Ontology based EMR for decision making in health care using SNOMED CT

Conference Paper · April 2012

DOI: 10.1109/ICRTIT.2012.6206787

2 authors:

Josephine Julina J. K.
Sri Sivasubramaniya Nadar College of Engineering
3 PUBLICATIONS **6** CITATIONS
SEE PROFILE

D. Thenmozhi
Sri Sivasubramaniya Nadar College of Engineering
12 PUBLICATIONS **16** CITATIONS
SEE PROFILE

Ontology Based EMR for Decision Making in Health Care Using SNOMED CT

J. Kulandai Josephine Julina
Department of Computer Science and Engineering
SSN College of Engineering
Chennai, India
josephinejulina@gmail.com

D. Thenmozhi
Department of Computer Science and Engineering
SSN College of Engineering
Chennai, India
theni_d@ssn.edu.in

Abstract— **Health care domain is gaining more focus in the field of Ontology and Semantic research. Ontology in medicine helps in proper representation and organization of clinical terminologies. An Electronic Medical Record (EMR) based on Ontology plays a significant role in decision making process. However there is a lack of standards in constructing Ontology, based on the current needs of the healthcare professionals, which results in poor quality of information extraction. This paper proposes the use of Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT), an existing medical Ontology which can be restructured to a high quality Ontology for achieving efficient information extraction. An EMR captures accurate patient information based on enriched Ontology thereby making it suitable for decision making with the help of Bayesian Belief approach.**

Keywords- Bayesian Network; Electronic Medical Record; Health Care; Lung Diseases; Ontology; SNOMED CT

I. INTRODUCTION

Health care is a complex field with large number of clinical terminologies which require structured storage and retrieval for maintaining consistency of information access. Health care is an important area of Ontology. Ontology provides a common understanding of a domain that can be communicated between people and heterogeneous and distributed systems. Tom Gruber defined Ontology as "explicit and formal specification of a shared conceptualization of a domain of interest". Ontology is basically a representation of domain in terms of concepts and relationships each with precise semantics described using the OWL (Web Ontology Language) and provide representation of knowledge.

Representing knowledge through building ontology for Electronic Medical Records (EMRs) is important to achieve semantic interoperability among healthcare information systems and to better execute decision support systems [1]. EMR becomes the primary mechanism of communication and its successful use reduces errors in data entry, reduces tedious paper work, secure patient identity, lowers medical costs, makes easy processing of data, enables quicker decision making, provides quicker results in patient treatment and management, and easier use of patient information.

Rest of the paper is organized as follows: Section II provides details about literature survey. Section III provides details about the architecture of the proposed method for decision making. Section IV provides experimental results and Section V concludes our work with references.

II. LITERATURE SURVEY

A. Ontology

Ontology is a philosophical discipline which deals with the nature and organization of reality [10]. Ontology learning (ontology extraction, ontology generation, or ontology acquisition) is a subtask of information extraction. Ontology learning can be made from unstructured sources (NLP techniques), semi-structured sources (XML, HTML) and structured data (SNOMED CT). The goal of Ontology learning is to (semi-)automatically extract relevant concepts and relations from a given corpus or other kinds of data sets to form Ontology. Ontology enrichment deals with refinement of existing ontology with new domain specific concepts. OWL can be used to explicitly represent the meaning of terms in vocabularies and the relationships between those terms [7]. This representation of terms and their interrelationships is called ontology.

With the fast growing clinical terminologies and changing needs of the clinicians in different hospitals the Ontology needs to be restructured so as to ensure good quality of information retrieval.

Some of the existing medical Ontologies include Unified Medical Language System (UMLS), GuideLine Interchange Format (GLIF), Generalized Architecture for Languages (GALEN), International Classification of Diseases (ICD), Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) [12] and so on out of which SNOMED CT is efficient and powerful.

B. SNOMED CT

SNOMED CT developed by the College of American Pathologists and United Kingdom's National Health Service maintains structured and consistent information that makes healthcare information useable and accessible. It is a comprehensive and precise reference medical Ontology with huge number of concepts and relationships [9].

SNOMED CT includes hierarchical "is-a" relationships and other relationships describing clinical attributes.

