Knowledge Representation and Sharing Using Visual Semantic Modeling for Diagnostic Medical Image Databases

Adrian S. Barb, *Student Member, IEEE*, Chi-Ren Shyu, *Member, IEEE*, and Yash P. Sethi

***Abstract*—Information technology offers great opportunities for supporting radiologists' expertise in decision support and training. However, this task is challenging due to difficulties in articulating and modeling visual patterns of abnormalities in a computational way. To address these issues, well established approaches to content management and image retrieval have been studied and applied to assist physicians in diagnoses. Unfortunately, most of the studies lack the flexibility of sharing both explicit and tacit knowledge involved in the decision making process, while adapting to each individual's opinion. In this paper, we propose a knowledge repository and exchange framework for diagnostic image databases called "evolutionary system for semantic exchange of information in collaborative environments" (Essence). This framework uses semantic methods to describe visual abnormalities, and offers a solution for tacit knowledge elicitation and exchange in the medical domain. Also, our approach provides a computational and visual mechanism for associating synonymous semantics of visual abnormalities. We conducted several experiments to demonstrate the system's capability of matching synonym terms, and the benefit of using tacit knowledge in improving the meaningfulness of semantic queries.**

***Index Terms*—Content-based image retrieval, knowledge exchange, knowledge representation, medical image database, radiology, semantic query.**

I. Introduction

In the medical domain, knowledge exchange is difficult, especially due to the autonomy of care providers and to the importance of its tacit component [33]. Domain experts, who usually carry this knowledge, have close concordance with their local environment, in which both previous experience and colleagues' opinions have a major influence. However, local knowledge is often limited and insufficient to deal with tough cases that have not been previously diagnosed [23]. The tradeoff between knowledge value and elicitation effort becomes very important since physicians have limited amount of time to respond to a case and/or share expertise with peers. Looking for knowledge beyond the local setting is necessary but difficult due to the differences in group cultures and to locally defined methods of encoding information into semantics.

Manuscript received April 17, 2004; revised December 10, 2004.

A. S. Barb is with the Computer Science Department, University of Missouri, Columbia MO 65211 USA (e-mail: Adrian@missouri.edu).

C.-R. Shyu is with the Computer Science Department and Health Management and Informatics Department, University of Missouri, Columbia MO 65211 USA (e-mail: ShyuC@missouri.edu).

Y. P. Sethi is with the Radiology Department, University of Missouri Hospital, Columbia MO 65211 USA (e-mail: SethiY@health.missouri.edu).

Digital Object Identifier 10.1109/TITB.2005.855563

Several systems that focused on knowledge exchange were developed [18], [19], [26]. Fox and Thomson [18] proposed a unified technology for clinical decision support and disease management that emphasizes integrated methodologies for developing clinical applications. Gardner *et al.* [19] designed a framework, using XML-derived schemas, which defines an interoperability standard for neuroscience informatics resources. The knowledge exchange framework developed by Kindberg *et al.* [26] addressed the issue of communicating through peer-to-peer networks, as well as methods of facilitating data and sharing knowledge. While it is true that knowledge-base systems cannot perform better than human experts [5], they are capable of filtering the information to be presented to experts for diagnoses. Economou *et al.* [17] proposed a computer aided medical system that allows a human-in-the-loop, step-by-step procedure for approximating the final diagnosis in different fields of medicine. In the radiology community, knowledge sharing is more difficult than other medical domains since it is very difficult to accurately describe visual patterns using plain text annotations. Therefore, instead of plain text, systems need a common base to share and exchange knowledge related to visual content of the abnormality present in diagnostic medical images. That is, no matter what annotations are associated with the images, if two medical images share similar visual abnormalities identified by physicians, they should also share similar visual contents detected by computer algorithms.

In the past decade, researchers have been developing several prominent content-based image retrieval (CBIR) systems for medicine [12], [13], [25], [27], [37], [47]. These CBIR systems mimic the domain knowledge to extract image contents and provide query methods for direct image (visual pattern) match using low level image features. The prototype by Cai *et al.* [12] retrieved positron emission tomography images based of their specific physiological kinetic features, and developed a methodology of image compression that supports fast content-based image retrieval. Chu *et al.* [13] developed a semantic model for content-based image retrieval for capturing the hierarchical, spatial, temporal, and evolutionary semantics of neural images in image databases. The system by Kelly *et al.* [25] associated each medical image a signature for capturing textures and histograms of pathologies, and retrieves images using query-by-example techniques. Fast query results for nearest neighbor search is addressed by Korn *et al.* [27] who used multidimensional indexing of medical tumors with similar shapes using an R-tree. The system proposed by Nah and Sheu [37] used operational semantics to ensure the meaningfulness of content-based retrieval of neuroscience images. In the ASSERT system [47], Shyu *et al.* designed a suite of computer vision algorithms to extract visual abnormalities and used multidimensional hashing approach to index pathologies of lung HRCT images. Robinson *et al.* [43] indexed shapes of cardiac boundary curves using a KD-tree.

The ultimate goal of these medical CBIR systems is to assist physicians' diagnoses. However, most of them used standalone knowledge and did little to encourage knowledge elicitation and exchange among groups of physicians. Knowledge, described as information with a productive component, is a very important component of the value generation process [52] in any organization. Tacit knowledge is an important part of human reasoning that evolves through human interactions with the surrounding environment. It is described by Reeves *et al.* in [42] as "the glue, texture, and backdrop for our interaction with people, places and things." It can help reach conclusions when explicit knowledge fails to capture full explanations of a phenomenon, but is very difficult to share, due to the human tendency to protect information that can give a competitive advantage over the other members of the organization.

Peer-to-peer networks have proven to be successful in tacit knowledge elicitation and exchange by facilitating alternative opinions and revisions [42]. Knowledge exchange through these networks has a very high creation potential due to its capabilities to create strong temporary connections. These connections are derived from existing weak ones to maximize knowledge generation through connecting different groups of users in an environment. It is rare to see a medical CBIR system that encourages physicians to define their own semantics to the database, as well as to adapt individual preferences of semantics to the common knowledge base in the medical domain. This aspect becomes very important in the medical domain because medical concepts, with their empirical characteristics, are subject to a continuous semantic and conceptual adaptation [11]. In practice, physicians use several perceptual categories [47] to make diagnoses. A major drawback of a system that tries to mimic this reasoning process is the subjective assignment of the mapping between semantic terms and image features. If there is a significant discrepancy between the similarity as assigned by the system and the notion of similarity in the physician's mind, the results are destined to be unsatisfactory.

Domain ontology can be used as a common framework for knowledge representation and exchange because it can connect patient information to concepts stored in the knowledge base. Leroy and Chen [29] developed a tool (Medical Concept Mapper) for facilitating access to online medical information that uses human-created ontologies such as unified medical language system (UMLS) [54] and WordNet [34] to improve the document retrieval performance. But the use of ontology requires consensus on ontological definitions among the community members to reduce the ambiguities in communication. However, such consensus may limit the individual user's possibility to view the knowledge according to his or her specific expertise. For this reason, experts should be able to customize their individual semantic terms in order to create a physician friendly environment for decision support.

Visual semantics used by the physicians for diagnoses are not always binary: existing or nonexisting. In practice, there is no hard boundary that separates two visually similar semantics, such as many and few nodular opacities. If a crisp threshold of a low level feature is set to distinguish two semantics, the threshold is always subjective and may not calibrate what is in the physician's mind [16]. Fuzzy logic could be a good tool to handle this subjectivity of semantic assignments. Approaches in domain of general image retrieval, such as [1], [28], [36] and [45], try to implement fuzzy logic concepts to increase the meaningfulness of the retrieval results. Aguilera *et al.* [1] developed a model for fuzzy image retrieval by expressing image features and user queries in terms of fuzzy sets. Madasani *et al.* [30] represented image regions and queries as fuzzy attributed relational graphs and use an efficient fuzzy algorithm for matching them. Mouaddib and Bonnano [36] developed a fuzzy relational schema that assigns to each tuple a degree of compatibility with the fuzzy constraints defined on the relation. Saint-Paul *et al.* [45] applied fuzzy semantic hierarchies and relationships among terms. Techniques proposed by these researchers will be valuable for modeling the knowledge environment to be able to integrate with physician's individual preference.

In this paper, we present Essence—a framework for knowledge representation and sharing for the radiology community. In Essence, we develop a shared ontology based on the common knowledge from expert radiologists, and information from two well-known references [51], [56]. This framework extracts and manages visual content of lung pathologies. Physicians can build their personalized semantic search criteria by customizing the degrees of satisfaction of features to existing semantic terms, and by adding new semantic terms to existing perceptual categories. The system is also capable of refining the shared ontology by adapting the assignment of semantic terms to image features based on individuals' preferences. We have chosen XML for information storage and exchange due to its flexibility and extensibility [31].

This paper is organized as follows. Section II introduces the architecture of knowledge representation in Essence. Section III presents procedures of mapping low-level image features to high-level semantic terms using fuzzy logic techniques. Section IV demonstrates the retrieval system using query by semantics methods. Section V describes procedures of semantics integration to identify synonymous semantic terms. Section VI shows how the users customize their semantic settings and exchange information. We then present experimental results with usability tests in Section VII, and conclude this paper in Section VIII.

II. Knowledge Representation

Most of the decisions in the medical domain are made by comparing the data in hand against existing domain knowledge. During the decision-making process, physicians base their diagnoses on a set of heuristics developed from different areas as a "multi-dimensional intuition" [41] in which tacit knowledge plays a very important role [10]. Web-based knowledge

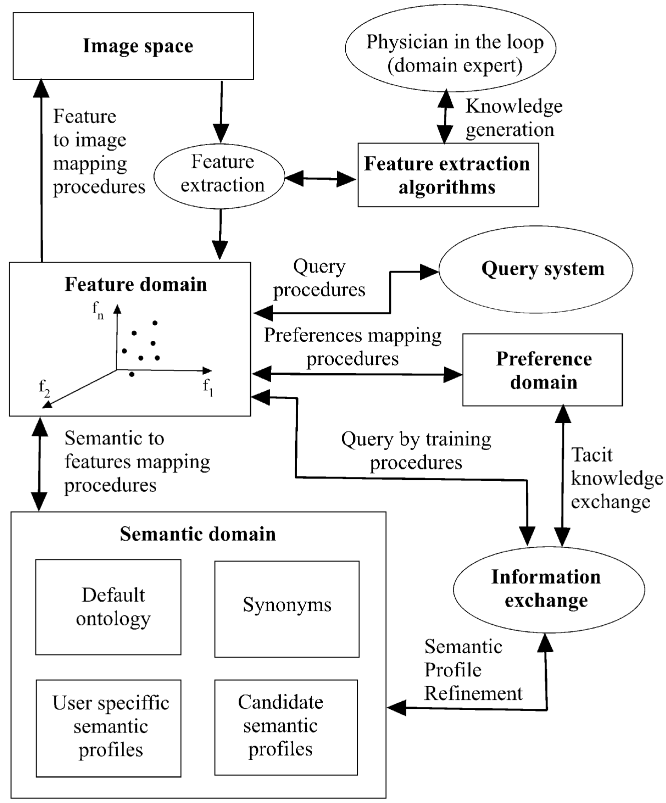

Fig. 1. Architecture of the knowledge base framework in Essence.

