The Hong Kong Polytechnic University

Department of Industrial and Systems Engineering

A CLINICAL DECISION SUPPORT SYSTEM FOR MEDICAL PRESCRIPTION PROCESS

TING Siu Lun

A thesis submitted in partial fulfillment of the requirements for

the Degree of Doctor of Philosophy

May 2011

Certification of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

Ting Siu Lun

Abstract

Nowadays enormous amounts of new drugs are developed and launched onto the market. With the increased complexity of drug information and a degree of uncertainty surrounding it, medicine prescription has become a vexing process. This is particularly true from the general practitioners' perspective as they are primarily responsible for providing general health care to patients, and keeping abreast with the latest drug information for treatment of mutating diseases. Many researchers have proposed the development of a decision support system for addressing these issues in recent years. However, the existing approaches lack flexibility to deal with the complex and dynamic changes in drug information. Thus, providing support to the medical prescription process remains a major challenge to be adequately addressed. The main problems are the development of adaptive mechanisms of knowledge acquisition and knowledge modeling in the medical prescription process.

This research proposes a new approach to addressing the above mentioned difficulties. Prescription practices differ from one physician to another. This leads to the existence of a large variety of methods, giving rise to considerable complexity in modeling the physician's decision logic. The proposed approach to modeling physicians' prescription logic takes into account both individual and collective wisdoms of a pool of physicians. It has been implemented in a Medical Prescription Decision Support System (MedicPDSS) that automatically generates drug suggestions after considering the medical information of a specific patient. The approach integrates computational intelligence and data mining techniques to model physicians' prescription behaviors, from which suggested options of safe medical prescriptions are generated. This system employs case-based reasoning to retrieve past prescription records that are related to treatment of the disease under consideration, and the association rules mining technique is then applied to integrate such results. The former step models the practices of individual physicians, and the latter one models their collective judgments. The suggested options of medical prescriptions are generated accordingly along with rankings of their appropriateness. Furthermore, the prescription selected by the physician is checked with data obtained by an automatic information retrieval engine to detect drug-drug interactions.

A prototype system that applies this approach has been built and tested in a medical organization – Humphrey and Partners Medical Services Limited (HPMS). With the implementation of the MedicPDSS in HPMS, a number of potential benefits are realized. The system produces significant improvement in a number of performance criteria, such as reducing the time for deciding on a prescription, minimizing prescription errors, optimizing the list of suggested drugs, and enhancing knowledge sharing of prescription practices. In addition, the physicians and nurses confirmed that MedicPDSS has helped them to improve quality of prescriptions.

Even though the proposed system has been developed for a typical healthcare environment, its concept can be customized for other applications in healthcare environments that deal with specific diseases. It can also be used together with other computer-aided diagnostic systems and clinical decision support systems to further enhance the medical prescription process. Moreover, the system can be employed to support training for medical professionals.

Publications Arising from the Thesis

In the process of conducting the study, **9** international journal papers and **3** conference papers have been published.

Journal Articles

1. **Ting, S.L.**, Shum, C.C., Kwok, S.K., Tsang, A.H.C. and Lee, W.B. (2009), "Data Mining in Biomedicine: Current Applications and Further Directions for Research", *Journal of Software Engineering and Application*, Vol. 2, No.3, pp. 150-159.
2. **Ting, S.L.**, Kwok, S.K., Tsang, A.H.C. and Lee, W.B. (2010), "CASESIAN: A Knowledge-based System Using Statistical and Experiential Perspectives for Improving the Knowledge Sharing in the Medical Prescription Process", *Expert Systems With Applications*, Vol. 37, No. 7, pp. 5336-5346.
3. **Ting, J.S.L.**, Kwok, S.K., Tsang, A.H.C., Lee, W.B. and Yee, K.F. (2010), "Experiences Sharing of Implementing Template-Based Electronic Medical Record System (TEMRS) in a Hong Kong Medical Organization", *Journal of Medical Systems*, Accepted (DOI: 10.1007/s10916-010-9436-9).
4. **Ting, S.L.**, Wang, W.M., Kwok, S.K., Tsang, A.H.C. and Lee, W.B. (2010), "RACER: Rule-Associated CasE-based Reasoning for Supporting General Practitioners in Prescription Making", *Expert Systems With Applications*, Vol. 37, No. 12, pp. 8079-8089.
5. **Ting, S.L.**, Kwok, S.K., Tsang, A.H.C. and Lee, W.B. (2011), "A Hybrid

Knowledge-based Approach to Supporting the Medical Prescription for General Practitioners: Real case in a Hong Kong Medical Center", *Knowledge-Based Systems*, Vol. 24, No. 3, pp. 444-456.