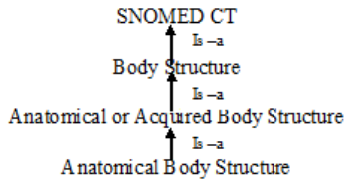

Figure 1. Hierarchical Relationships in SNOMED CT

SNOMED- CT is a concept-oriented and machine-readable medical terminology used in electronic medical records. Information retrieval and information extraction are significant issues in the medical and health care domains where the accuracy of the retrieved information and obtaining it in a time critical situation are extremely important [9]. The clinical descriptions in practice can be transformed into the SNOMED CT concepts and descriptions which compose the final clinical terminology [12]. Formal vocabulary is maintained in the knowledge base which includes the existing SNOMED CT Ontology in addition to the new domain specific concepts provided by clinicians. SNOMED CT simplifies the search for diseases and symptoms; however if a physician wishes to use this system they can find only simple hierarchies of SNOMED CT codes, which they must then interpret [4]. Hence the existing SNOMED CT needs to be restructured to meet specific needs of the clinicians.

C. EMR for Decision Making

The current system in hospitals whereby doctors enter patient information using paper charts is cumbersome, time consuming and does not facilitate knowledge sharing types of information, and valuable time is often lost trying to correlate data in order to diagnose and treat patients [11].

Electronic Health Record provides reliable and better quality health care. The EMR becomes inefficient when it fails to capture data which is not useful for semantic matching or may be inefficient in arriving at appropriate decisions. This paper focusses on creation of EMR which records accurate patient information based on the refined Ontology.

The standard artificial intelligence methods like Decision Trees and Random Forests, Artificial Neural Networks, Bayesian Networks or Gaussian processes can be used to build a decision support system. Many comparisons between artificial intelligence methods have been made, however none of these comparisons included which method would best search SNOMED CT data [4].

The algorithms deal with the execution of mathematical or statistical models. The system would learn old and new relationships within the SNOMED CT database and apply probabilities to any relationships creating a Bayesian network [4]. Bayesian Belief networks are used to get the desired results as it deals with uncertainties commonly present in clinical practice.

III. SYSTEM ARCHITECTURE

The Fig. 1 shows the overall architecture of the system depicting the complete representation of modules. The data used in the project is obtained from SNOMED CT dataset, an existing medical ontology [5]. The data is preprocessed making it suitable for use in Electronic Medical Record. The information about lung diseases is extracted from dataset and relevant ontology is created.

Visualization of Ontology is done using protégé editor. Protégé is based on Java, is extensible, and provides a plug-and-play environment that makes it a flexible base for rapid prototyping and application development.

The patient information is recorded based on the enriched ontology in EMR. The specific classes which aid in decision making are assigned appropriate probability values thereby forming a Bayesian network. The concepts are semantically matched with patient details in the knowledge base and are given as input to the intelligent agent. The decision support system considers patient information and possible treatments so as to enable the healthcare professionals to take decisions effectively. The system is proposed to provide recommendations to clinicians and patient based on the posed queries that can help in diagnosis and analysis of patient data for improved health care.