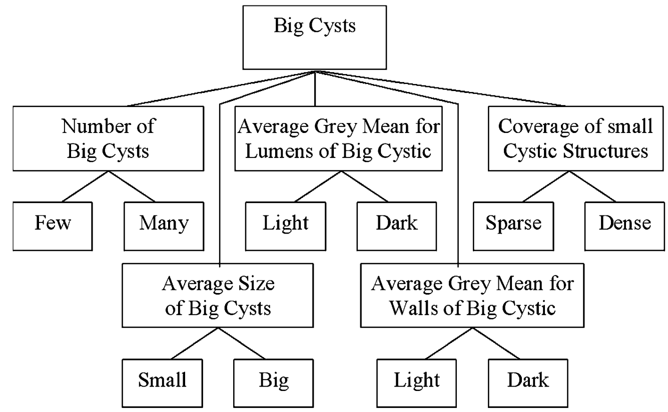

Fig. 2. Hierarchical structure of linguistic variables.

management systems have the unique feature of going beyond the typical boundaries of groups of experts [39]. They have to deal with different settings for users and with complicated information exchange procedures [28]. Such systems should effectively support most of the strategies used by physicians so the decision process is not constrained [41] by system limitations. Brown *et al.* described in [9] a knowledge-based approach to HRCT image segmentation by using anatomical structure and various domain specific knowledge. The knowledge base developed by Tayar [53] focused on data consistency and incremental development by dividing the knowledge base into layers. The model developed by Wei *et al.* [57] focuses on representing the complex heuristics and data intensive knowledge specific to the medical domain that facilitates interactions among heterogeneous and autonomous medical data sources. All these approaches bring novel ideas in knowledge management, but do little to address means of customizing the settings to the users' preferences.

The goal of our knowledge base is to provide a visual and cooperative environment that facilitates the knowledge exchange in more intuitive ways. Our framework uses XML to categorize information in a semistructured format, and also offers several methods of fast and accurate access [4]. As depicted in Fig. 1, the main components in Essence are: 1) *Semantic domain*; 2) *Images space*; 3) *Feature extraction algorithms*; 4) *Feature domain*; 5) *Preference domain*; 6) *Query system*; and 7) *Information exchange module*. Knowledge components are represented in rectangles, and knowledge-driven actions, such as search and discovery, are represented in oval shapes.

The *Semantic domain* is organized as a local-as-view data integration subsystem [15]. This system let users build, refine, and further decompose their semantics independently, with minimum effort, on the top of the shared ontology [3], [20]. The shared ontology is to be exploited by all users who will access the system. Along with the *Preference domain*, the *Semantic domain* represents the expert's knowledge in an XML format. Using a similar format, the framework represents the knowledge of a specific case, a medical image, in *Feature domain*. Each element in the *Feature domain* is a signature of a medical image in the *Image space*. The signature is computed by executing the *Feature extraction algorithms* designed by computer vision and image processing researchers. The *Query system* searches the knowledge base, selects relevant images, and translates the result into a human-readable format. It provides two mechanisms to access the knowledge: 1) query by semantics and 2) synchronization of semantic terms [2]. The *Information exchange module* facilitates knowledge exchange among users through peer-to-peer and centralized channels.

A. Semantic Domain

Physicians use several perceptual categories for recognizing pathologies in high-resolution computed tomography (HRCT) images of the lung. We define linguistic variables to model those perceptual categories used by physicians. Each of these linguistic variables is assigned a set of semantic terms that represents a semantic assignment for lung pathology. The linguistic variables and their semantic terms are arranged in a hierarchical structure, as depicted in Fig. 2. For example, in Fig. 2, the linguistic variable *Number of Big Cysts* has been assigned a semantic term set $\{$*Few, Many*$\}$. The union of all hierarchical structures of linguistic variables constitutes a semantic profile that is used to query the image space.

1) Linguistic Variables: The linguistic variables, as defined by physicians, are tuples l in the form $\langle u, c, c_1, d\rangle$ where u is the user that defined the linguistic variable; c is an indexing code; c_1 is the indexing code of the parent linguistic variable if any; and d is a description of the linguistic variable. For example, the instance of the knowledge base in Table I shows some linguistic variables defined by user *adrian*. The linguistic variable *cysb*

TABLE I
INSTANCE OF KNOWLEDGE BASE FOR STORING LINGUISTIC VARIABLES

User	Code	Parent code	Description
adrian	lngs		Lungs pathologies
adrian	cys	lngs	Cysts
adrian	cysb	cys	Big cysts
adrian	cysbn	cysb	Number of big cysts
adrian	cysbs	cysb	Average size of big cysts

TABLE II
INSTANCE OF THE KNOWLEDGE BASE FOR STORING SEMANTIC TERMS

Code	Description	Type	Parent variable	Semantic assignment function	Scope
cysbnf	Few big cysts	Right bounded	Number of big cysts	$<\lambda^{R1}=4, \lambda^{R2}=4, \lambda^{R3}=2.1>$	public
cysbnm	Many big cysts	Left bounded	Number of big cysts	$<\lambda^{L1}=12, \lambda^{L2}=5, \lambda^{L3}=1.8>$	private
cysbsav	Big cysts	Right bounded	Size of big cysts	$<\lambda^{R1}=11, \lambda^{R2}=3, \lambda^{R3}=0.9>$	private
cysbsv	Very big cysts	Left bounded	Size of big cysts	$<\lambda^{L1}=17, \lambda^{L2}=3.5, \lambda^{L3}=2.1>$	public

is described by the tuple $\langle$*adrian, cysb, cys, Big cysts*$\rangle$, and has the meaning: "User *adrian* describes the *cysb* linguistic variable as *Big cysts*." This linguistic variable is defined in the semantic tree as a descendant of the linguistic variable *cys* (*cysts*).

2) Semantic Terms: The semantic terms associated to a linguistic variable are defined as tuples s with the form $\langle c, l, d, a, o, t \rangle$ where c is an indexing code; l is the linguistic variable to which the semantic term is attached; d is a description of the semantic term; a is the description of a function that defines the semantic assignment for lung pathologies; $o \in \{$*private, protected, public*$\}$ is the scope of the semantic term and shows to what extent other users have access; and t is the type of the semantic term which will be discussed in Section III. The assignment of lung pathologies is done by specifying a matching degree to all the linguistic variable measurements in relation to the semantic term used.

When adding new semantic terms, our system follows the principles for designing ontology: 1) parsimony—semantic terms are added only if strictly necessary; 2) clarity—semantic terms should effectively communicate the intended meaning; and 3) coherence—all new terms should be locally consistent. Also, each semantic term should be mapped to a normalized possibility distribution (PD). That is, there should exist at least one image that fully matches the semantic term [55].

For example, the first row of Table II lists a semantic term with the following attributes: *cysbnf*—indexing code, *Few big cysts*—description, and *Right bounded primitive*—type. This term is a child node of the linguistic variable *Number of big cysts* (indexing code *cysbn*). The PD is described by a series of coefficients λ, which will be explained in the next section. The fifth column indicates that this semantic term is available for all users.

3) Semantic Profiles: To adapt itself to users' preferences, our system creates four semantic profile types: *default, candidate, user-specific*, and *working*. The first two are designed for all users; the last two for an individual user. These profiles are stored in the knowledge base in XML format. In this semantic profile, each nonleaf node holds a linguistic variable as described in Section II-A1, while a leaf-node holds a semantic term as described in Section II-A2.

For a new user or knowledge depositor, the *user-specific* profile is initially empty and the user inherits the parameters from the *default* one which contains all the linguistic variables and semantic terms commonly agreed by the existing users. Also, the new users have access to all the other semantic terms from existing users by using the *candidate* profile, which is updated only when new linguistic variables or semantic terms are added to the system. When a user customizes his or her personal settings, the new parameters are saved in the *user-specific* profile. To retrieve database images by semantics, a *working semantic* profile is created on the fly. This profile inherits all the linguistic variables and semantic terms from the *default* profile and appends all new variables and terms from the *candidate* profile. However, settings in the *user-specific* profile are mandatory to overwrite those in both *default* and *candidate* profiles. Fig. 3 shows the process of building a *working* profile for user *adrian*. The *working* profile, Fig. 3(d), inherits the *default* profile with double-circle nodes and appends the *candidate* profile with thin-circle nodes. The bold single-circle nodes are from the *user-specific* profile.

B. Image Space and Feature Extraction Algorithms

The raw information processed by our system is a collection of HRCT images of lung. To extract high level semantics from images, a suite of computer vision and image processing algorithms are designed to identify visual abnormalities of lung pathologies. To have a concise presentation of the main theme of this paper in knowledge sharing and semantic modeling, we only briefly discuss the algorithms that were designed to extract two perceptual categories (out of 24): small *nodular opacities* and *cystic structures*.

1) Algorithms to Extract Nodular Opacities: An example lung disease resulting in nodular opacities on HRCT images is *sarcoid* [52]. Important features to describe nodule opacities include: 1) the gray values associated with nodules since the values carry important information with regard to whether the tissue is benign or malignant; 2) the size and spatial distributions associated with the nodular opacities; and 3) the roundness of high grey-level objects.

To extract image features related to nodular perceptual category, we have implemented the following procedure.

a) Extract the lung regions [46] and apply Otsu thresholding [40] on them.
b) Apply labeling to high pixels.
c) Compute the roundness of labeled objects by

