for certain. The mean scores in the file information fields tended to be lower, but only one field, *image resolution* had a mean below the midpoint. One contributing factor to the low score for this field was that, unlike in real use, the users were given images that had *already* been compressed, and were not given the information about that process. Thus, while more sophisticated users may have been able to determine the resolution by looking at the image file sizes, others may not have known how to make a judgement. Indeed, at least one comment, "image resolution -> hard for me to judge", supports this notion.

Terminology Matching

A significant portion (72%) of the clinical terms used to describe the original images in the teaching collection was successfully matched to terms in the SNOMED 3.5

vocabulary during the Phase I investigation. Considering the simplicity of this automated approach, a 72% success rate is quite promising, indicating that SAPHIRE might well prove satisfactory as a way into the SNOMED vocabulary of the UMLS. Since 95% of the concepts were represented in SNOMED, it may serve as appropriate terminology for this type of application.

Terminology deficits comprised 5% of the matching failures, and can only be improved by addition of those terms to the vocabulary. User term phrasing/spelling is a somewhat thornier issue for this particular application. While the use of pick lists during the indexing process would theoretically reduce this problem, the sheer breadth of the general medicine domain reduces the feasibility of this approach.

One potential limitation in the Phase I investigation was the role of the author as judge of the matching success. The question of who is best suited to judge terminology matching resembles the question in the information retrieval domain about who best to judge relevancy of retrieved items[39]. The individual who documents image content using specific terms may be attempting to express a concept that would not be recognized by a third party. In this sense, the person best suited to determine the success of a terminology match may be the same individual who entered the term.

In Phase II, users were given the opportunity to judge the appropriateness of the SNOMED codes returned by the algorithm, as well as indicate their opinions on the successfulness with which the matching process was integrated with the rest of the indexing process. Generally speaking, participants did not feel that the addition of the SNOMED term matching process was too time-consuming, and felt that the potential benefits outweighed any extra time required. There was strong agreement with the sentiment that the matching process was well integrated in the application. In 80% of all images, there was agreement with the statement that the application had identified appropriate codes. Terms that users indicated matched poorly did so for a variety of reasons. These terms were informally reviewed, and errors were found to be due to spelling, SNOMED terminology deficits, and mapping errors in which the concept existed *both* in the UMLS and in SNOMED but had not been retrieved. Substantive conclusions cannot be drawn with such a small set of terms, but this may represent an opportunity for future analysis.

Using SAPHIRE as an entry point into the UMLS Metathesaurus with the ultimate goal of identifying SNOMED codes may have certain advantages. By originating a terminology matching in the Metathesaurus, and generating a CUI list for a search, mapping to other vocabularies is allowed. One potential disadvantage is that a successful match must rely on accurate mappings between the Metathesaurus and SNOMED. The mappings, however, are frequently one-to-many, and in some cases not sufficient to identify concepts in the target vocabulary. Indeed, some of the preliminary matching failures were somewhat puzzling, in that the algorithm was able to identify an appropriate CUI, but not able to find the correct SNOMED code, even though it could be retrieved manually. Additionally, because SNOMED allows compositional concept representation, concepts in the Metathesaurus may not have pre-coordinated codes in SNOMED. As an example, the term "squamous cell lung cancer" has a CUI in the UMLS Metathesaurus, but in SNOMED must be represented by the combination of codes for "squamous cell carcinoma", "has-topology", and "lung". Currently the UMLS terminology mapping table is not able to equate these two concepts perfectly.

Limitations

Results from this two phase investigation must be interpreted somewhat cautiously for several reasons. In the Phase I evaluation, the author was the sole judge of the success of the automated SNOMED matching. The potential for bias exists, as well as the opportunity for incorrect assignments without knowing for certain the intended meaning of the users' terms. Additionally, the classification of concepts into semantic types was done only by the author, and may have inaccuracies for similar reasons.

The Phase II investigated was the first evaluation of a prototypical system. Occasional "bugs" were noted by users and may have interfered with their data entry. Since the application was not yet in active use, users had to deal with a "learning curve" as they began using the system. Some participants received a brief training session, while others did not (often, schedules would not permit it).