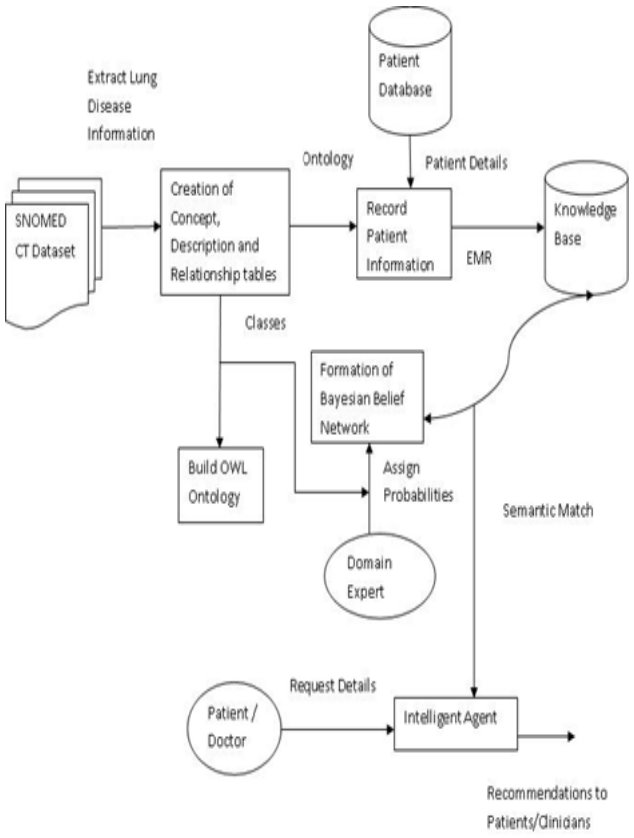

Figure 2. System Architecture

A. System Configuration and Ontology Creation

System Configuration activities include creation of login for care provider and patient roles and necessary SNOMED tables for use in main application. Ontology helps in knowledge acquisition and consistent storage of information. Ontology provides clarity, coherence and models relationships. For ontology to work as a knowledge base, the individual instances of the concepts have to be combined with a set of concepts [8].

The raw data obtained from SNOMED CT contains lots of information some of which may not be useful for the analysis of lung diseases. Data cleaning is done which involves removal of irrelevant and redundant entries, correction of data so as to make it efficient for further analysis and processing. This process not only reduces the file size but also increases the quality of available data.

The OWL Web Ontology Language is used to describe the classes and relations between them. Ontology is a data model representing a set of concepts within a domain and the relationship between these concepts. Ontology development is an iterative process. Classes and class hierarchies are defined using following methods.

- Top-down – Terms of interest
- Bottom-up – Generalization of terms
- Combination of both

The technique used in identification of concepts uses combination of both top down, when defining with the domain experts the concepts to consider and bottom up approach when identifying generalizations for some of the concepts [2].

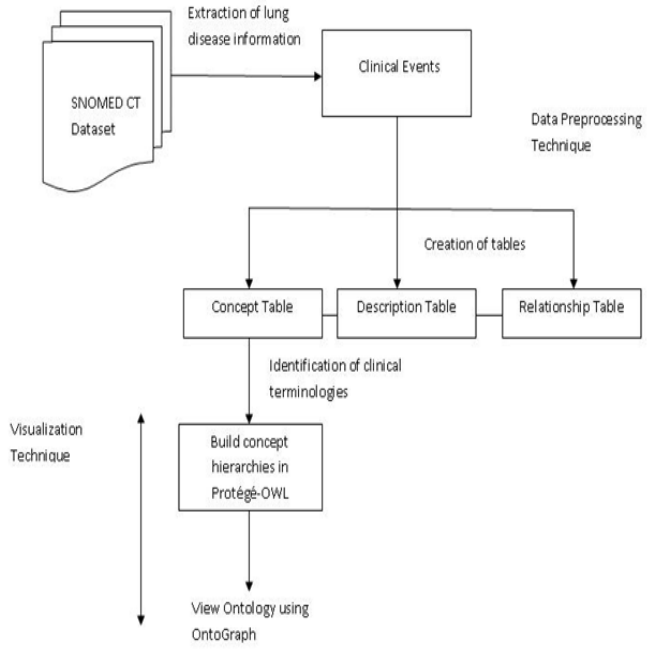

Figure 3. Data Pre-processing and Visualization

Algorithm:

Input: SNOMED CT Dataset [5]

Output: Concept, Description and Relationship tables and visualization of lung diseases in Protégé-OWL.

- Create login for care provider and patient roles.
- Remove unwanted columns from the original SNOMED CT dataset in concept, description and relationship tables.
- Identify concept codes related to lung diseases using a SNOMED browser.
- Clean the data by removing irrelevant and redundant entries.
- Form the enriched SNOMED CT tables using the core concepts.
- Organize the data under main categories namely body structure, procedure list, observable entity, situation, regime/therapy, finding and disorder.
- Construct OntoGraf using protégé tool.
- Create OWL(Web Ontology Language) files.

B. EMR based on Ontology for Decision Making

As the use of electronic records increases, healthcare providers should be expected to make the best possible use of the wealth of computable patient data that systems will contain [3].