$$\text{roundness} = \frac{4 * \text{Area}}{\pi * \text{Diameter}^2}.$$

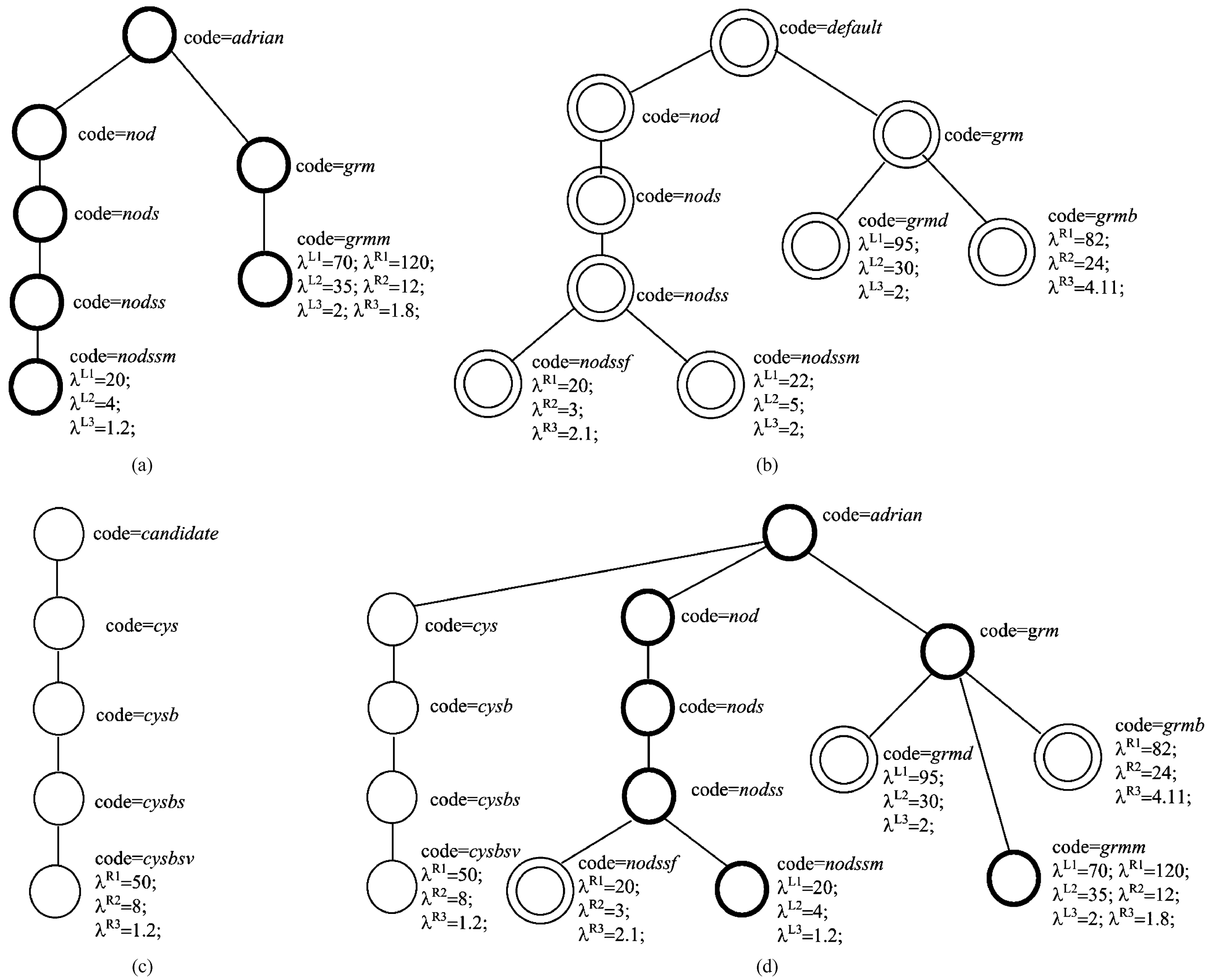

Fig. 3. Example of semantic profiles. (d) The *working semantic* profile is the result of combining (a) the *user-specific* profile, (b) the *default profile*, and (c) the *candidate* profile.

d) Group labeled objects into small nodules and big nodules based on two measurements: roundnesses and sizes of labeled objects. Both thresholds were learned from the training data.

Effective feature measurements for images with this type of pathology include: 1) number of small nodules; 2) roundness mean of small nodules; 3) grey mean of small nodules; 4) average nearest neighboring distances (NNDs); 5) standard deviation of NNDs; and 6) histogram of NNDs partitioned into six bins. In this paper, we show the semantic term derived from the first feature. Other semantic terms, such as "uniformly distributed small nodules" and "skew distributed small nodules" are modeled by using features 4–6.

2) Algorithms to Extract Cystic Structures: To identify the presence and absence of cystic structures, we have applied the following procedure.

a) Extract the lung regions and apply a dual-thresholding on the regions to highlight potential cyst walls from high pixels and possible lumens from low pixels.
b) Set all pixels outside the lung regions to zero. These outside pixels and pixels of the lumens have the same gray value.
c) Apply watershed algorithm [44] to repair the broken cystic walls.
d) Apply component labeling to high pixels.
e) Compute the sizes of the labeled objects and the average grey-scale mean difference between labeled objects (potential walls) and pixels bounded by the labeled objects (possible lumens).
f) Prune out those labeled objects that fall at least one standard deviation away from the means of the sizes and grey mean differences of training cystic structures.

Effective attribute measurements for images with this type of pathology include: 1) number of cystic cells; 2) average size of cells; and 3) coverage of cystic structures within the lung regions.

A more comprehensive discussion for all perceptual categories used in Essence can be found in [46]. Each perceptual

TABLE III
INSTANCE OF THE KNOWLEDGE BASE FOR STORING IMAGE FEATURES

Image	Linguistic variable	Measurement
Essence-01	Number of small nodules	25
Essence-01	Average size of small nodules	1.55
Essence-02	Number of small nodules	44
Essence-02	Average size of small nodules	0.77

category studied in this paper has a set of relevant image features which were tested by multivariate analysis of variance (MANOVA) [52] and empirically proven to be efficient [48] to distinguish categories from each other. A multidimensional feature vector is then formed for each raw image. Whenever a new linguistic variable is defined, Essence either reuses the existing algorithms or asks the computer vision/image processing researchers to develop a new feature extraction algorithm that is dedicated to this new variable.

C. Feature Domain

For each new image in the database, feature extraction algorithms [50] are applied and an image feature profile is created. This profile has a hierarchical structure similar to the combined structure of the *default* and *candidate* profiles, which were discussed in Section II-A3, except that the semantic terms are replaced by feature measurements. The knowledge base describes image features as tuples f with the form $\langle i, l, m \rangle$ where i is image, l is the linguistic variable, and m is the measurement assignment for lung pathologies. The example in Table III shows an instance of a knowledge base that stores information of image features. In this example, the value for the linguistic variable *Number of small nodules* of image id *Essence-01* is 25.

D. Preferences Domain

The uniqueness of our system comes from its self-adaptive functions that utilize relevant information provided by the users. These functions are used to provide query statistics and to update the *user-specific* and *default* profiles. Both profile updates utilize physicians' feedback, which involve a rating process to evaluate the relevance of retrieved images for certain semantic terms used in the queries.

The information stored from the rating process, which will be discussed in detail later, includes: 1) user's rating preference k in the form $\langle i, s, r \rangle$ where i is the image id, s is the semantic term that was evaluated during the rating process, and r is the rating on a scale from 0 to 10 and 2) user's query preference q in the form $\langle i, S, z_1, z_2 \rangle$ where i is the image that was retrieved by the query result, S is the set of semantic terms used when querying, z_1 is the number of times the image was retrieved in query results, and z_2 is the number of times users have selected the image as a valid result. Table IV shows an instance of a user's rating preferences for the semantic term *Many small nodules*, and Table V shows an instance of the corresponding query preferences.

TABLE IV
INSTANCE OF THE KNOWLEDGE BASE FOR STORING CUSTOMIZING OPTIONS

Image id	Semantic term	Rating
Essence-01	Many Small Nodules	8
Essence-02	Many Small Nodules	2
Essence-03	Many Small Nodules	10
Essence-04	Many Small Nodules	6

TABLE V
INSTANCE OF THE KNOWLEDGE BASE FOR STORING QUERYING OPTIONS

Image id	Semantic term	z_1	z_2
Essence-001	Many Small Nodules	6	2
Essence-002	Many Small Nodules	12	0
Essence-003	Many Small Nodules	10	8
Essence-004	Many Small Nodules	2	1

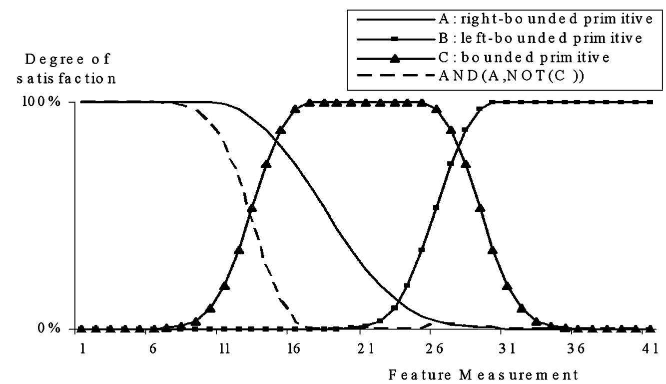

Fig. 4. Example fuzzy set for different types of semantic terms.

III. MAPPING IMAGE FEATURES TO SEMANTIC TERMS

In this section we discuss our approach of mapping the image features to semantic terms. The mapping process uses three types of information, which are: 1) semantic information; 2) image feature information; and 3) user preferences. The possibility distribution that maps semantic terms to image features is expected to capture a user's preferences in a computational way. Mitiam *et al.* [35] analyzed different types of shapes in fuzzy set theory by testing how these shapes can approximate different testing functions. Although the best shape is subjective and data/application dependent, this research concludes that there are set functions that could approximate better than the triangular or trapezoid ones.

For the purpose of our model, we extended Mitiam's research by adding an asymmetric property to the PDs of semantic terms for perceptual categories. This property is believed to be better in fitting user's semantic preference than commonly used symmetric functions. There are three parameters that control the shape of the possibility distribution, which are: 1) the center of the function (λ^1); 2) the width factor (λ^2); and 3) the exponential factor (λ^3). For example, in Fig. 4, the sigmoid part of the possibility function noted A has the parameters $\lambda^1 = 10, \lambda^2 = 8$, and $\lambda^3 = 2$.

Each PD is used to model a semantic term for a perceptual category, which is presented by a linguistic variable. Let L be the set of linguistic variables assigned to a database

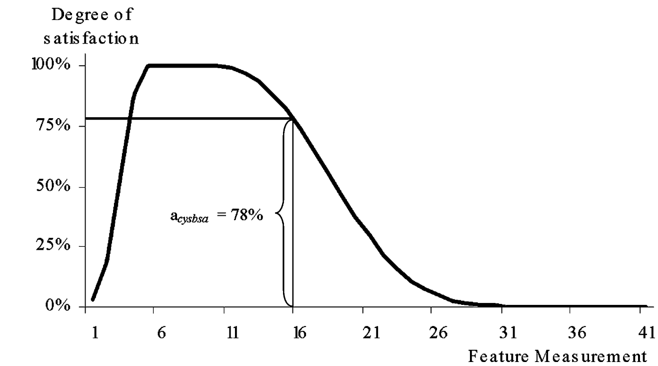

Fig. 5. Computation of the degree of satisfaction.

image, $s = \langle c_s, l, d, a_s(m), o_s, t_s\rangle$ be a primitive semantic term defined by a user for the linguistic variable $l \in L$, and NOT $(s) = \langle c_{\rm sn}, l, d, 1 - a_s(m), o_s, t_s\rangle$ be the Boolean function that is true when semantic term s is absent from a query. The semantic term s associates to the linguistic variable l a PD function $a_s : U^l \rightarrow [0, 1]$ defined by a user over the universe U^l of the linguistic variable l. For example, we can define the semantic term *Average number of big cysts* (indexing code *cysbna*) for the linguistic variable *Number of Big Cysts* (indexing code *cysbn*) $l = \langle adrian, cysbn, cysb, number of\ Big Cysts\rangle$ as $s = \langle$*cysbna*, l, *Average number of big cysts*, $a_{cysbsa}(m)$, *public*, *bounded*$\rangle$. Fig. 5 shows an example of PD for the semantic term s and the degree of satisfaction (78%) of the measurement $m = 16$ to s (*Average size of big cysts*).

We define three types of PDs to model semantic terms shown in Fig. 4: *Left-bounded primitive, Right-bounded primitive*, and *Bounded primitive*. We also define a complex semantic term that is composed of multiple primitive terms concatenated by logical operations.