The survey, while piloted by several colleagues, was a newly-developed one that was not fully validated. Questions that targeted similar concepts tended to have similar responses, suggesting internal consistency of the instrument.

Additionally, a small set of images was indexed. It is difficult to draw substantial conclusions about both the adequacy of the data model as well as the success of the terminology matching with only forty-eight images. Additionally, the indexing process may not have accurately modeled the approach users will take when submitting images. Most likely there will be more images from a single patient in one set, reducing the need for data entry compared to six to seven images from different patients. Questions about how much time the process took led to instructive answers, but no data is available on how much time the process actually took, or how much time users would have available to submit images during their routine schedules.

While some meaningful data was gathered, some of the responses were associated with high degrees of variability, making conclusions difficult to draw. A qualitative approach to the evaluation would have offered the opportunity to clarify some of the respondents' perceptions and attitudes about the data model.

Future Directions

The potential benefits of incorporating a standard terminology into any application will primarily be associated with the retrieval process. While this application currently does not support retrieval based on terminology matching, this is an area for future work. Now, users are asked to spend additional time at the outset by identifying appropriate SNOMED codes. It is imperative that a retrieval mechanism be implemented, allowing users to retrieve concepts using synonyms and relationships, and receive a return on their initial "investment". While the application currently retrieves codes using the 1999

UMLS Metathesaurus, which only points to SNOMED 3.5 codes, the tables that associate codes with concepts, and concepts with each other (explicitly representing their relationships), are SNOMED-RT tables. Thus, the current system will be able to take advantage of the description logic and multiple explicit hierarchies used in the newest SNOMED version to improve retrieval, even before the Metathesaurus is updated with new SNOMED terms.

As mentioned earlier, interpretation of this data must be done somewhat cautiously with such a small sample size of images. More data is needed to comprehensively evaluate both the matching algorithm as well as the adequacy of the content coverage of SNOMED for this application. As more images with more terms are added to the database, this is a potential area for future study. Currently, users are asked to submit the terms that failed to match to SNOMED. Assuming compliance with this step, a potentially useful list of "trouble" terms will emerge and may allow targeted efforts to improve the matching. Similarly, as greater diversity is encountered, there may be need for additions or modifications of the data model. This can be much more meaningfully studied when more images are added.

Another area of potential study involves the comparison of a system that offers a free text approach for indexing with one that uses the potentially more restrictive model described above. Based on the results of this evaluation, a "comment" field will be developed, allowing users to write a brief clinical narrative. It will be interesting to see if this freedom leads users to abandon data entry in the more specifically labeled *diagnosis*, *finding*, and *anatomic location* fields. Additionally, it would be interesting to compare the success of the matching algorithm in these two systems. Since SAPHIRE has been used to parse textual reports in the radiology domain with reasonable success, it may offer valuable benefits for coding the information in such a free-text field.

The desire for, an implementation of, the *patient code* attribute has many interesting elements. From an ethical standpoint, more effort should be done focusing on the question of whether it represents a threat to patient confidentiality. If so, its use may well be affected by the upcoming enactment of HIPAA regulations. It also would be interesting to note how it affects image retrieval, and how often it's used for this purpose.

SUMMARY

The well-known aphorism, "A picture is worth a thousand words", is as true in medicine as it is in art or journalism. Choosing which of these "thousand words" is most important in representing image content and image-related information is a vexing challenge for those involved in the development of image databases.

Careful selection of the data model used to store information has the potential to significantly affect the success of image retrieval. Teaching image collections, which often rely on manual indexing, require a submission process which is not overly cumbersome or time-consuming. Also of concern when creating a data model is that the model must be consistent with how users conceptually organize the relevant information about an image and its acquisition.

Using a standard terminology to represent clinical concepts also can have a large impact on the accuracy of indexing and retrieval in a teaching image database. In order to be useful, the terminology must have adequate content coverage to represent the users' concepts, and utilize a matching approach that not only identifies correct codes, but is well integrated with the indexing process so as not to impose undue burden on the busy clinician-user.