Ontology provides unique representation of knowledge. It enables unambiguous recording of data in the knowledge base. It is essential to capture signs and symptoms about lung diseases and store information about tests to be carried out to confirm a particular disease. Signs and symptoms of lung diseases include heart failure, cracking sound while breathing, bluish color in skin and lips, indication of fluid filled in lungs (pneumonia), low oxygen levels in blood [6]. The EMR is created for both new patient and also existing patient.

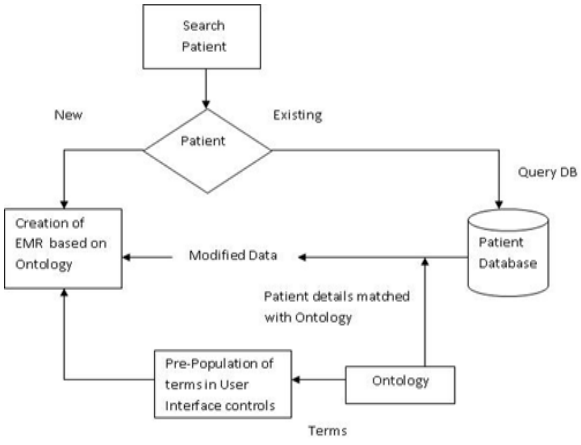

Figure 4. EMR based on Ontology

Algorithm:

Input: Clinical information about Patient

Output: EMR ontology capable of representing knowledge about lung disease

search patient
if found
i, pre-populate SNOMED terms
ii, modify patient details
if not found
i, create a new record
ii, ensure SNOMED-CT data entry

Semantic relationships enable reliability and consistency of data retrieval. It uses cost functions based on health goals and correlate patient information and possible treatments [3].

Two classes of Bayesian networks in supervised learning: Naïve Bayesian networks (NB) and Tree-Augmented Naïve Bayesian networks (TAN) in which TAN is more similar to the current organization of SNOMED CT [4].

A Bayesian network is a probabilistic graphical model that uses a set of random variables to specify a joint probability distribution.

i, assign probability values to specific concepts
ii, construct conditional probability tables for decision making
iii, perform probabilistic analysis
iv, pose a query
if user a care provider
several alternatives according to probability values are listed
else
the highest probability value is given as recommendation to patient
v, refine diagnosis based on the result.

The system incorporates querying and reasoning about alternative courses of action in order to make context management and interpretation easy. The retrieval component provides multi modal representation of query results either textual or graphical.

IV. Experiments and Results

The concepts related to lung diseases are extracted from the SNOMED CT dataset and concept, description and relationship tables are created through a user interface screen. These tables have reduced columns from the original SNOMED CT tables. This is depicted in Table I. The Table II shows the reduced entries after data cleaning process. The sample entries of modified SNOMED CT tables are depicted in Fig. 5.

TABLE I
Table Information

Tables	Fields
Concept	ConceptID, ConceptName
Description	DescriptionID, ConceptID, DescriptionName
Relationship	RelationshipID, ConceptID1, ConceptID2

TABLE II
Data Pre-processing Technique

Table Name	Entries in original dataset	Entries after data cleaning
Concept	393072	2313
Description	1048576	1910
Relationship	1048576	1910