A. Left-Bounded Primitive Semantic Term

The left bounded primitive semantic terms assign a full degree of satisfaction to all the measurements that are greater than a specified value. Semantic terms such as *big, many*, and *huge* fall into this category. The following equation is used to model this type of primitive semantic term

$$a_s(m, \lambda_s^{L1}, \lambda_s^{L2}, \lambda_s^{L3}) = \begin{cases} \dfrac{2}{1 + e^{((\lambda_s^{L1} - m)/\lambda_s^{L2})^{\lambda_s^{L3}}}}, & \text{for } m < \lambda_s^{L1}. \\ 1, & \text{otherwise} \end{cases} \quad (1)$$

It is defined as the union of a constant function and a sigmoid function. The sigmoid function is centered at λ_s^{L1} and has width factor λ_s^{L2}, and an exponential factor λ_s^{L3}. The degree of satisfaction of the left bounded primitive semantic term equals 1 for any measurement $m \geq \lambda_s^{L1}$.

B. Right-Bounded Primitive Semantic Term

The right-bounded primitive semantic terms assign a full degree of satisfaction to all the measurements that are smaller than a specified value. Semantic terms such as *small, few, little* fall in this category. The following equation is used to model this type of primitive semantic term

$$a_s\left(m, \lambda_s^{R1}, \lambda_s^{R2}, \lambda_s^{R3}\right) = \begin{cases} 1, & \text{for } m \leq \lambda_s^{R1} \\ \dfrac{2}{1 + e^{((m - \lambda_s^{R1})/\lambda_s^{R2})^{\lambda_s^{R3}}}}, & \text{otherwise} \end{cases}. \quad (2)$$

It is defined as the union of a constant function and one sigmoid function. The sigmoid function is centered at λ_s^{R1} and has width factor λ_s^{R2}, and an exponential factor λ_s^{R3}. The degree of satisfaction of the right bounded semantic term equals 1 for any measurement $m \leq \lambda_s^{R1}$.

C. Bounded Primitive Semantic Term

The bounded primitive combines the characteristics of the previously defined semantic terms. It assigns a full degree of satisfaction to all the measurements in an specified interval. Semantic terms such as *average* and *medium* fall in this category. The following equation is used to model this type of primitive semantic term

$$a_s\left(m, \lambda_s^{L1}, \lambda_s^{L2}, \lambda_s^{L3}, \lambda_s^{R1}, \lambda_s^{R2}, \lambda_s^{R3}\right) = \begin{cases} \dfrac{2}{1+e^{((\lambda_s^{L1} - m)/\lambda_s^{L2})^{\lambda_{L3}}}}, & \text{for } m < \lambda_s^{L1} \\ 1, & \text{for } m \in \left[\lambda_s^{L1}, \lambda_s^{R1}\right]. \\ \dfrac{2}{1+e^{((m - \lambda_s^{R1})/\lambda_s^{R2})^{\lambda_s^{R3}}}}, & \text{for } m > \lambda_s^{R1} \end{cases} \quad (3)$$

It is defined as the union of a constant function and two sigmoid functions. The sigmoid functions are centered at λ_s^{L1} and λ_s^{R1} with width factors λ_s^{L2} and λ_s^{R2}, and the exponential factors λ_s^{L3} and λ_s^{R3}. The degree of satisfaction of the bounded semantic term equals 1 for any measurement $m \in [\lambda_s^{L1}, \lambda_s^{R1}]$.

D. Complex Semantic Term

Let s_1, s_2 be two semantic terms and a_{s_1}, a_{s_2} be the associating PDs. We define a set of logic operators for these functions $O_p = \{\text{AND}, \text{OR}, \text{NOT}\}$ where

$$\text{AND}\,(a_{s_1}, a_{s_2}) = \min\,(a_{s_1}, a_{s_2})$$

$$\text{OR}\,(a_{s_1}, a_{s_2}) = \max\,(a_{s_1}, a_{s_2})$$

and

$$\text{NOT}\,(a_{s_1}) = 1 - a_{s_1}.$$

A complex semantic term is defined as $s = \langle c, l, d, o_p(s_1, s_2, \ldots, s_i), o, t\rangle$, where $o_p(s_1, s_2, \ldots, s_i)$ is the rule to compose multiple primitive semantic terms or other complex terms using logic operators in O_p, and all other variables are defined in Section II. For example, we can construct a complex semantic term s—*Many, above average size, with sparse coverage calcified regions* by combining the possibility distributions of two primitive semantic terms and one complex semantic term. s_1—*Many calcified regions* and s_2—*Sparse coverage of calcified regions* are primitive terms. The semantic term *Above average size calcified region* is defined by a user who wanted to find images with calcified regions that are either big or average size. Such term is not in the collection of the primitive semantic terms defined in the *Semantic domain*. Therefore, an intermediate complex term s_3—*Above average size calcified regions*, is constructed by applying OR logic to another two primitive terms: s_4—*Average size calcified regions* and s_5—*Big calcified regions*. The PD for s_3 is

$$a_{s_3} = \text{OR}\left(a_{s_4}, a_{s_5}\right) = \max\left(a_{s_4}, a_{s_5}\right) \tag{4}$$

where a_{si} is the PD for s_i. Subsequently, the PD for s is expressed by

$$\begin{aligned} a_s &= \text{AND}\left(a_{s_1}, \text{OR}\left(a_{s_4}, a_{s_5}\right), a_{s_2}\right) \\ &= \min\left(a_{s_1}, a_{s_2}, \max\left(a_{s_4}, a_{s_5}\right)\right). \end{aligned} \tag{5}$$

IV. Query System

The main tasks performed by the *Query system* are: 1) processing semantic query constraints from the user's input; 2) searching image databases by semantics; and 3) accumulating the query history for updating the user's preferences. For a given query, such as "retrieve lung images with big cysts," the *Query system* first finds the semantic term *Big cysts* from the semantic profile tree, see Fig. 3, and then forms a possibility distribution for this semantic term on-the-fly. The system ranks the qualified images based on the descending order of the degree of satisfaction by substituting the measurement, in this example the size of cysts, into the PD function. These three tasks are implemented using the following pseudocode:

```
01 Query_system(user)
02   QU = nil; // initialize query constraints set
03   RIS = nil; // initialize retrieved images set
04   // semantic term selection
05   do
06     if selected term is a linguistic variable then
07       display term's child nodes;
08     else if term is semantic term then
09       add term to QU
10     end if
11   while user selects more terms
12   // image ranking
13   for i=1 to size[image database] do
14     aqu[i]=0;// initialize overall degree of satisfaction
for image i
15     DS = array[length[QU]] // set of degrees of
satisfaction for image i
16     for j=1 to length [QU] do
17       Ds[j]=asu(ms[i]); // degree of satisfaction of image
i to semantic term j
18     end for
19     aqu[i] ← min(DS[j]),1 ≤ j≤ length[QU];
20   end for
21   // display top ranked images
22   rank images on aqu[.] in descending order
23   RIS ← top ranked images
24   display images in RIS
25 end query_system;
```

A. Selecting Semantic Terms

To select a set of semantic terms, users access their *working* semantic profiles and select linguistic variables (*Size of cysts*) with semantic terms (*big*) as shown in lines 04 to 11 of the Query_system function. In our system a query constraint is defined as the PD a_s assigned to a semantic term s in the *working* profile of user u. The *user query* $Q_u = \{s1, s2, \ldots, sb\}$ is formed by the set of all b querying constraints defined by the user in the semantic query. $Q_u = \{cysbsa, nodsm\}$ is an example shows a user's query for the semantic terms *Average-size cysts*, and *Many small nodules*.

B. Querying Semantic Databases

Once the search criteria have been defined by the user and processed by the system, the system ranks database image signatures in the *Feature domain* (lines 11 to 23 of the Query_system function.) Let N_I be the number of images archived in the database. The query module parses the system's feature profile, introduced in Section II, for each image and determines degree of significance $a_s(m_l(i))$ between the semantic terms $s \in Q_u$ defined over the linguistic variable l and the measurement $m_l(i)$ from the feature domain of images $i \in [1, N_I]$. In the case of querying for multiple constraints, the degree of joint satisfaction of an image i is defined as $a_{Q_u}(i) = \min[a_s^u(m_l(i))], \forall s \in Q_u$. Fig. 6 shows an example of query system output for the semantic terms *Many cysts* and *Big cysts*.

There are three types of query solutions the system tries to construct: 1) *perfect*, 2) *good*, and 3) *partial* solutions. First, the system tries to deliver a *perfect solution*, that is retrieving only images with a full degree of joint significance $a_{Q_u}(i) = 100\%$. This constraint will define a b-dimension hyper cubical area in the feature domain,assuming we have b semantic terms in the user's query. If there are not enough images that are qualified for the perfect solution, the system will relax the constraints to a *good solution*, that is to also include images with partial degree of joint significance $0\% < a_{Q_u}(i) < 100\%$. The constraints can be further relaxed to a *partial solution* by removing some constraint(s) from the query. The constraint removal criterion is based on the selectivity of the PDs.

V. Synchronization of Semantic Descriptions

Physicians may use different descriptions for the same pathology due to their training and geographical locations. For example, the *Tree-in-bud* (TIB) pattern is a direct CT scan finding of bronchiolar disease. The same pattern could also be called *Finger-in-glove* [50]. In order to effectively accommodate different users, and ensure accurate and timely results, our system needs to address this semantic-variation issue because it can negatively affect the system performance.

It is quite possible that users may have a clear visual picture of a candidate semantic term that describes a desired perceptual category but be unfamiliar with the linguistic variables and/or semantic terms used by others and deposited to our system in the shared ontology. In addition, the same semantic meaning may already exist in the shared ontology, but be described differently. In such cases, querying the system by selecting semantic terms from the shared ontology will have limited relevance to the user. For these situations, we provide a system module to synchronize the meaning of the semantic terms between the user's semantics and the shared ontology if any inconsistency in wording exists. This module identifies linguistic variables or semantic terms that refer to the same perceptual category in the knowledge base and creates a synonym database for information exchange.

This process has an iterative approach that includes two steps: *image set selection* and *semantic set refinement*. To reduce the burden of the user, only representative images that cover all meaningful semantics in the shared ontology will be displayed. To accomplish this, we partition the semantic terms into groups. For each group, images displayed to the user maximize the relevance of the associating semantic terms. The user will be asked to rate the images on a scale from 0 to 10, with 0 corresponding to "*Excellent Counter-example*" and 10 to "*Excellent example*." After several iterations, the system is expected to converge to the most relevant semantic term from the shared ontology. This module is implemented using the following pseudocode:

```
01 Sync_semantics (user)
02 NSG ← number of semantic groups
03 DIS ← database image set
04 STS ← semantic term set
05 STG ← semantic term groups
06 do
06   // image set selection
07   for each image in DIS do
08     for each term in STS do
09       DS[image, term]=asu(ms[image]);
//degree of satisfaction of image to term
10     end for
11     for each group in STG do
12       γ[image, group] = min(DS[image, s]), s ∈ group;
13     end for
14   end for
15   for each group in STG do
16     select top 2 ranked images on γ[., group]
17   end for
18   display images to user
19   //Semantic set refinement
20   for each term in STS
21     for each rated image do
22       r=|rating−5|/5;
23         if rating > 5 then // positive example
24             βe[i][j] = min(r, asu (ms[rated image}]));
25         else // negative example
26             βc[i][j] = max(r, asu (ms[rated image]));
27         end if
28       end for
29       compute Be[term], BC[term], B[term]// see Eqs.
(11)-(13)
30       if B[term] < a preset threshold then
31         remove term from STS
32       end if
33     end for
34   while length[STS] > 1 and some term removed from STS;
35   return term;
36 end sync_semantics
```

TABLE VI
EXAMPLE OF USER RATING

Image Name	Rating	Meaning
Essence-054	2	Good Counter-example
Essence-418	9	Very Good Example
Essence-809	10	Excellent Example
Essence-941	1	Very Good Counter-example
Essence-980	0	Excellent Counter-example
Essence-520	3	Fair Counter-example
Essence-963	6	Poor Example
Essence-565	7	Fair Example

A. Image Set Selection

As mentioned previously, image selection maximizes the relevance of displayed images to the semantic terms in the image set. Let N_M be the number of relevant semantic terms from the shared ontology. The semantic terms are partitioned into N_G groups, where

$$N_G = \lceil \min(\psi \cdot \ln(N_M), N_M) \rceil. \tag{6}$$

Equation (6) ensures that when the semantic set is small, semantic terms are grouped individually. It also limits the number of groups when the semantic set is big by using the logarithmic function. The parameter ψ is used to scale up the result of the logarithmic function so it determines a reasonable number of groups to be used. For each group in N_G, the system will display two images. For example, if the relevant semantic term set includes 53 terms, and considering $\psi = 3.5$, the system will display 28 images, partitioned into 14 groups. The semantic terms with the highest correlated degree of satisfaction will be clustered in the same group. The degree of correlation among semantic terms is computed offline every time new images are added to the database.