A WWW-based image database was developed to facilitate the indexing, storage, and retrieval of clinical photographs within the internal medicine domain. Careful attention

was paid to the development of a data model designed to allow accurate and complete representation of relevant image-related information. A SAPHIRE based matching algorithm was developed to match user-entered clinical terms to the SNOMED 3.5 vocabulary.

Preliminary evaluations on a limited set of images indicate that users felt they were able to store the most important information about images that they added to the database. Users generally perceived the submission process as easy, and not taking "too much time". While users were satisfied with the amount of patient-specific information (patient age, gender, etc), they were interested in representing more clinical information than was generally allowed by the database, including a clinical history. While mean scores indicated that attributes such as *photographer, image resolution, file type,* etc. were less important, there was significant variation, limiting the conclusions that can be drawn.

Using SNOMED to represent the clinical concepts in this database appeared feasible. Users perceived that it was well integrated with the indexing process, and did not take "too long" to identify potential codes and allow selection of the most appropriate ones. In the 48 images indexed, users felt that the SAPHIRE-based algorithm identified codes correctly 80% of the time. Encouragingly, even without the current ability for SNOMED-based retrieval, participants felt that the potential benefits of the terminology outweighed the time necessary for the matching.

REFERENCES

1. Magner LN. A History of Medicine. New York, New York: Marcel Dekker, Inc.; 1992.
2. Neuse WH, Neumann NJ, Lehmann P, Jansen T, Plewig G. The history of photography in dermatology. Milestones from the roots to the 20th century. Arch Dermatol 1996;132(12):1492-8.
3. http://www.kbnet.co.uk/rleggat/photo/, Robert Leggat MA ME, Ph.D., FRPS, FRSA. A History of Photography.
4. Raymond SL, Algermissen VL. A retrieval system for biomedical slides using MeSH. Bull Med Libr Assoc 1976;64(2):233-5.
5. Prakash UB. Filing and retrieving a personal collection of 35-mm teaching slides in medicine. A simplified system. Archives of Internal Medicine 1985;145(9):1680-2.
6. Hindel R. Digital image storage technology. Investigative Radiology 1993;28(5):454-8.
7. O'Brien MJ, Sotnikov AV. Digital imaging in anatomic pathology. Am J Clin Pathol 1996;106(4 Suppl 1):S25-32.
8. Diepgen TL, Eysenbach G. Digital images in dermatology and the Dermatology Online Atlas on the World Wide Web. Journal of Dermatology 1998;25(12):782-7.
9. http://www.nyee.edu/diglib/cfm/index.htm, The Digital Atlas of Ophthalmology. Department of Opthalmology, The New York Eye and Ear Infirmary.
10. Balter S. Digital images. Catheter Cardiovasc Interv 1999;46(4):487-96.
11. Bidgood WD, Jr., Bray B, Brown N, Mori AR, Spackman KA, Golichowski A, et al. Image acquisition context: procedure description attributes for clinically relevant indexing and selective retrieval of biomedical images. Journal of the American Medical Informatics Association 1999;6(1):61-75.
12. Browning GC, Liang Y, Buckwalter KA, Kruger RA, Aisen A. World Wide Web interface to digital imaging and communication in medicine-capable image servers. Journal of Digital Imaging 1996;9(4):178-84.
13. Hatton C, Woods J, Dhir R, Bastacky S, Epstein J, Miller G, et al. Application of UMLS indexing systems to a WWW-based tool for indexing of digital images. Proceedings / AMIA Annual Fall Symposium 1997:420-3.
14. Zaidel M, Hopper K, Iyriboz T. Interactive web-based radiology teaching file. Journal of Digital Imaging 1999;12(2 Suppl 1):203-4.
15. Eysenbach G, Bauer J, Sager A, Bittorf A, Simon M, Diepgen T. An international dermatological image atlas on the WWW: practical use for undergraduate and continuing medical education, patient education and epidemiological research. Medinfo 1998;9(Pt 2):788-92.
16. Bidgood WD, Jr., Horii SC, Prior FW, Van Syckle DE. Understanding and using DICOM, the data interchange standard for biomedical imaging [see comments]. J Am Med Inform Assoc 1997;4(3):199-212.
17. Lowe HJ, Antipov I, Hersh W, Smith CA. Towards knowledge-based retrieval of medical images. The role of semantic indexing, image content representation and knowledge-based retrieval. Proceedings / AMIA Annual Symposium 1998:882-6.
18. Bidgood WD, Jr. Documenting the information content of images. Proc AMIA Annu Fall Symp 1997:424-8.