6. **Ting, J.S.L.**, Tsang, A.H.C., Ip, A.W.H. and Ho, G.T.S. (2011), "RF-MediSys: A Radio Frequency Identification-based Electronic Medical Record System for Improving Medical Information Accessibility and Services at Point of Care", *Health Information Management Journal*, Vol. 40, No. 1, pp. 25-32.

7. **Ting, J.S.L.**, Kwok, S.K., Tsang, A.H.C. and Lee, W.B. (2011), "Critical Elements and Lessons Learnt from the Implementation of an RFID-enabled Healthcare Management System in a Medical Organization", *Journal of Medical Systems*, Vol. 35, No. 4, pp. 657-669.

8. **Ting, J.S.L.**, Wang, W.M. Tse, Y.K. and Ip, W.H. (2011), "Knowledge Elicitation Approach in Enhancing Tacit Knowledge Sharing", *Industrial Management & Data System*, Vol. 111, No. 7, Vol. 111, No. 7, pp. 1039-1064.

9. **Ting, S.L.**, Ip, W.H. and Tsang, A.H.C. (2011), "Is Naïve Bayes a Good Classifier for Document Classification?", *International Journal of Software Engineering and Its Applications*, Vol. 5, No. 3, pp. 37-46.

Abstract and Conference Proceedings

1. **Ting, J.S.L.**, Tsang, A.H.C., Kwok, S.K. and Lee, W.B. (2007), "Mobile Electronic

Medication Management for Psychiatric Rehabilitants: A Case Study in a Hong Kong Psychiatric Hospital", *Proceedings of the 5th International Conference on Quality and Reliability (ICQR) 2007*, Chiang Mai, Thailand, 5-7 November 2007, pp. 267-271.

2. **Ting, S.L.**, Kwok, S.K., Tsang, A.H.C. and Lee, W.B. (2009), "Development of a Dynamic Medical Prescription Education System (MEDPRES)", *Proceedings of International Technology, Education and Development Conference (INTED) 2009*, Valencia, Spain, 9-11 March 2009, pp. 2958-2965.

3. **Ting, S.L.**, Kwok, S.K., Tsang, A.H.C. and Lee, W.B. (2010), "An RFID-based Drug Management System: A Case In Medical Organization", *Proceedings of the 1st International Conference on e-Health Services and Technologies (ICEHST) 2010: Part of INNOV, the International Multi-Conference on Innovative Developments in ICT. 2010.*, Athens, Greece, 29-31 July, 2010, pp. 99-107.

Acknowledgments

I would like to express my sincere gratitude to my chief supervisor, Dr W.H. Ip, for the continuous support of my PhD study and research, for his patient guidance, enthusiastic encouragement, useful critiques, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. Furthermore, I would like to thank my co-supervisors, Prof. W.B. Lee and Dr Albert H.C. Tsang, and Dr S.K. Kwok (my ex-supervisor) for their encouragement, insightful comments, and hard questions.

My sincere thanks also go to Dr. Francis K.Y. Liu, Dr. Peter N.B. Lo, Dr. C.W. Lo, Miss Maggie Poon and all the staff of the Humphrey and Partners Medical Services Limited for providing me invaluable field data and information for this research.

The financial assistance extended by the Research Committee of the Hong Kong Polytechnic University, through a research studentship award, is greatly appreciated. Also, I would like to extend my appreciation to my colleagues in the Department of Industrial and Systems Engineering: Dr George Ho, Dr Wang, Miss Meina Cheng, Mr. Gabriel Lee, Mr. Felix Ng, Mr. Ocean Ng, Mr. Burly Tan, and Miss Tina Wang, for their thoughtful discussion and assistance.

Last but not least, I dedicate this thesis to my family who has jovially supported me and my endeavors and whose vision infect shaped me.

Table of Contents

Page

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Descriptions
1. ADE	Adverse Drug Event
2. AI	Artificial Intelligence
3. ANNs	Artificial Neural Networks
4. ARM	Association Rules Mining
5. BN	Bayesian Network
6. BoDs	Board of Directors
7. BS	Billing System
8. CAD	Coronary Artery Disease
9. CAL	Computer-assisted Learning
10. CBR	Case-based Reasoning
11. CDOE	Computer-based Drug Order Entry
12. CDPD	Chronic Disease Prognosis and Diagnosis
13. CP	Conditional Probability
14. DAG	Directed Acyclic Graph
15. DICP	Drug Information Classification Process
16. DIDS	Drug Interaction Detection System
17. DM	Data Mining
18. DMSS	Data Mining Surveillance System
19. DSS	Decision Support System
20. DT	Decision Trees
21. DTS	Diagnostic and Treatment System
22. EMR	Electronic Medical Records

