An AI Approach to Transform the E-Patient Records into Clinical Cases for Medical Diagnostic Systems

Mohammed F. Audi, Rania A. Hodhod, Abdel-Badeeh M. Salem
Computer Science Department, Faculty of Computer and Information Sciences
Ain Shams University
Cairo, Egypt
Mo.fekry@gmail.com

Abstract - **A major problem for researchers in medical informatics is the presence of 'useful' Electronic Patient Record (EPR) and its use in providing important information to and about the patient's diseases and treatment. This paper discusses the importance of medical informatics and its role in health procurement branch. It also proposes an AI approach to transform the E-Patient records into clinical cases for medical diagnostic systems. This approach is based on using medical documents to generate E-Patient records, then mapping these records to build medical cases to form a cases repository that would be utilized in a CBR-medical diagnostic system.**

Keywords- MedicalInformatics ; Electronic Patient Records ; Medical Clinical Caes; Case Based Reasoning; Knowledge Acquisition ; XML

I. INTRODUCTION

Medical informatics is a special field that focuses on the use of computer technology to help, organize, analyze, manage and use information in improving health care. It aims to enhancing the usability of health information systems to facilitate communication and manage flow of health information from physicians to patients and vice versa. Health information for and about patients are collected as Electronic Patient Records (EPR) [7].

EPR can be defined as an organized collection of electronic health information about patients and populations. It may include data in comprehensive or summary form, including medical history medication and allergies, radiology images and personal stats like age and weight information. The presence of an electronic format allows the EPR to be shared across different health care sites, and to be embedded in network-connected enterprise-wide information systems like medical case based reasoning systems (CBR) [5].

CBR-based medical diagnostic system requires up-to-date diagnostic-quality operable clinical cases (OCCs). The availability of high-quality OCC is the bottleneck towards the incorporation of medical CBR systems in any real life medical diagnostic environment.

Preparing CBR clinical cases is a great challenge; it requires medical experts to map their knowledge and experience to an unfamiliar computational formalism. Alternatively, generating these CBR-compliant cases from electronic patient records (EPRs) is a good solution that can be achieved by applying an artificial medical knowledge acquisition approach.

The availability of XML-medical documents on the internet gives us a magic solution to build electronic patient records [9]. Getting proper medical documents is a challenge in itself. However, the use of automated medical concept recognition could serve as a solution for medical documents retrieval and text mining research [2]. Different approaches were developed in order to do this. For example, depending on generating conceptual terms by permuting words in a sentence, an approach was developed using phrase-based indexing to represent similar concepts, and query expansion to improve matching query terms with the terms in the document [10]. Another approach was followed in which two main techniques were utilized. The first technique is Key phrase Identification program (KIP) used for identifying topical concepts from medical documents (i.e. Noun Phrase & Key Phrase Extractor systems). KIP combines two functions: noun phrase extraction and key phrase identification [2]. The second technique is based on Semantic Similarity Retrieval Model (SSRM) which correlates documents containing semantically similar terms. MEDSEARCH is a complete retrieval system for medical informatics implemented to retrieve medical documents from USA midline library [3].

The work done considers the presence of well formatted XML-based EPR to extract the needed patient records information. To the extent of our knowledge there is no single system that attempted to do both the preparation stage of these patients' records through retrieving the medical documents, transfer them to XML-formatted EPR and then extract the needed patient records information to be utilized in a medical diagnostic system. This paper proposes a single approach that attempt the above three stages; extraction, transformation and utilization of EPR.

II. ELECTRONIC PATIENT RECORDS

The Electronic Patient Record (EPR) is a file kept on a computer containing information about the patient's health. Previously, patient records were kept as hard copies in physical files. The movement of physical files towards the electronic forms allows physicians to query, transfer and handle patients' information in an easy way. An electronic record is created for each service a patient receives from clinical departments, such as radiology, laboratory, or pharmacy, or as a result of administrative action (e.g., creating a claim). Storing and transferring patient information

electronically reduces clinical errors and improves patient safety as well as allowing clinicians to communicate more quickly and accurately and identifying relevant information in an easy way [5].

EPR can be viewed as a clarification of the physician's problem-solving strategy as it contains a problem situation and its physician solution (action). Information contained can be divided into three main parts:

- *Problem situation*: the state of problem description documented by the physician.
- *Solution*: the physician solution given as diagnosis and treatment.
- *Outcome*: the state resulted when solution is applied on the problem stated before.