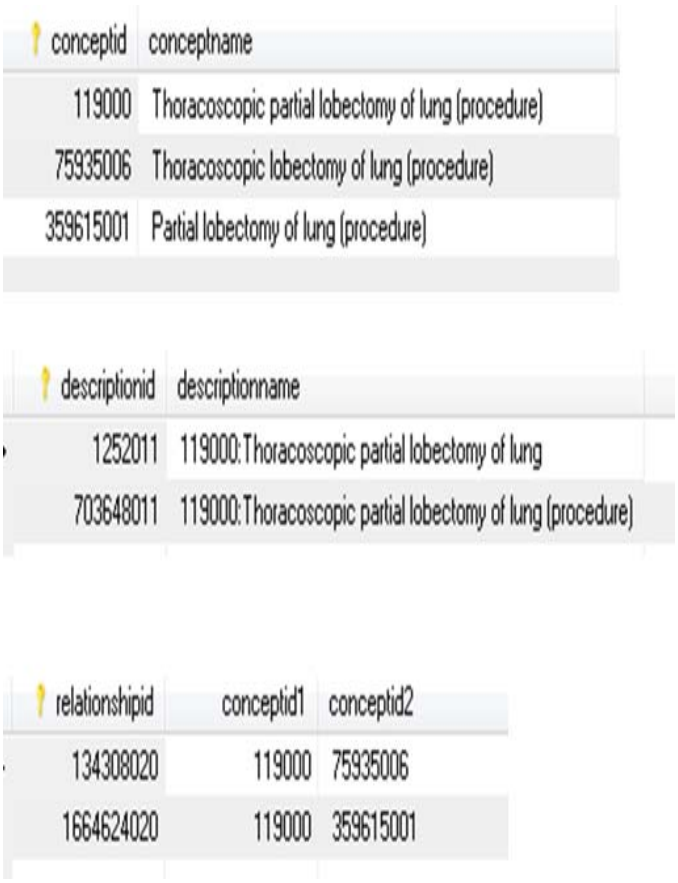

Figure 5. Sample entries in SNOMED Tables

The core concepts are identified and data is classified into different categories such as Finding, Observable Entity,

Situation, Disorder, Procedure, Body Structure and Regime/Therapy. The OWL (Web Ontology Language) file with different disease categories is created and visualized in Protégé editor as shown in Fig. 6. This helps in semantic matching of query terms and concepts stored in knowledge base thereby making decision process easy. The information about lung diseases is also visualized in Protégé which is shown in Fig. 7.

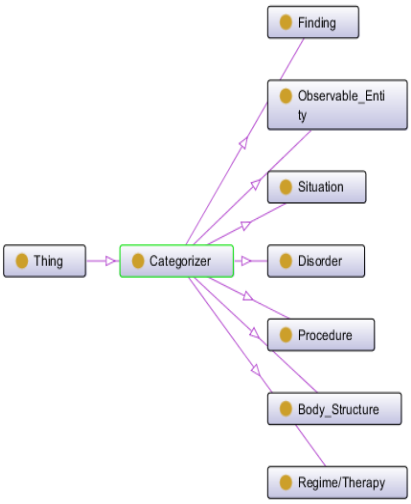

Figure 6. Concept Categorizer

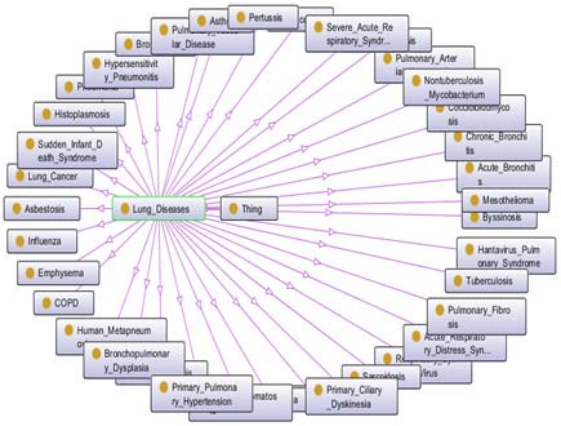

Figure 7. Concept Categorizer

The patient details include demographics and disease information and these are stored in the patient database. The electronic medical record created for a patient which helps in maintaining consistency of patient data which helps in semantic interoperability. The disease information for a patient is recorded by using SNOMED CT concept codes as shown in Fig. 8. There is a possibility of storing multiple disease entries for a single patient. This can be searched and modified as depicted in Fig. 9.