The system selects relevant images to the gth group G_g (lines 11 to 13 in Sync_semantics) by computing a degree of relevance γ^i of each image i to the semantic terms in the group, using

$$\gamma_g^i = \min\left(a_{s_g}\left(m(i)\right)\right) \mid s_g \in G_g. \tag{7}$$

This approach guaranties that all the other semantic terms in the group will have the degree of satisfaction greater than or equal to γ_g^i. Then, we maximize γ_g^i among all the images in the database. As shown in lines 15 to 17 in the Sync_semantics function, an image i is selected to be displayed in G_g if

$$\gamma_g^{i*} = \max(\gamma_g^i) | i \in [1, N_I]. \tag{8}$$

We repeat the same image selection process for other semantic groups, without including the already selected images. After images are selected for all the groups, the system displays them to the user for rating. Table VI shows an example of user rating on a scale from 0 to 10, with 0 corresponding to *Excellent Counter-example* and 10 to *Excellent example*. Once the system receives the user's ratings for this image set, it further evaluates the relevance of each term in the semantic set to decide the next relevant semantic term set. This iterative process stops when only one semantic term was determined to be relevant.

B. Semantic Set Refinement

The initial semantic set selected by the system is often too general to finalize the synchronization of the semantic meaning. The system will take the ratings of positive examples and counterexamples from the user's feedback to select a more significant set of images for the next iteration. This process intends to create a much smaller set of semantic terms from the shared ontology. Once a new set of semantic terms is defined, a new set of images is presented to the user. The user can follow the same process described in the previous sub-section to refine the synchronization results.

Let i_e be a positive example image, s be a term in the relevant semantic set, and r be the rating factor as described in line 22 of the pseudocode. We define $\beta_e\ (s, i_e)$ as the relevance degree of the positive example i_e to the semantic term s with

$$\beta_e(s, i_e) = \min(a_s(m(i_e)), r). \tag{9}$$

Similarly, let i_c be a counterexample image selected by user, s a semantic term, and r the rating factor as described in line 22 of the pseudocode. We define $\beta_c(s, i_c)$ as the degree of dissimilarity of the counterexample i_c with the semantic term s

$$\beta_c(s, i_e) = \min(1 - a_s(m(i_c)), r). \tag{10}$$

From (9) and (10) we can compute $B_e(s)$ as the degree to which there exist at least one highly representative example for s. We also compute $B_c(s)$ as the degree to which all highly representative counterexamples are irrelevant to s

$$B_e(s) = \max(\beta_e(s, i_e)), \forall i_e \tag{11}$$

$$B_e(s) = \min(\beta_c(s, i_c)), \forall i_c. \tag{12}$$

We can estimate the overall degree of relevance for a semantic term s to a set of rated images by computing the following:

$$B(s) = \min(B_e(s), B_c(s)). \tag{13}$$

A semantic term will be excluded from the set if the overall relevance falls below a threshold. This process helps us to select the most relevant semantic terms that matches with the user's candidate semantic term. If the process doesn't converge to the most relevant one, the system applies this entire process for the next iteration until no more semantic terms from the shared ontology are excluded.

C. Updating the Knowledge Base

When the query refinement is completed, the user is presented with an option to enter his or her description of the candidate semantic term. The new description is then populated into the knowledge base using entries with the attributes: $\langle$*indexing-code, type, value*$\rangle$. For example if the most meaningful semantic term in the shared ontology synchronized with the user's candidate term is *fin_gl_big*, and the user description is *Big finger-in-glove*, the following new description will be populated into the knowledge base: $\langle$*fin_gl_big*, "synonym", *Big finger-in-glove*$\rangle$

VI. Information Exchange

Information exchange is very important in any shared system. While preference customization is important for end-users, it also makes information exchange difficult. Our system provides two types of information exchange which are: 1) system-level information organization and sharing and 2) peer-to-peer information exchange.

A. System-Level Information Organization and Sharing

There are several reasons that a semantic retrieval could lead to an unacceptable result. In diagnostic image retrieval, the process of articulating perceptual categories, as well as quantifying the associated semantic terms, proves to be highly subjective. Therefore, a robust semantic search engine should allow the users to modify the quantification of existing semantic terms and to add new ones if needed. Upon reviewing the retrieval results of query Q_u, if the user u decides that the results are not satisfactory, he or she can either modify the possibility distribution of each semantic term in Q_u, or add a new semantic term to the linguistic variable, with the help of the system's web-based interface.

The flow of events for customizing the PD of a semantic term s is: 1) the system displays k training images having the measurement evenly distributed over the universe of the linguistic variable l and 2) the user rates the displayed images on a scale from 0 to 10. If the user's selection is not informative enough (rated few images or all low ratings), the system will repeat the similar process based on the high rated images selected in the previous iteration.

Determining the PD that best matches the user's preferences (Fig. 7) could be achieved by ensuring: 1) distribution completeness; 2) user preferences compensation; and 3) distribution regression. Let $m(i)$ be a measurement associated with the feature of an image i, r_i be the rating given by the user to image i, and s be the semantic term to be refined. As mentioned previously in Section II-A2, the function a_s must assign full degree of significance for at least one measurement m in the universe of discourse. The system ensures the completeness by computing $b = \max(\text{median}(r_i, r_{i+1}, \ldots, r_w)), \forall i \in [1, k]$ with varying w). The new membership function is then computed using $a_s(m(i)) = \min(1, r_i\ /\ b)$. At this point λ^{L1} and λ^{R1} can be determined using the following equations:

$$\lambda^{L1} = \min(m(i))\,|\,a_s(m(i)) = 1 \tag{14}$$

$$\lambda^{R1} = \min(m(i))\,|\,a_s(m(i)) = 1. \tag{15}$$

The sigmoid functions that best match the adjusted user's preferences are computed using a nonlinear least square fitting algorithm, and then the parameters $\lambda^{L2}, \lambda^{L3}, \lambda^{R2}$, and λ^{R3} are decided. This setting is saved in both the *user-specific* and *candidate* semantic profiles, while the user selections are saved in the *user preferences* knowledge base.

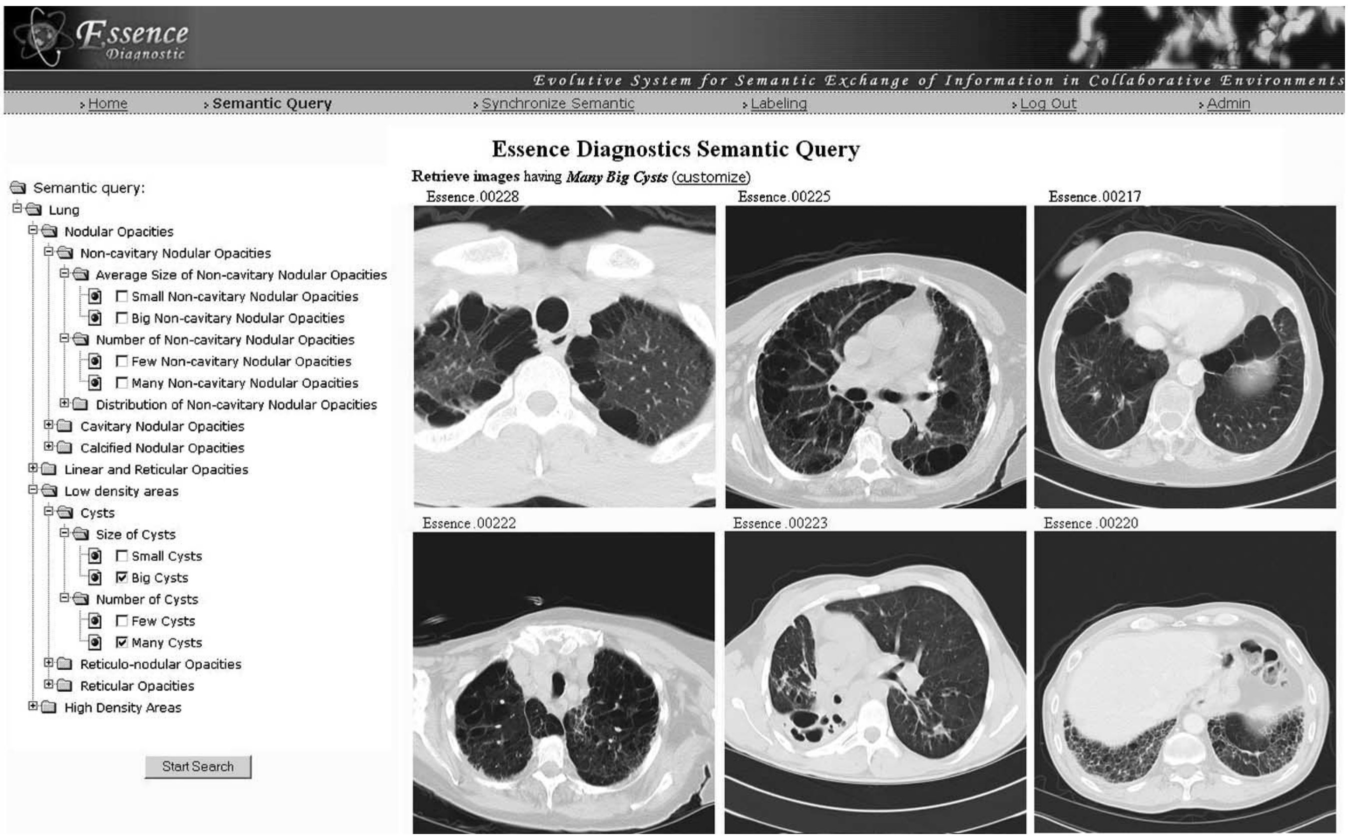

Fig. 6. Set of images retrieved upon querying for the semantic terms *Big Cysts* (λ_{cysb}^{L1}=6, λ_{cysb}^{L2}=1.25, λ_{cysb}^{L3}=2) and *Many cysts* (λ_{cysm}^{L1}=9, λ_{cysm}^{L2}=2, λ_{cysm}^{L3}=2).