From the above EPR properties, EPR is seen as quite similar to the medical case contents. Accordingly, EPR can be viewed as an abstraction of a clinical case 'Problem Solving' knowledge system. Table 1 shows the similarity of EPR to a clinical case.

As a result of above similarity, EPR can act as a magic solution for making clinical cases that are required for medical case based reasoning systems (CBR). The transformation process of EPR to Clinical Case (CC) is based on establishing a conceptual equivalence at both the structure and contents levels. This process consists of two main phases;

- Structural equivalence that attempts to make structural mapping such as determining for each CC attribute a conceptual equivalent EPR attribute.
- Content equivalence that attempts to make content (value) mapping such as unifying the value of each mapped EPR attribute with respect to vocabulary of corresponding CC attribute. [9]

TABLE I. AN ELECTRONIC PATIENT RECORD VERSUS A CLINICAL CASE

EPR Contents		Medical Case Content
Problem	-Personal information -Admission Information -Nursing Documentation - Symptoms - patient History	- Age - Gender - Vital signs -Temperature, pulse, Blood pressure - Report
Solution	Clinician requests Progress notes	Diagnosis Procedure
Out-come	Physician examination Diagnosis reports	Assessment Progress status

III. MEDICAL CASE BASED REASONING

Case-based reasoning (CBR) is an analogical reasoning method that provides a methodology problem solving and decision-support problems [7]. It is consistent with much that psychologist have observed in the natural problem solving that people do. CBR problem solving technique includes planning, diagnosis and design. It refers to a number of concepts and techniques that can be used to record and index cases and then search them to identify the ones that might be useful in solving new cases when they are presented. In addition, there are techniques that can be used to modify earlier cases to better match new cases and other techniques to synthesize new cases when they are needed.

The diagram in Fig. 1 can be summarized in the following processes:

- *Assign index*: each input case is assigned an index characterizes its properties
- *Retrieve*: retrieving the most similar case that matches the input case
- *Modify*: modify the retrieved case's solution to adapt the current case
- *Test*: apply the retrieved modified solution to the current case and check if it solves the current problem
- *Assign and Store*: if the solution succeeds, then assign an index to it and store it in the case base
- *Explain Repair and Test*: if the solution fails, then explain the failure and trying to repair the solution to be tested again. [6]

So, the challenge lies in the existence of appropriate clinical cases in the case base to be searched against the input case. In practical terms, finding medical clinical case is quite a challenge. To overcome this problem, medical clinical cases had to be prepared first. The procurement of medical clinical case is a big challenge, as this requires medical experts to map their experiential knowledge to an unfamiliar computational formalism [1]. Instead, by making use of the existence of EPR over the internet, medical clinical cases could be generated.

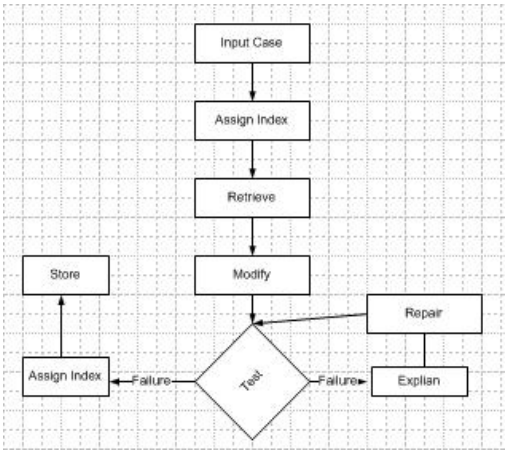

Figure 1: Case based reasoning phases

So the challenge now is to get the appropriate medical documents to generate the needed EPR in order to generate the required medical clinical cases. [4].

IV. Medical Document Retrieval

Automated medical concept recognition is important for medical informatics such as medical document retrieval and text mining. The need of retrieving specific medical documents increased because of the unavailability of ready electronic patient record. Not only the medical document retrieval is important but also the identification of topical concepts from retrieved medical documents to build electronic medical records. Different approaches have been developed to handle the medical document retrieval and the knowledge identification [2].

Based on generating conceptual terms by permuting words in a sentence, an approach has been developed using phrase-based indexing to represent similar concepts, and query expansion to improve matching query terms with the terms in the document. This approach consists of three main components: *Index Finder*, *Phrase-based VSM* and *Knowledge-based query expansion*. The Index finder extracts terms from free-text, and generates conceptual terms by permuting words in a sentence. In Phrase-based VSM, each document is divided into a set of phrases, each phrase is represented by a concept defined in the controlled vocabulary and the corresponding word stems. Knowledge-based query expansion expands terms related to the specified scenario and provides answers and retrieved content correlated medical documents. But this approach has a problem with semantics [10].