19. http://www.dermis.net, Diepgen TL SM. Dermatology On Line Atlas; 1999.
20. Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR. The content coverage of clinical classifications. For The Computer-Based Patient Record Institute's Work Group on Codes & Structures. Journal of the American Medical Informatics Association 1996;3(3):224-33.
21. http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html. UMLS Metathesaurus, United States National Library of Medicine.
22. UMLS Knowledge Sources Documentation. 8th ed: US Department of Health and Human Sercies, National Institutes of Health, National Library of Medicine; 1997.
23. Lowe HJ. Image Engine: an object-oriented multimedia database for storing, retrieving and sharing medical images and text. Proceedings - the Annual Symposium on Computer Applications in Medical Care 1993:839-43.
24. Chute CG, Cohn SP, Campbell JR. A framework for comprehensive health terminology systems in the United States: development guidelines, criteria for selection, and public policy implications. ANSI Healthcare Informatics Standards Board Vocabulary Working Group and the Computer-Based Patient Records Institute Working Group on Codes and Structures. J Am Med Inform Assoc 1998;5(6):503-10.
25. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med 1998;37(4-5):394-403.
26. Campbell JR, Carpenter P, Sneiderman C, Cohn S, Chute CG, Warren J. Phase II evaluation of clinical coding schemes: completeness, taxonomy, mapping, definitions, and clarity. CPRI Work Group on Codes and Structures [see comments]. Journal of the American Medical Informatics Association 1997;4(3):238-51.
27. Campbell KE, Oliver DE, Spackman KA, Shortliffe EH. Representing thoughts, words, and things in the UMLS. Journal of the American Medical Informatics Association 1998;5(5):421-31.
28. Campbell KE, Oliver DE, Shortliffe EH. The Unified Medical Language System: toward a collaborative approach for solving terminologic problems. J Am Med Inform Assoc 1998;5(1):12-6.
29. Lindberg DA, Humphreys BL, McCray AT. The Unified Medical Language System. Methods Inf Med 1993;32(4):281-91.
30. Lowe HJ AI, Hersh W, Smith CA, Mailhot M. Automated Semantic Indexing of Imaging Reports to Support Retrieval of Medical Images in the Multimedia Electronic Medical Record. Methods of Information in Medicine 1999;38:303-7.
31. Spackman K. Evaluating Clinical Terminology. Portland, OR; 1999.
32. Spackman KA, Campbell KE, Cote RA. SNOMED RT: a reference terminology for health care. Proceedings / AMIA Annual Fall Symposium 1997:640-4.
33. Spackman K. Evaluating Clinical Terminology. Portland, OR; October 1999.
34. http://www.medsight-info.com/en/Products/Snocode.htm, Medsight Informatique I. SNOCODE.
35. http://www.lxt.com/Guide.content.html, Lexical Technologies I. Metaphrase Guide.
36. K.A. Spackman M, PhD, C. Broverman, PhD, C. Masarie, MD, C.Chute, MD, L.M. Lau, MD, PhD. Report of Experiences Using the Beta Version of SNOMED-RT. In: AMIA '99 Annual Symposium; 1999; Washington, D.C.; 1999.