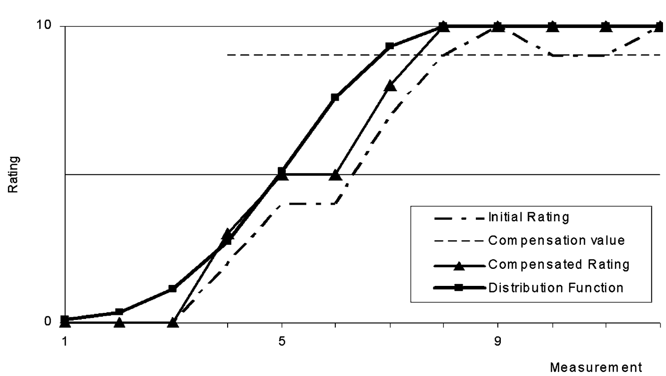

Fig. 7. The process of customizing PD to accommodate user's preferences.

Periodically, the system automatically triggers a learning component that updates PDs for the *default* profile. To do that, the system searches *user preferences* for the most recently updated distributions that are highly correlated to the *default* profile (correlation greater than 0.7). It then computes two weights: $w_{\mathrm{un}} = \log_2(n_{\mathrm{un}})/(\log_2(n_{\mathrm{un}}) + \log_2(n_{\mathrm{ue}}))$ for the qualified users' new ratings, and $w_{\mathrm{default}} = \log_2(n_{\mathrm{ue}})/(\log_2(n_{\mathrm{un}}) + \log_2(n_{\mathrm{ue}}))$ for the default PD. In these ratios, n_{ue} is the number of users that have already contributed to the *default* profile and n_{un} is the number of qualified users that will contribute to it. This approach progressively increases w_{default} to ensure the stability of the *default* profile. On the other hand, this system should be able to keep accepting new inputs from users even with a large number of users who previously contributed to the *default profile*. To deal with this, the logarithm function works by limiting the influence of w_{default} when n_{ue} is large. After both updated weights are computed, the system builds a new nonparametric PD by taking a weighted average from the default PD and the ratings from all qualified users. This is to adjust the default PD. An algorithm similar to the one described previously for *user-specific* profiles is then applied to form a new parametric default PD for the linguistic variable.

B. Peer-to-Peer Information Exchange

If, during query process, a user considers that the result has a high degree of relevance, the user can save the result in his or her *user preferences*. The user can share the results of this successful query with peers by sending them a reference to this query. Peers are able to visualize the resulting images directly, without an actual query action. A peer user could adopt the same PD in his or her *user-specific* profile for future retrievals.

VII. Experiment Results

To demonstrate the performance of our model, we designed three experiments. The first experiment tests the improvement in retrieval precision after the system evolves its shared knowledge settings by adapting to domain expertise. The second

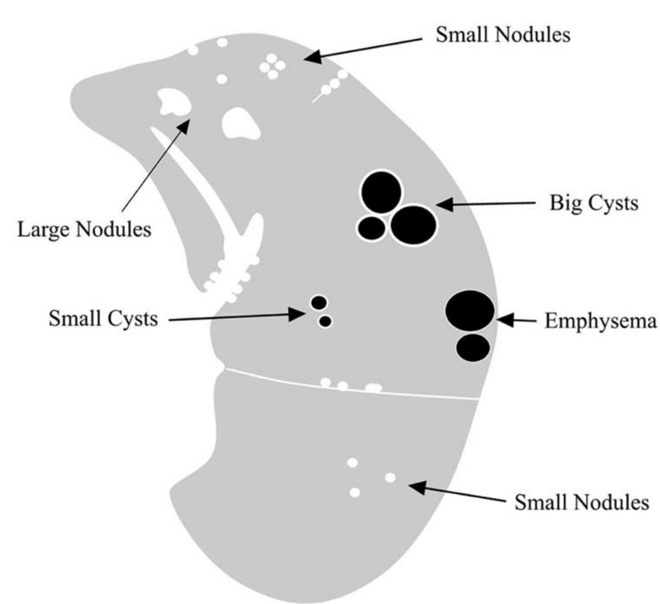

Fig. 8. Typical appearance of different lung pathologies used in our experiment.

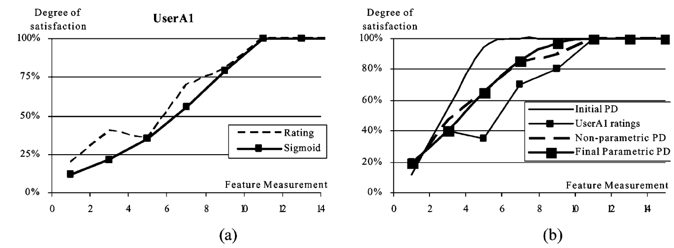

Fig. 9. PD for: (a) *UserA1*. (b) Shared ontology at *stage* 1 after *userA1* rating.

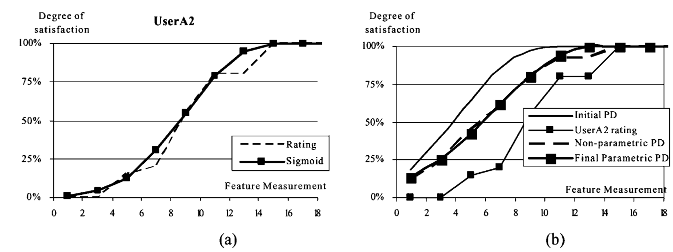

Fig. 10. PD for: (a) *UserA2*. (b) Shared ontology at stage 2 after *userA2* rating.

demonstrates the appropriateness of using the sigmoid functions described in Section III to quantize the semantic modeling. Finally, the third experiment evaluates the performance of the semantic integration mechanism described in Section V when searching for synonymous semantic terms. Upon completion of the experiments, the users were asked to fill a usability test. The results of the test are then discussed at the end of this section.

A. Simulated Scenario for Experiments

Users were shown, both visually and semantically, the typical appearance of *cysts* and *nodules* using the sketch shown in Fig. 8 and comparing them with other similar lung pathologies such as *emphysema*. Each user is then instructed by a domain expert, using a training image set, to identify the visual abnormalities of these pathologies on real HRCT lung images. This process emphasizes on the semantic terms that will be used in the experiment such as *Many small nodules*, and *Many big cysts*. For example, the term *Cyst* is used to refer to a lesion of a lung having the following characteristics [57]: well defined, circumscribed, air-containing, and thin-walled with size greater than 3 mm. It differentiates from *Emphysema* by the fact that the latter show very thin and less defined walls. From HRCT images, a *cyst* (perceptual category) with a diameter between 10–20 pixels (range of values) might be classified as medium size. The term *small nodule* [57] refers to a rounded density that does not correspond to vessels and is represented by a spherical structure having less than 1 cm in diameter.

All the experiments reported in this paper require users to rate the relevance of HRCT images of lung for one of these semantic terms. We used a rating scale from 0 to 10, where 0 corresponds to *Excellent Counter-example* and 10 to *Excellent Example.*

B. Improving the Retrieval Precision Through Adapting the Shared Ontology Settings

In this experiment, users were assigned into two groups: Group 1: *active users*, including the two domain experts and three computer scientists (*userA1* to *userA5*), and Group 2: *inactive users*, including two computer scientists (*userA6* and *userA7*). We assumed that the *inactive users*, although they have the expertise to customize their setting, prefer to use only the shared ontology. The purpose of this experiment is to evaluate the improvement in retrieval precision for the *inactive users* by benefiting from the *active users*' domain expertise. This process involves a system level knowledge exchange as described in Section VI. During this experiment, 869 images were rated by both *active* and *inactive* users.

To capture the evolving nature of this process, *active users* were asked to customize their settings for *Many big cysts* at different time intervals. At the end of each time interval, the system updated its default setting after each customization process. Each time interval represents a stage in the evolution of the shared ontology settings for this term. In this experiment, *userA1* customized his semantic profile at stage 1, *userA2* at stage 2, and *userA3, userA4*, and *userA5*, at stage 3.

Fig. 9(a) shows the shape of the user-specific possibility function after stage 1. It shows both the degree of satisfaction derived from the rating and its sigmoid approximation. For the shared PD shown in Fig. 9(b), the system uses both the initial PD and *userA1*'s ratings to determine the updated nonparametric distribution, and later its sigmoid parametric approximation. Fig. 10 follows the same idea but uses the initial possibility function computed at stage 1 as the initial function. In Fig. 11, the rating from three users contributes along with the previously determined possibility function to determine the shared PD.

At each stage, the *inactive users* were asked to query the database for *Many big cysts*, using the shared ontology settings.

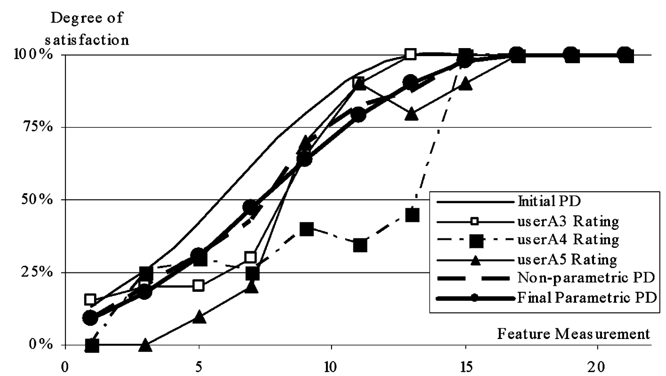

Fig. 11. Default PD after obtaining ratings from *userA3*, *userA4*, and *userA5*.

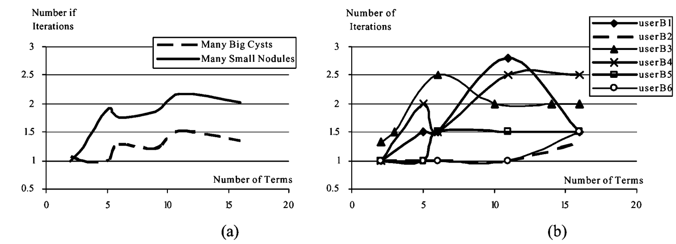

Fig. 12. Average number of iterations for each (a) semantic term and (b) user.

Then, they evaluated the retrieval result by rating the displayed images. From their ratings, we computed the retrieval precision as the percentage of images rated above 7 in the retrieval result (*Good* to *Excellent Example*). It improved from 25% in the initial stage to 65% after all training experts updated their possibility distribution. We conclude that using domain expertise to evolve the shared semantic settings can improve the retrieval precision for new and inactive users. The average time to complete this task, which required ratings of 20 images, was 110 s for computer scientists (standard deviation of 53 s) and 184.5 s for physicians (standard deviation of 42 s). From the above observations, the efficiency of the semantic customization process is acceptable for users. This is also consistent with the usability test which will be discussed in details later in this section.