Another approach is the Key Phrase Identification Program (KIP) that aims to identifying topical concepts from medical documents. It combines two functions: Noun Phrase Extraction (NPE) and Key Phrase Identification (KPI). NPE extracts automatically noun phrases from medical literature. KPI assigns weights to extracted noun phrases for a medical document based on how important they are to that document and how domain specific they are in the medical domain. The experimental results show that noun phrase extractor is effective in identifying noun phrases from medical documents, so is the key phrase extractor in identifying important medical conceptual terms. But this approach suffers from problems related to semantics [2]

Semantic Similarity Retrieval Model (SSRM) is another model that overcomes the problem of semantics; it has the ability of associating documents containing semantically similar terms. SSRM suggests discovering semantically similar terms in documents and queries using term taxonomies (ontologies) and by associating such terms using semantic similarity methods. [3]

V. Proposed Approach

This study aims to shed light on the importance of medical records and the need to improve the medical services provided to the patient. The proposed implementation is a CBR-based medical diagnostic system that supports the diagnosis and treatment of diseases, for example in the heart domain. It will apply the evidence-based medicine approach, which is concerned with ensuring that strategies of proven clinical effectiveness are adopted by the system. It will be based on the mechanism of case based reasoning in dealing with input cases to find the most similar case from its case-base and adapting it in order to apply the proper solution (treatment) to it. A basic step in building such medical CBR system is building the case repository with clinical medical cases. Unfortunately it is so difficult to find a repository of medical clinical case. To overcome this, the proposed solution is to make use of the various repositories of electronic patient records over the internet. As not all electronic medical documents will be useful to the proposed conversion mechanism, a knowledge acquisition technique will be used to retrieve the desired medical document only.

The proposed methodology shown in Fig. 2 works as follows:

A. Using Semantic Similarity Retrieval Model (SSRM) to get the proper needed medical documents in order to use them in getting the electronic patient records. SSRM discovers documents terms that are semantically similar and based on term taxonomies (ontologies), it performs the querying process by associating such terms using semantic similarity methods. SSRM will act as follows:

- Reduce long plain English sentences to specific words.
- Determine all the terms to be included.
- Determine which search feature should be applied (i.e. AND, OR)
- Include relevant information to the search query
- Evaluate the produced search result.

B. Proposing a methodology to convert the electronic patient records EPR to medical clinical cases as follows:

- *Name matching*: direct mapping by finding similar EPR attributes to medical case attributes.
- *Structural equivalence*: it will be for non-mapped above attributes. For each attribute in the medical case, there will be a list of equivalent synonyms. Those synonyms will be used in establishing a synonyms matching with EPR attribute.
- *Content equivalence:* By defining vocabulary and numeric range for each medical case attribute-value then unifying the EPR values with respect to the pre-defined medical case contents

C. Indexing the generated medical clinical cases.

- Case indexing involves assigning indexes to cases to facilitate their retrieval. The following methods are the most famous indexing methods such as (1) Paralleled

indexing, (2) Bitwise indexing, (3) similarity indexing, (4) cluster indexing.

D. Building the case-base

- Building the case-base is very critical task. One of the famous techniques that are used is the Genetic Algorithm (GA). It based on feature selection and weighting.

E. Applying the case base in a CBR-based medical diagnostic system. The CBR-based medical system will act as follows:

- A new patient case will be accepted
- Retrieving the most similar problems that match the input case
- Adapting the retrieved solution to match the input case
- Testing the adapted retrieved solution and see if it solves the current case
- If the solution succeeds, the current input case with the tested solution will be indexed and saved in the case-base to be used in future.
- If the solution fails, then it will be reviewed, retained, and tested again.