23. EMRS	Electronic Medical Records System
24. ENT	Ear, Nose and Throat
25. GPs	General Practitioners
26. HL7	Health Level Seven
27. HPMS	Humphrey & Partners Medical Services Limited
28. HTML	HyperText Markup Language
29. ICT	Information and Communication Technology
30. IDF	Inverse Document Frequency
31. KBS	Knowledge-based System
32. KEP	Keyword Extraction Process
33. KDD	Knowledge Discovery in Databases
34. KR	Knowledge Repository
35. MedicPDSS	Medical Prescription Decision Support System
36. NB	Naïve Bayes
37. NN	Neural Network
38. NNR	Nearest-neighbor Retrieval
39. PDS	Prescription Dispensing System
40. PMS	Patient Management System
41. RACER	Rule-associated Case-based Reasoning
42. RBR	Rule-based Reasoning
43. RI	Rule Induction
44. RS	Reporting System
45. SQL	Structured Query Language
46. SVM	Support Vector Machines

47. TEMRS	Template-based Electronic Medical Records System
48. TF	Term Frequency
49. TF-IDF	Term Frequency – Inverse Document Frequency
50. U.R.T.I.	Upper Respiratory Tract Infections
51. VMS	Visit Scheduling System
52. WWW	World Wide Web
53. XML	Extensible Markup Language

Chapter 1 Introduction

1.1 Research Background

Prescription is an essential element of medical practice. The aim of prescription is to administer the most appropriate medicines for a particular patient or a population of patients in order to achieve the desired therapeutic results with minimal adverse drug effects and to improve their conditions by given the available clinical information (Galland, 1997). Quaglini et al. (1992) described that the prescription process consists of repetitive cycles that the physicians are capable to: (i) obtain data regarding the state of patient; (ii) interpret these data to make diagnostics hypotheses and therapy for remedy; (iii) evaluate and refine the therapy; (iv) predict the progress; and (v) remove drug-drug interactions in each therapy. In each action within the process, good observation and experiences of the physicians are demanded. In contrast, patient data misinterpretation and insufficient knowledge in medicines are the main reasons for the cause of medication errors.

Furthermore, the decision to prescribe is influenced by many other factors, such as interactions between physician and patient, cost issues, uncertainty of the diagnosis, and tremendous amount of medicine information (Bradley, 1991; Bradley, 1992; Chen and Landefield, 1994; Gill et al., 1995; Denig et al., 1998; Coscelli, 2000; Lundin, 2000; Wazana, 2000). Bradley (1991), and Greenhalgh and Gill (1997) discussed that the act of issuing medicines is the culmination of a complex chain of decisions along with biomedical, historical, psychosocial, and commercial influences. Substantially, an appropriate prescription is difficult to be made and medication errors often occur when there is any improper use of medication.

According to the study conducted by the Institute of Medicine (2006), around 1.5 million people are injured and 7,000 died each year in the United States because of medication errors. On average, every hospital produces at least one medication error every day. Carter (2004) discussed that the most likely prescription mistakes made by physicians are: (i) interactions between the prescribed medicines and the medicines the patient has already taken, or the foods that the patient commonly eats; (ii) lack of the considerations of medicine allergy; (iii) failure to recognize the side effect; and (iv) incorrect dose. In particular, General Practitioners (GPs) need to diagnose and treat a wide range of health conditions and diseases. Most of the patients go to consult GPs instead of specialists during their first visit. In other words, GPs must be knowledgeable in interpreting patients' conditions as well as deciding which kind of treatments should be conducted (i.e. either prescribing medicines or referring the patient to other health professionals). Such services increases the challenge of GPs to provide effective treatment, especially in cases where they are not familiar with.

Attempts to respond to these issues, Oren et al. (2003) investigated that technology-based intervention plays an important role in avoiding medication errors and improving patient safety. Decision Support System (DSS) has been proposed as one of the most effective ways of medication errors reduction, since it integrates both knowledge-based and expert-based concepts to support GPs in selecting and deciding appropriate medicines to cure the patient (Garg et al., 2005). DSS is a computerized system which provides an interactive and user-friendly interface. It makes use of historic patient data and elements of relevant medical knowledge (such as the information provided in biomedical literatures) to reach the required conclusion. However, the existing approach lacks flexibility to manage the complex and dynamic changes in drug information, providing support for the medical prescription process remains a challenging and difficult task. The problems are the development of adaptive mechanisms of knowledge acquisition, and modeling in the medical prescription process. Consequently, the major issues are how and when to capture the necessary knowledge from physicians during the prescription process.