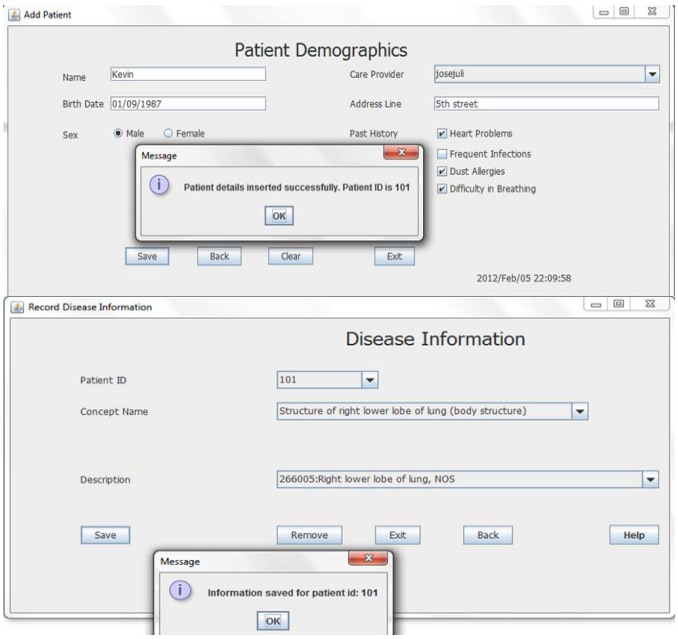

Figure 8. Electronic Medical Record

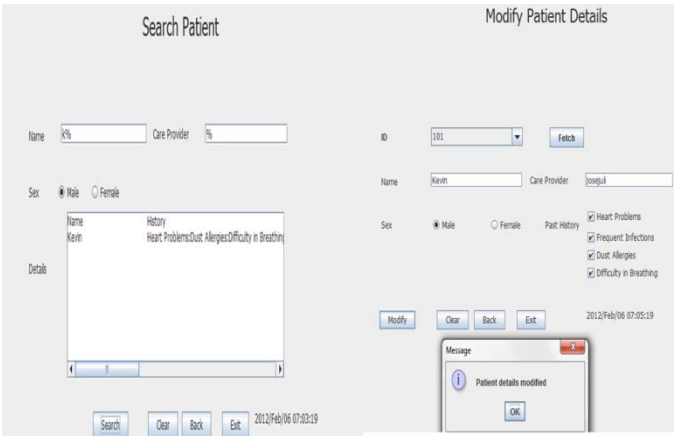

Figure 9. EMR Refinement

V. Conclusions and Future Work

The overall system architecture is designed which depicts data collection and preprocessing from SNOMED CT dataset, visualization of classes in Protégé editor, creation of EMR and decision making using Bayesian approach. In this paper we have proposed Ontology construction from SNOMED CT dataset with visualization in protégé and creation and refinement of electronic medical record suitable for decision making. The future work includes formation of probability tables and Bayesian Belief Network and creation of query and retrieval components.

REFERENCES

[1] Ashraf Mohammed Iqbal, Michael Sheperd and Syed Sibte Raza Abidi (2011) 'An Ontology-Based Electronic Medical Record for Chronic Disease Management'. 44th Hawaii International Conference on System Sciences.

[2] Caita M Machado, Francisco Couto, Alexandra R Fernandes, Susana Santos, Nuno Cardim and Ana T Freitas (2010) 'Semantic Characterization of Hypertrophic Cardiomyopathy Disease'. IEEE International Conference on Bioinformatics and Biomedicine Workshops.

[3] David E. Robbins, Varadaraj P. Gurupur, John Tanik (2011) 'Information Architecture of a Clinical Decision Support System'. Proceedings of IEEE.

[4] Ewelina Ciolko, BHSc, Fletcher Lu (2010) 'Intelligent Clinical Decision Support Systems Based on SNOMED CT'. 32nd Annual International Conference of the IEEE EMBS.

[5] http://www.nlm.nih.gov/research/umls/licensedcontent/snomedctfiles.html

[6] http://www.nhlbi.nih.gov/health/public/lung/index.htm

[7] http://www.w3.org/TR/owl-features/

[8] Hyun-Je Song, Seong-Bae Park, Se-Young Park. An Automatic Ontology Population with a Machine Learning Technique from Semi-Structured Documents. Proceedings of the 2009 IEEE International Conference on Information and Automation June 22-25, 2009, Zhuhai/Macau, China

[9] John Patrick, Min Li (2009) 'Intelligent Clinical Notes System: An Information Retrieval and Information Extraction System for Clinical Notes'. Healthcom 2009.

[10] Nicola Guarino, Pierdaniele Giaretta. Ontologies and Knowledge Bases, Towards a Terminological Clarification.

[11] Prathima D, Koushike Parvathy Devi B (2011) 'Computing Technology Used in Retrieval of Patient's Record'. Proceedings of the National Conference on Innovations in Emerging Technology.

[12] Shuli Yuwen, Xiaoping Yang (2010) 'Research on the Clinical Terminology Construction based on SNOMED'. Seventh International Conference on Fuzzy Systems and Knowledge Discovery.