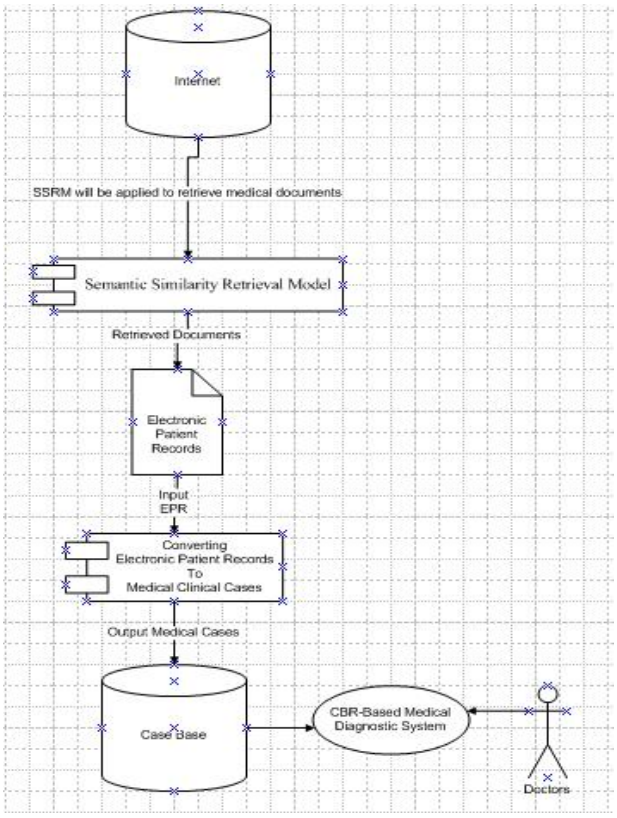

Figure 2: Proposed methodology

VI. EVALUATION

This research aims to clarify the meaning of medical records to the patient and the extent of the difficulties faced by doctors in the use of such records to improve medical services. In the past, patient's medical records were treated as archived records only. However, the importance of storing and manipulating these records electronically has been manifested in the past few years. Also we illustrated on the importance of medical applications and highlighted the need of well formatted medical clinical cases, which can be considered a big challenge.

Generating the medical clinical cases from available electronic patient records available over the internet is the main target to overcome the problem of finding required/desired medical clinical cases for the CBR medical diagnostic system. We preferred using the SSRM model as a knowledge acquisition approach to get the needed medical document in order to extract the electronic patient records as it overcomes the problem of word semantics that exists in other models.

Another challenge is to find an appropriate indexing method that would serve the storing and retrieving processes efficiently.

VII. CONCLUSION

This paper presented an AI approach that aims to generate medical clinical cases to be utilized by a CBR medical diagnostic system. This approach consists of five main steps. (1) Gathering the needed medical documents to generate EPR by using Semantic Similarity Retrieval Model (SSRM). Then these documents will be used to get the electronic patient records. (2) Converting the electronic patient records EPR to medical clinical cases by using a proposed methodology to map each EPR attribute to the appropriate medical case attribute. (3) Indexing the generated medical clinical cases to facilitate their retrieval. (4) Building the case-base which is considered a very critical task. This step will be done by using genetic algorithm(s) (GA). (5) Employ the case base in a CBR-based medical diagnostic system.

The paper also proposed a new approach based on Semantic Similarity Retrieval Model (SSRM) in retrieving medical documents, in which overcomes the semantics problem. It has the ability to associate documents containing semantically similar terms, by suggesting discovered semantically similar terms in documents and queries using term taxonomies (ontologies) and by associating such terms using semantic similarity methods. The EPR-OCC attributes mapping overcomes the problem of non-matched attributes names by defining vocabulary and numeric range for each medical case attribute-value then unifying the EPR values with respect to the pre-defined medical case contents. The approach manages to transfer the medical records to useful EPR that would be transformed to medical cases to build the case repository and then utilized in a CBR-medical system.

References

[1] R.Y. Baeza, B.N. Ribeiro, Modern information retrieval. England : Massachusetts, 2007.

[2] Y. Brook, Identifying important concepts from medical documents, Journal of Biomedical Informatics, 2006.

[3] A. Hliaoutakis, G. Varelas, E. Petrakis, E. Milios, A retrieval system for medical information based on semantic similarity, 2006.

[4] M. Luis, Retrieving medical records using bayesian networks, 2009.

[5] A.L. Rector, W.A. Nolan, S. Kay, Foundations for an electronic medical record, Methods of Information in Medicine, 30, 86-179, 1991.

[6] B. Salem, Case based reasoning technology for medical diagnosis, World Academy of Science, Engineering and Technology 31, 2007.

[7] R. Schmidt, L. Gierl, Medical case-based reasoning systems, Institute for Medical Informatics and Biometry, University of Rostock, Germany, 2001.

[8] C. Shihchieh, An information retrieval system for medical records & documents, 2008.

[9] S. Sibte, Leveraging XML-based electronic medical records to extract experiential clinical knowledge, An automated approach to generate cases for medical case-based reasoning systems, International Journal of Medical Informatics, 2002.

[10] W. Wesley, A knowledge-based approach for retrieving scenario-specific medical text documents, 2005.