Expert knowledge assists physicians in formulating better medical judgements and making proper decisions, such as identifying the symptoms to be observed on the patient, and the medicines to be prescribed. However, such knowledge is often difficult to capture for sharing purposes. Typically, the knowledge of a healthcare organization is possessed by a few experts who acquired such knowledge through their medical practices. Physicians in the same specialty may use different knowledge and approaches to treat patients. Furthermore, there is no unique way to determine the optimal regime of medical treatments that applies to a specific patient. All the medical decisions are based on hypotheses, each associated with a probabilistic estimate determined from the physician's experience and knowledge in available options for treatment (Pizzi, 2009). In addition, most knowledge held by experts is tacit knowledge accumulated from experience; it is subjective and personal, making it more difficult to extract when compared to explicit knowledge (Ford and Sterman, 1998). Even when knowledge has been successfully elicited, much effort may still be needed to transform it into explicit form to gain an in-depth understanding of the knowledge.

Unlike other medical domains (such as cancer diagnosing), the conclusion of prescription is complex that consists of a number of medicines. Comparing to diagnosis, which always considers only two classes (e.g. either positive or negative) or multiple classes (e.g. one disease out of different diseases), each medicine out of hundreds of medicines can be a part of the solution in prescription making. However, the multiple values solution deduced a difficult in modeling the prescription behavior of physicians in such domain.

1.2 Problem Statements

Medical prescription is an important task for physicians. In order to provide a complete therapy to patient, an accurate diagnosis is not enough; selection of appropriate medication is also an important factor. According to the study of Riou et al. (1999), prescriptions rely on the expertise of the physicians and at best on careful consideration of the increasing number of drugs, and also with attention to new laws and rules about prescription. Moreover, a modern physician is expected to justify drug selection and know how potential beneficial effects are brought about (Van Hyfte et al., 2001). All of these are knowledge intensive tasks, which in other words depend heavily on the know-how and experience of the physicians. Even though numerous new perspectives on medical prescription are available publicly, physicians may not be able to obtain all of this ongoing knowledge. So if a physician practices drug therapy only by their dated experience and knowledge, he/she cannot provide the standard of care required (Anis et al., 1996; Laurence et al, 1996).

Numerous methods have been investigated for improving the knowledge acquisition and modeling process in medical prescription decision support (Wickramasinghe et al., 2005). In essence, this process is mostly represented in the form of research articles, forum discussions and clinical guidelines. Jabr (2007) argues that this kind of knowledge-sharing process is not well constructed and that problems are still mounting. One challenge for physicians is the limited time they have available for acquiring the relevant knowledge because of the demanding nature of their work and the speed and quality of the transfer process. This acknowledges that there is a pressing and burning need to develop a new approach to facilitate time-efficient, effective knowledge sharing and information exchange for medical prescription.

As a backdrop to the above mentioned considerations, several researchers have utilized knowledge-based system, collective intelligence methods, and data mining techniques to provide computer-assistance and decision support for medical prescription (Spenceley et al., 1997; Warren et al., 1998; Riou et al, 1999; Susan and Warren, 2000; Van Hyfte et al., 2001). These “intelligent and expert” systems provide the same principal contribution: flexibility and the potential for the reuse of knowledge. This is mainly due to the declarative nature of the knowledge embodied in such systems. However, the prescription of medicine is a complicated process, usually complicated with a large amount of information (e.g. new drugs for facing the mutating diseases). The existing systems are not capable of modeling the relationships between the variables, which are between the diagnosis, symptoms, history of patients and drug information. Generally, physicians have developed their own prescription style and behaviors based on their knowledge and experience. In this situation, the problem solving is presented in a single looping process that generates a solution prescribed by the physician himself/herself previously. As a means of knowledge acquisition and modeling, this approach is not suitable because physicians do not share what they know with peers. Even though each physician has the knowledge to make the prescription, it is important for them to learn from others' experiences as well. Thus, an external method is required to enhance the sharing process between physicians, thereby supporting the peer-based comparison determined in collective intelligence perspectives.

In addition, most research on designing an intelligent medical prescription system lack a way of automatically and autonomously capturing the up-to-date information about drug contra-indication. As a result, the current computer-aided medical prescription system may not come up with the most appropriate prescription. According to the above survey, there remain four major problems to be solved. They are summarized as follows:

(i) To design a reliable mechanism to acquire the tacit knowledge of physicians and model their prescription behavior so as to facilitate the selection of medicines;

(ii) To acquire physician's medical therapy decision using an electronic means, such as Electronic Medical Records System (EMRS);

(iii) To identify the drug contraindication of possible prescription solutions by taking into consideration the updated medicine report available in the literature; and

(iv) To create a knowledge access environment so that physicians obtain the right knowledge that is used for decision making on the diagnostic situation.