C. Evaluating the Usage of Sigmoid Functions to Approximate the Possibility Function

This experiment evaluates appropriateness of using the sigmoid function in approximation of the possibility function. For comparison we use a linear function $f_l(m) = \max(0, \min(1, (m-b)/(a-b)))$, in which a and b are the values of low-level image measurements m with the degree of satisfaction 100% and 0%, respectively. To measure the efficacy of both functions, we computed the approximation error for both linear and sigmoid functions

$$E_{\text{linear}} = \sum_{\text{all ratings}} |f_l(m) - r(m)| \tag{16}$$

$$E_{\text{sigmoid}} = \sum_{\text{all ratings}} |a_s(m) - r(m)| \tag{17}$$

where $r(m)$ is the user's rating for the measurement m and $a_s(m)$ is the possibility function discussed in Section III.

The approximation performance was evaluated in 11 cases—seven of them were related to *user-specific* possibility functions and four to *default* ones. The sigmoid function outperformed the linear function in ten out of 11 cases by decreasing the error rate by 31% on average.

D. Evaluating the Semantic Integration Mechanism When Searching for Synonymous Semantics

We asked three physicians (*userB1* to *userB3*) and three computer scientists (*userB4* to *userB6*) to rate the existence of a candidate semantic term by using a set of HRCT images, an initial set of semantic terms, and the synchronization mechanism. The searched term has a visual pattern associated with terms archived in our shared ontology. For this experiment, we used candidate semantic terms that are synonymous to our targeting terms—*Many big cysts* or *Many small nodules*. The images presented to the user were selected according to the algorithm in Section V-A to cover all the significant terms. By asking users to search for these visual patterns, we evaluated both the accuracy and the rate of convergence in matching the candidate semantic terms with the targeting ones. The rate of convergence is defined as the number of iterations needed by the algorithm to converge to a unique semantic term.

Our experiments show that this process accurately converges to the targeting semantic term in 92.8% of the cases (26 out of 28 synchronizations). For both targeting terms the process converged in approximately two iterations on average, which demonstrates the viability of our approach in semantic set refinement discussed in Section V. However, the convergence rates differ between these two targeting terms. Fig. 12(a) shows that synchronizing a candidate term to *Many small nodules* requires 26% more iteration on average than synchronizing to *Many big cysts*. There are three reasons for this result, which are: 1) *many small nodules* is more likely to be co-existed with other semantic terms; 2) *many big cysts* is more easily recognizable than the *many small nodules* even without in-depth training; and 3) the behavior of each user can differ depending on their subjectivity. Fig. 12(b) shows the convergence rate of this process for each user with different sizes of initial set. The average time required to complete this task was 203 s for computer scientists (standard deviation of 118 s) and 292 s for medical experts (standard deviation of 59 s). The time required to do semantic synchronization is higher due to its recursive nature. However, the times measured in this experiment are reasonable to learn a new perceptual category without knowing the exact associating semantic term.

E. Usability Evaluation

Due to the fact that *Essence* is used by both medical experts and computer scientists, it is very important to evaluate how easy is for them to collaborate in such an environment.

TABLE VII
USABILITY TEST RESULT

Score	Computer scientists		Medical experts		Overall	
	Percent	Cumulative percent	Percent	Cumulative percent	Percent	Cumulative percent
5	26.25	26.25	32.81	32.81	29.16	29.16
4	57.50	83.75	43.75	76.56	51.38	80.55
3	7.50	91.25	17.18	93.75	11.80	92.36
2	6.25	97.50	6.25	100.00	6.25	98.61
1	2.50	100.00	0.00	100.00	1.38	100.00

To achieve this goal, we developed a usability questionnaire based on the SUS usability scale [8]. However, the ten questions in the original test are too general for the purpose of our study. We added six more questions from other usability questionnaires [14], [38] that addressed some more specific issues such as terminology, functionality and usefulness of retrieved images. Nine subjects-five computer scientists and four physicians-rated the usability according to the guidelines of the SUS test. The system was trained by experts before the experiment in order to stabilize the semantic assignments used in the experiment. Subjects filled the questionnaire at the end of the experiments discussed previously. The data was collected and further studied using analysis of means and variances of the usability ratings over the 16 questions. A perfect system would receive score five ratings. As listed in Table VII, we received 29.16% of score five ratings and 80.55% of score 4 or better ratings.

The lowest overall score was given to question "I understand the terminology used in the system" which received a score of 3. From the feedback provided by our subjects, medical terms are not intuitive to the computer scientists, while the interface terminology is not straightforward to the physicians. A noteworthy observation to report is that using the search tree (as shown on the left panel of Fig. 6) for semantic queries was new for most of the users at the beginning of the experiment. However, all users were successful in subsequent searches due to the intuitiveness of this type of search. On average, each semantic query takes 70.1 s to construct. The highest overall scores the system received were for its function integration and for effectiveness. Computer scientists also appreciated more the consistency of the system, while the medical experts appreciated the manageability and results of the queries.

The system was also evaluated on the SUS usability scale. The SUS scale yields results between 0–100, with 0 for poor perceived usability and 100 for high perceived usability. The study of Nielsen and Levy [38] shows that a system with average usability gets a score around 64 on such a *Likert* scale even though 50 represents neutral. The average SUS score for Essence was 77.22. Medical experts rated the system higher (average SUS score of 78.12) compared to the score of computer scientists (average SUS score 76.5).

All the physicians in this usability test considered the results of the semantic queries satisfactory without customizing their semantic assignment. Once the system is trained by domain experts, most of the physicians do not need to create new linguistic variables or customize existing semantic assignments. Under the condition when a new perceptual category is needed for certain newly discovered diseases, the community will ask for contributions from users. This knowledge exchange procedure consists of semantic synchronization and customization, and is believed to be acceptable by physicians who are enthusiastic about sharing their expertise with the databases.

VIII. Conclusion

In this paper, we have presented the Essence framework for knowledge representation and sharing in the radiology domain. It offers methods for physicians to refine their semantic settings on top of a shared ontology. This framework would be valuable for training and differential diagnosis, and could be the foundation of building a novel and flexible model for diagnostic medical image retrieval that uses physician-defined semantics. It accomplishes these tasks by assigning customized possibility distributions for each semantic term defined, and/or by adding new semantic terms. Although the physician's decision-making process relies upon precise, scientific tests and measurements, it also incorporates evaluations of symptoms and relationships among human perception and semantic terms in a fuzzy and intuitive manner. The framework also facilitates knowledge exchange among physicians through peer-to-peer and centralized channels. There are three keys that make our work unique, which are: 1) knowledge sharing and semantic setting customization; 2) physicians' defined linguistic variables closely related to known pathologies; and 3) more desirable results obtained by customizing the semantic terms attached to these linguistic variables. Currently, there is no truly successful system for knowledge exchange among physicians for diagnostic image databases. Although our framework is applied specifically to HRCT lung images, we believe this approach is likely to be accepted by physicians. With appropriate extensions, the Essence framework can also be adapted to other modalities of medical images.

Our future work includes extending our current shared ontology by integrating existing standards for the radiology domain, such as semantic networks in UMLS, and more comprehensive testing on different linguistic variables for different perceptual categories.

Acknowledgment

The authors would like to thank the reviewers who provided many constructive comments for us to improve the quality of this paper. The authors also would like to thank Dr. R. Singh, from CDC, Dr. P. Pancoast, from HealthLink Inc., Dr. G. Arthur, and Dr. D. Mitchel, from the Health Management and Informatics Department of the University of Missouri for assignment of linguist variables and system evaluation. Appreciation also goes to W. He from the School of Information Science and Learning Technology of the University of Missouri for fruitful discussions in developing the usability test.

REFERENCES

[1] A. Aguilera, A. R. Subero, L. Martinez, A. Subero, and L. J. Tineo, "Fuzzy image retrieval system," in *IEEE Int. Fuzzy Systems Conf.*, 2001, pp. 1247–1250.

[2] A. Barb, C. R. Shyu, and Y. Sethi, "Semantic integration and knowledge exchange for diagnostic medical image databases," in *Proc. IEEE 4th Symp. Bioinformatics and Bioengineering (BIBE2004)*, Taichung, Taiwan, R.O.C., May 2004, pp. 19–21.

[3] A. S. Barb and C. R. Shyu, "Semantics modeling in diagnostic medical image databases using customized fuzzy membership functions," in *Proc. IEEE Int Conf. Fuzzy Systems*, St. Louis, MO, May 2003, pp. 1159–1164.

[4] ——, "A semantic query approach for diagnostic image database management—Content-based access and knowledge exchange using XML," in *Proc. 6th World Multiconference on Systemics, Cybernetics, and Informatics*, vol. XVIII, Orlando, FL, Jul. 2002, pp. 162–167.

[5] P. G. Baker, C. A. Goble, S. Bechhofer, N. W. Paton, R. Stevens, and A. Brass, "An ontology for bioinformatics applications," *Bioinformatics*, vol. 15, no. 6, pp. 510–520, 1999.

[6] S. Berchtold, C. B. Shm, and H.-P. Kriegel, "The pyramid-technique: Towards breaking the curse of dimensionality," in *Proc. Int. Conf. Management of Data (SIGMOD'98)*, Seattle, WA, 1998, pp. 142–153.

[7] B. Bradshaw, "Semantic based image retrieval: A probabilistic approach," in *Proc. 8th ACM Int. Conf. Multimedia*, 2000, pp. 167–176.

[8] J. Booke, "SUS—A quick and dirty usability scale," in *Usability Evaluation in Industry*, P. W. Jordan, B. Thomas, B. A. Weerdmeester, and I. L. McClelland, Eds. London, U.K.: Taylor and Francis, 1996, pp. 189–194.

[9] M. S. Brown, M. F. McNitt-Gray, J. G. Goldin, and D. R. Aberle, "An extensible knowledge-based architecture for segmenting computed tomography images," in *Medical Imaging 1998: Image Processing Proc. SPIE–Int. Soc. Opt. Eng.*, K. M. Hanson, Ed., vol. 3338, 1998, pp. 564–574.

[10] B. Brehmer, "Models of diagnostic judgments," in *Human Detection and Diagnostic of System Failures*, J. Rasmussen and W. B. Rouse, Eds. New York, NY: Plenum, 1981, pp. 231–241.

[11] A. Burgun, G. Botti, M. Fieschi, and P. L. Beux, "Sharing knowledge in medicine: Semantic and ontologic facets of medical concepts," in *IEEE Proc. Syst., Man, Cybern.*, 1999, pp. 300–305.

[12] W. Cai, D. Feng, and R. Fulton, "Content-based retrieval of dynamic PET functional images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 4, no. 2, pp. 152–158, Jun. 2000.

[13] W. W. Chu, I. T. Ieong, and R. K. Taira, "A semantic modeling approach for image retrieval by content," *VDLB J.*, pp. 445–477, 1994.

[14] K. Cogdill, MEDLINE plus interface evaluation, Final Rep. University of Maryland Human-Computer Interaction Laboratory, Aug. 1999. [Online]. Available: http://citeseer.ist.psu.edu/cogdill99medlineplus.html

[15] I. F. Cruz and A. Rajendran, "Semantic data integration in hierarchical domains," *IEEE Intell. Syst.*, vol. 18, no. 2, pp. 66–73, Mar.–Apr. 2003.

[16] D. Dubois, "Fuzzy logic technologies in multimedia database querying: A preliminary investigation of potentials," *IEEE Trans. Knowl. Data Eng.*, vol. 13, no. 3, pp. 383–392, May/Jun. 2001.