37. Hersh WR, Greenes RA. SAPHIRE--an information retrieval system featuring concept matching, automatic indexing, probabilistic retrieval, and hierarchical relationships. Comput Biomed Res 1990;23(5):410-25.
38. Greenes RA, McClure RC, Pattison-Gordon E, Sato L. The findings--diagnosis continuum: implications for image descriptions and clinical databases. Proc Annu Symp Comput Appl Med Care 1992:383-7.
39. Tague-Sutcliffe JM. Some Perspectives on the Evaluation of Information Retrieval Systems. Journal of the American Society for Information Science 1996;47(1):1-3.

Appendix I: PERL Script for Terminology Matching to SNOMED Using SAPHIRE

```
sub Coder
## This subroutine takes a phrase and returns an array of potential
##matching
## SNOMED codes and terms. It builds an array of CUI's first, then
##uses these
## to map to SNOMED codes using the MRSO table of the UMLS.  Finally
##it looks up
## these codes in a hash table to identify the SNOMED concept term
##that goes
## with the code.

{
$saph = "/mnt/d4/Saphire4.3/saphire4.3 -d /mnt/d4/Saphire4.3";
$line = $_[0];
$line =~s/[',-]/ /;
$CUI_limit = 10;
$SNM_limit = 7;
@snm=();
@snmMatch=();
@CUI=();
@SNMI98results=();
@SNM2results=();
@Grepresults=();
## First, try to get match right out of SNMI98
## We'll fill the results array with a bunch of CUI's that we know
##will
##    map successfully to SNOMED

@SNMI98results = `$saph -B -n5 -v SNMI98 $line | cut -f1 -d "|"`;
@SNM2results  = `$saph -B -n5 -v SNM2 $line | cut -f1 -d "|"`;
@Grepresults  = `$saph -B -n5 -K $line | grep SNM | cut -f1 -d"|"`;
@UMLSresults  = `$saph -B -n5 $line | cut -f1 -d"|"`;
for ($i=0; $i<@SNMI98results; $i++)
        {
        chomp $SNMI98results[$i];
        if ($SNMI98results[$i] =~ /^\s/) {$SNMI98results[$i]
=~s/^\s//;}
      if (($SNMI98results[$i]!~/!|No matches found|-/) &&
($SNMI98results[$i] ne ""))
            {
            push (@CUI,$SNMI98results[$i]);
              }
      }
```

```
## Pass 2: Now, if we don't have enough CUI's, we'll go after SNM2

$size = @CUI;
if (@CUI < $CUI_limit)
	{
	$i=0;
	while ((@CUI < $CUI_limit) && ($i<@SNM2results))
		{
		chomp $SNM2results[$i];
		if ($SNM2results[$i] =~ /^\s/) {$SNM2results[$i]
=~s/^\s//;}
		if (($SNM2results[$i]!~/!|No matches found|-/) &&
($SNM2results[$i] ne "") && (!CUI_in_Array($SNM2results[$i])))
			{
			push (@CUI,$SNM2results[$i]);
			}
		$i++;
		}
	}
## Pass Three - display the source codes and grep out those CUI's
with SNM

if (@CUI < $CUI_limit)
	{
	$i=0;
	while ((@CUI < $CUI_limit) && ($i<@Grepresults))
		{
		chomp $Grepresults[$i];
		if ($Grepresults[$i] =~ /^\s/) {$Grepresults[$i]
=~s/^\s//;}
		if (($Grepresults[$i]!~/!|No matches found|-/) &&
($Grepresults[$i] ne "") && (!CUI_in_Array($Grepresults[$i])))