[17] G.-P. Economou, D. Lymberopoulos, E. Karavatselou, and C. Chassomeris, "A new concept toward computer-aided medical diagnosis—A prototype implementation addressing pulmonary diseases," *IEEE Trans. Inf. Technol. Biomed.*, vol. 5, no. 1, pp. 55–65, Mar. 2001.

[18] J. Fox and R. Thomson, "Decision support and disease management: A logic engineering approach," *IEEE Trans. Inf. Technol. Biomed.*, vol. 2, no. 4, pp. 217–228, Dec. 1998.

[19] D. Gardner, K. H. Knuth, M. Abato, S. M. Erde, T. White, R. DeBellis, and E. P. Gardner, "Common data model for neuroscience data and data model exchange," *J. Amer. Med. Inform. Assoc.*, vol. 8, no. 1, pp. 17–33, 2001.

[20] R. Giacobazzi and I. Mastroeni, "Compositionality in the puzzle of semantics," in *Proc. ACM SIGPLAN Symp. PEPM*, Portland, OR, Jan. 2002, pp. 87–97.

[21] P. Hansen, "An exploratory study of IR interaction for user interface design," in *ACM SIGIR'97 20th Int. ACM SIGIR Conf. Research and Development in Information Retrieval*, Philadelphia, PA, Jul. 1997, pp. 1–55.

[22] C. C. Hsu, W. W. Chu, and R. K. Taira, "A knowledge-based approach for retrieving images by content," *IEEE Trans. Knowl. Data Eng.*, vol. 8, no. 4, pp. 522–532, Aug. 1996.

[23] J. Ingenerf, "Telemedicine and terminology: Different needs of context information," *IEEE Trans. Inf. Technol. Biomed.*, vol. 3, no. 2, pp. 92–100, Jun. 1999.

[24] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall, 1998.

[25] P. M. Kelly, T. M. Cannon, and D. R. Hush, "Query by image example: the CANDID approach," *Storage and Retrieval for Image and Video Databases III*, vol. 2420, pp. 238–248, 1995.

[26] T. Kindberg, N. Bryan-Kinns, and R. Makwana, "Supporting the shared care of diabetic patients," in *Proc. Int. ACM SIGGROUP Conf. Supporing Group Work*, Phoenix, AZ, Nov. 1999, pp. 91–100.

[27] F. Korn, N. Sidiropoulos, C. Faloutsos, E. Siegel, and Z. Protopapas, "Fast and effective retrieval of medical tumor shapes," *IEEE Trans. Knowl. Data Eng.*, vol. 10, no. 6, pp. 889–904, Nov. 1998.

[28] K. Kruempel, "Making the right (interactive) moves for knowledge-producing tasks in computer-mediated groups," *IEEE Trans. Prof. Commun.*, vol. 43, pp. 185–195, Jun. 2000.

[29] G. Leroy and H. Chen, "Meeting medical terminology needs-the ontology-enhanced medical concept mapper," *IEEE Trans. Inf. Technol. Biomed.*, vol. 5, no. 4, pp. 261–270, Dec. 2001.

[30] S. Madasani and R. Krishnapuram, "A fuzzy approach to content-based image retrieval," in *Proc. Fuzzy Systems Conf. FUZZ-IEEE*, vol. 3, 1999, pp. 1251–1260.

[31] C. Marietti, "Extensibility sensibility. Is XML the solution for electronic medical records data exchange?," *Healthcare Informatics*, pp. 51–64, Mar. 1998.

[32] V. Maojo and C. A. Kulikowski, "Bioinformatics and medical informatics: Collaborations on the road to genomic medicine?," *J. Amer. Med. Inform. Assoc.*, vol. 10, no. 6, pp. 515–522, 2003.

[33] D. May and P. Taylor, "Knowledge management with patterns," *Commun. ACM*, vol. 46, no. 7, pp. 94–99, 2003.

[34] G. A. Miller, "WordNet: A lexical database for english," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, Nov. 1995.

[35] S. Mitaim and B. Kosko, "The shape of fuzzy sets in adaptive function approximation," *IEEE Trans. Fuzzy Syst.*, vol. 9, no. 4, pp. 637–656, Aug. 2001.

[36] N. Mouaddib and N. Bonnano, "New semantics for the membership degree in fuzzy databases," in *IEEE Proc. ISUMA-NAFISP*, 1995, pp. 655–660.

[37] Y. Nah and P. C.-Y. Sheu, "Image content modeling for neuroscience databases," in *Proc. Int. Software Engineering and Knowledge Engineering Conf.*, Italy: Ischia, Jul. 2002, pp. 91–98

[38] J. Nielsen and L. Levy, *Measuring Usability: Preference vs. Performance, Commun. ACM*, vol. 37, no. 4, 1994, pp. 66–75.

[39] D. E. O'Leary, "Enterprise knowledge management," *IEEE Computer*, vol. 31, no. 3, pp. 54–61, Mar. 1998.

[40] N. Otsu, "A threshold selection method from gray-level histogram," *IEEE Trans. Syst. Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.

[41] J. Rasmussen, "Diagnostic reasoning in action," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 4, pp. 981–992, Jul.–Aug. 1993.

[42] B. N. Reeves, F. Shipman, and E. D. Mynatt, "Tacit knowledge: Icebergs in collaborative designs," *ACM SIGOIS Bull.*, vol. 17, no. 3, pp. 24–33, 1996.

[43] G. P. Robinson, H. D. Tagare, J. S. Duncan, and C. C. Jaffe, "Medical image collection indexing: Shape-based retrieval using KD-tree," *Computerized Medical Imaging and Graphics*, vol. 20, no. 4, pp. 209–217, 1996.

[44] J. C. Russ, *The Image Processing Handbook*, 4th ed. Boca Raton, FL: CRC Press, 2002.

[45] R. Saint-Paul, G. Raschia, and N. Mouadib, "Prototyping and browsing image databases using linguistic summaries," in *IEEE Int. Conf. Fuzzy Systems*, vol. 1, Hawaii, 2002, pp. 476–481.

[46] C. R. Shyu, "A physician-in-the-loop content-based image retrieval system for medical image databases," Ph.D. dissertation, Purdue Univ., West Lafayette, IN, 1999.

[47] C. R. Shyu, C. E. Brodley, A. C. Kak, A. Kosaka, A. M. Aisen, and L. S. Broderick, "ASSERT: A physician-in-the-loop content-based image retrieval system for HRCT image databases," *Comput. Vis. and Image Understanding*, vol. 75, no. 1/2, pp. 111–132, 1999.

[48] C. R. Shyu, C. Pavlopoulou, A. C. Kak, C. E. Brodely, and L. S. Broderick, "Using human perceptual categories for content-based retrieval from a

medical image database," *Comput. Vis. and Image Understanding*, vol. 88, no. 3, pp. 119–151, 2002.
[49] C. R. Shyu, J. Harnsomburana, Y. Sethi, R. Singh, and J. Reid, "A hybrid query approach for radiology education—Integrating content-based medical image retrieval and text-based information retrieval," *Int. Comput. J. Appl.*, vol. 24, no. 2, pp. 83–92, 2002.
[50] E. J. Stern and J. Collins, "HRCT: Know your buzz words," in *Society of Thoracic Radiology Annu. Meetings*, 2000, pp. 205–207.
[51] E. J. Stern and S. J. Swensen, *High Resolution CT of the Chest: Comprehensive Atlas*, 2nd ed. Philadelphia, PA: Lippincott Williams & Wilkins, 2000.
[52] A. Tiwana, "Affinity to infinity in peer-to-peer knowledge platforms," *Commun. ACM*, vol. 46, no. 5, pp. 77–80, May 2003.
[53] N. Tayar, "A model for developing large shared knowledge bases," in *IEEE Proc. Information and Knowledge Management*, Washington, DC, Nov. 1993, pp. 717–719.
[54] D. A. B. Lindberg and B. L. Humphreys, *UMLS Knowledge Sources*, 14th ed. Bethesda, MD: Na. Library Med., 2003. 2003AB.
[55] J. V. E. O. Oliveira, "Semantic constraints for membership function optimization," *IEEE Trans. Syst. Man, Cybern A, Syst.*, vol. 29, no. 1, pp. 128–138, Humans, Jan. 1999.
[56] W. R. Webb, N. L. Muller, and D. P. Naidich, *High-Resolution CT of Lung*. Philadelphia, PA: Lippincott-Raven, 1996.
[57] C. P. Wei, P. J. H. Hu, and O. R. L. Sheng, "A knowledge-based system for patient image pre-fetching in heterogeneous database environments—Modeling, design, and evaluation," *IEEE Tran. Inf. Technol. Biomed.*, vol. 5, no. 1, pp. 33–45, Mar. 2001.

Adrian S. Barb (S'04) received the B.S. degree in industrial engineering from the School of Engineering, University of Bucharest, Bucharest, Romania, in 1990, and the M.B.A. degree from the University of Missouri, Columbia (MU) in 2002. He is currently, working toward the Ph.D. degree in computer science at MU.

He also worked as a Database Programmer Analyst for the Information Access and Technology Services at MU. His research interests include knowledge representation and exchange in content-based retrieval systems, semantic modeling and retrieval, capturing and predicting conceptual change in knowledge-base systems, ontology integration, and expert-in-the-loop knowledge exchange.

Chi-Ren Shyu (S'89–M'00) received the M.S.E.E. and Ph.D. degrees from the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, in 1994 and 1999, respectively. Between 1999 and 2000, he was a Postdoctoral Research Associate of the Robot Vision Laboratory, at Purdue University. Since October 2000, he has been an Assistant Professor in the Computer Science Department, University of Missouri, Columbia (MU). Currently, he is an Adjunct Faculty of the Health Management and Informatics Department and a core faculty for the National Library of Medicine pre-doctoral and post-doctoral training program at MU. He is the Founding Director of the Medical and Biological Digital Library Research Laboratory. He also serves as a Chief Computer Engineer for the Family Practice Inquiry Network (FPIN) and an Executive Committee Member for the Information Delivery System (IDS) of FPIN. His research interests include biomedical informatics, geospatial information management, medical imaging, computer vision, and pattern recognition.

Dr. Shyu is a member of Association for Computing Machinery (ACM) and American Medical Informatics Association (AMIA).

Yash P. Sethi received the M.B.B.S. and M.D. degrees from Delhi University, New Delhi, India, in 1988 and 1993, respectively.

His U.S. residency training includes internal medicine from Downstate SUNY-HSC, Brooklyn, NY, 1995–1996; radiology residency from Wayne State University, Detroit Medical Center, MI, 1996–1998; and radiology residency from New York Medical College, St. Vincent's Hospital, NY, 1998–2000. Subsequently, he completed the Fellowship in cross sectional imaging from University of Missouri and American Board of Radiology (Member of American Boards of Medical Specialties) in 2001. He is since been a Clinical Assistant Professor of Radiology at the University of Missouri, Columbia. He is also the Radiology Residency Associate Program Director and has been co-awarded best teacher in all four years of his faculty position. His research interests include radiologic imaging with a focus on computer applications, 3-D imaging, and advanced imaging.

Dr. Sethi is a Member of the Radiological Society of North America (RSNA), American College of Radiology (ACR), American Medical Association (AMA), and the Missouri State Medical Association.

He was also awarded Diplomat of National Board, Government of India, in Radiology in 1993.