			{
			push (@CUI,$Grepresults[$i]);
			}
		$i++;
		}
	}
if (@CUI < $CUI_limit)
	{
	$i=0;
	while ((@CUI < $CUI_limit) && ($i<@UMLSresults))
		{
		chomp $UMLSresults[$i];
		if ($UMLSresults[$i] =~ /^\s/) {$UMLSresults[$i]
=~s/^\s//;}
		if (($UMLSresults[$i]!~/!|No matches found|-/) &&
($UMLSresults[$i] ne "") && (!CUI_in_Array($UMLSresults[$i])))
			{
			push (@CUI,$UMLSresults[$i]);
			}
		$i++;
		}
	}
```

```
sub CUI_in_Array
	{
	$j=0;
	$inThere = "";
	while (($j<@CUI) && ($inThere eq ""))
		{
		chomp $CUI[$j];
		if ($CUI[$j] eq $_[0]) {$inThere = "true";}
		$j++;
		}
	return $inThere;
	}
$i=0;

#Set up oracle environment variables
$ENV{"ORACLE_HOME"}="/pub6/app/oracle/product/7.3.2";
$ENV{"ORACLE_SID"}="meta";

my $dbh = DBI-
>connect('dbi:Oracle:','lymanja','hoodoo',{AutoCommit=>0})
		or die "DBI::err:$DBI::errstr\n";
while (($i<@CUI) && (@snm < $SNM_limit))
	{
	@qry_ary=();

	#Open process and select data
	my ($str);
	$str = "select SNM from CUI_to_SNM
		where CUI = '$CUI[$i]'";

	my $sth = $dbh->prepare($str)
		or print "$DBI::err: $DBI::errstr";
	my($filename);

	$sth->execute
		or print "$DBI::err: $DBI::errstr";

	my ($qry_ary);
	while (@qry_ary = $sth->fetchrow_array)
	{
	$limit = @qry_ary;
	if (@qry_ary)
		{
		for ($k=0;$k<@qry_ary;$k++) {
			push(@snm, $qry_ary[$k]);
			}
		$termLimit = 4;
		}
	}
	$i++;
$sth->finish
	  or die "$DBI::err: $DBI::errstr\n";
	}
```

```
chomp @snm;
$size = @snm;
$dbh->disconnect
        or die "$DBI::err: $DBI::errstr\n";

#print @snm;
#print "\n";
@snmTerm =();
for ($i=0; $i<@snm; $i++)
        {
        chomp $snm[$i];
        if (($snm[$i] ne "|") && ($snm[$i] ne ":"))
            {
            $term = "";
            push (@snmTerm, $snm[$i]);
            push (@snmTerm, "|");
            $term = &getTerm($snm[$i]);
            if ($term eq "")
                  {
                  $term = `cat ../srt_en_us_terms.dat | grep $snm[$i]
| cut -f3 -d"|"`;
                  push (@snmTerm, $term);
                  }
            else
                  {
                  push (@snmTerm, $term);
                  }
            }
        else
            {
                push (@snmTerm, $snm[$i]);
            push (@snmTerm, "not found!");
            }
      push (@snmTerm, "|");
      }
chomp @snmTerm;

sub getTerm {
dbmopen (%concepts, "../myCon",0666) || die "Can't DBMopen myCon:
$!";
$line=$_[0];
$S_Term="";
$S_Term = "$concepts{$line}";
dbmclose(%concepts) || die "can't dmbclose myCon: $!";
return $S_Term;
}
return @snmTerm;
}
sub tryAltered {
        $orig_term = $_[0];
        $orig_term =~s/s\Z//;
        $orig__term =~s/s\s/ /g;
        @CodeResult = Coder($orig_term);
        return @CodeResult;
        }
```

Appendix II. Paper Survey and Instructions

The Physical Examiner Survey
April 2000

Directions:

You should have received a set of 5-10 images (on a floppy disk or via e-mail). In order to evaluate the submission process of the database application, I'd like you to submit these images one at a time, filling out all the indexing information you feel is appropriate.

On the next page are scenarios for each image. Please use this information as you go through the indexing process. If you recognize the image and want to use real information, please feel free. I'm not interested in your diagnostic accuracy! What's more important is that you try to identify what you think is important about each image and try to store that information in the appropriate field.

1. Go to the web site listed in the top right corner of this page. Log in with the ID, password shown. Hit the "submit an image" button.
2. In real life, you'll be sitting down to submit the image with a consent form that has the patient name, consent given, etc. In order to link the image with the consent information, the database uses consent numbers. To submit an image from a new patient, first "request a new consent number". This will retrieve the next consent number from the database. ***Please write this number down on the next page!*** In real life, you can write this info on the actual consent form; this will allow you to later find the consent information for a given image, and also save you some typing for the patient's additional images.
3. If you're adding an additional image for a patient who's already be entered, enter in the consent number that you were given for that patient.
4. After entering the new patient information, or confirming the existing information continue on to the next page. Here you'll browse to find the test images I've given you. Upload the image.
5. After confirming the image, enter the information in the fields shown. ***Please fill in your subject number (above in the left hand corner) in the photographer field*** and hit the 'submit' button when complete.
6. The next page will try to find the SNOMED codes for your clinical terms. Click all appropriate questions, and answer the questions at the bottom about how well your terms were matched to SNOMED.
7. Again, hit the "submit" button, confirm your data on the next page, and then hit the "submit" button once more. This will add your data to the database, and prompt you to answer some questions about the submission process. Please answer these, and then click on the link to return to the home page.
8. Repeat the process until all images are submitted.
9. The last step is to fill in the paper survey that I've included. If you'd like to experiment with retrieving the images as well, feel free!

YOUR IMAGES

abdMass.jpg: This is a 65 y.o. male named Ralph Tiller, seen November 1998 with an abdominal mass. He also has CHF. CONSENT NO:________

hirsutism.jpg: A 53 y.o. female named Georgia Reed, seen here in March 2000.
CONSENT NO:________

palmar_erythema1.jpg: A 39 y.o. female named Chris Thomas, in August 1999.
CONSENT NO:________

psoriasis2.jpg: A 47 y.o. male named Roger Conway, seen here January 14, 2000.
CONSENT NO:________

psoriasis5Kaynard.jpg: The same patient as above.

skinlesion-smallpox.jpg: A 34 y.o. man named Tom Greps in June of 1997.
CONSENT NO:________

General Questions	Strongly disagree				Strongly agree
1. Overall, the submission process was easy.	1	2	3	4	5
2. In general, submitting an image takes too much time.	1	2	3	4	5
3. Overall, I was able to store the information that I felt was most important about each image.	1	2	3	4	5
4. In general, there were too many fields to fill out.	1	2	3	4	5
5. Putting the clinical concepts in the separate diagnosis, findings, and anatomic location fields was too restrictive and inflexible.	1	2	3	4	5
6. The SNOMED matching process was well integrated with the rest of the submission process.	1	2	3	4	5
7. The SNOMED matching process took up too much time.	1	2	3	4	5
8. Overall, the application took too long to generate SNOMED codes.	1	2	3	4	5
9. The extra time required to match to the SNOMED codes is worth the potential benefits.	1	2	3	4	5
	Too much		**Just right**		**Too Little**
10. In general, the amount of patient information required was:	1	2	3	4	5
11. In general, the amount of clinical information required was:	1	2	3	4	5
12. In general, the amount of file information (resolution, date, photographer, etc.) required was:	1	2	3	4	5

13. **For each field in the indexing process, please indicate how important you think it is to store with the image.**

Field	Very unimport-ant				Very important
Diagnosis 1	1	2	3	4	5
Diagnosis 2	1	2	3	4	5
Diagnosis 3	1	2	3	4	5
Finding 1	1	2	3	4	5
Finding 2	1	2	3	4	5
Finding 3	1	2	3	4	5
Diagnostic Certainty	1	2	3	4	5
Anatomic Location	1	2	3	4	5
Patient Code	1	2	3	4	5
Patient Age	1	2	3	4	5
Patient Gender	1	2	3	4	5
Consent Number	1	2	3	4	5
Image Resolution	1	2	3	4	5
Photographer	1	2	3	4	5
Date Image Taken	1	2	3	4	5
Image Quality for Teaching	1	2	3	4	5

File Type	1	2	3	4	5
Department	1	2	3	4	5

14. Are there fields that you think are unnecessary and would like to see removed? If so, what fields?

15. Are there additional fields that you would like to see added? If so, what fields?

16. What, in your opinion, are the benefits of using SNOMED to encode the clinical concepts?

17. Do you have any additional comments, suggestions, etc?

Thank You for Your Time!