1.3 Research Objectives and Scope

Medical errors in prescription are globalized disasters over the past decade. One of the major areas of medical error is the improper administration of medication. Every day numerous patients suffer and even die of different types of adverse drug reactions; therefore health care professionals are responsible for applying the 'five rights' of medication administration as a standard of care - the right drug, the right dose, the right time, the right route and the right patient (Chan, 2006). To escort this rule into medical organizations, a Medical Prescription Decision Support System (MedicPDSS) has been developed and introduced to facilitate the drug safety during the drug selection process. The research objectives can be summarized as follows:

(i) To study the existing methods and technologies employed in medical prescription decision support;

(ii) To develop a knowledge acquisition and modeling method for managing physicians' prescription behaviors and automatically constructing a list of medication for prescription;

(iii) To develop computational intelligence algorithms for the implementation of the MedicPDSS and validate the performance of the algorithms by a series of experiments;

and

(iv) To develop a prototype of the MedicPDSS and evaluate the performance of the MedicPDSS through a trial implementation at a selected reference site.

The benefits and significance of developing and implementing the MedicPDSS in medical organization include:

(i) Pinpointing appropriate drug selection for a given disease;

(ii) Reducing the errors occurring in medical prescription;

(iii) Retaining prescription information and knowledge;

(iv) Discovering the latest drug contraindication information automatically;

(v) Acting as a training program for medical organizations and medical students;

(vi) Focusing on and responding to complex medication markets accurately; and

(vii) Recognizing the relationship between patients, diseases and medication.

MedicPDSS is defined as a list of computer modules, algorithms and systems interacting with users in the organization. They aim to facilitate the enhancement of the quality of the medical prescription by modeling the collective prescription wisdom and discovering the drug-drug interaction from trusted and secure sources available to the public (such as the published literature). Furthermore, the current research study is focused on tackling the general diseases, such as upper respiratory tract infections, bronchitis, and gastroenteritis, faced by the GPs

1.4 Research Methodology

The methodology of the research framework follows the conventional procedure for the development of a decision support system (Uzoka, 2009). The research procedures are shown in Figure 1.1. Based on the above research motivation and objectives, this study first involves review of existing literature in the area of medical prescription knowledge, factors affecting medical prescription decisions, and procedures of prescription determination. Research is further conducted on the possibility and plausibility of development of a DSS for the medical prescription process. A literature review is the bedrock of a study that can integrate different opinions, criticize previous scholarly works, build bridges between related topic areas, and/or identify the central issues in a field (Naoum, 1998; Fellows and Liu, 2003). The research further reviews the concepts underlying the development of a DSS. As stated by Naoum (1998), this step is important to obtain insights into the performance of DSS and the ability to support human expertise in specific problem domains. After reviewing the existing literature, the fundamental and basic concept of DSS is identified in the conceptualization stage to guide in the determination of the various elements, theories, technologies, and DSS techniques for the recognized problem domain (see Chapter 2). The primary objectives of these preliminary works attempts to:

(i) Emphasize the significance of DSS for supporting the physician's prescription determination through introducing the current situation of prescription process;

(ii) Discuss the importance of developing DSS culture for the improvement of mediation errors, such as reducing the occurrence of drug-drug interactions;

(iii) Explore the DSS technologies and practice to identify effective solutions to develop a DSS for prescription decision support.

DSS is all about storing, codifying and modeling the human expertise and knowledge into processing rules and logic for further problem solving. Thus, knowledge acquisition is the heart of DSS (Gebus and Leiviska, 2009). After the problem domain was determined in the previous phase, the knowledge and information based of the problem was then acquired. The analysis of the knowledge acquired in the knowledge acquisition phase led to the design of the system which includes the overall architecture of the system, the programming part as well as the interfaces. The system design integrates technologies from the fields of computer science, software engineering, knowledge engineering, and module-based system architecture, that based on the insight derived from the previous phase on the best approach for representing the specific expert knowledge and support capabilities in the DSS. In particular, the method used in the inference engine for processing the system knowledge is also identified in this phase. In this study, an automatic knowledge elicitation approach is developed to acquire the tacit prescription knowledge, whereas a hybrid approach (that modified from the case-based reasoning and association rules mining) is developed to model the captioned prescription knowledge and process the knowledge for decision support. Further details of the proposed methodological framework are explained in Chapter 3.

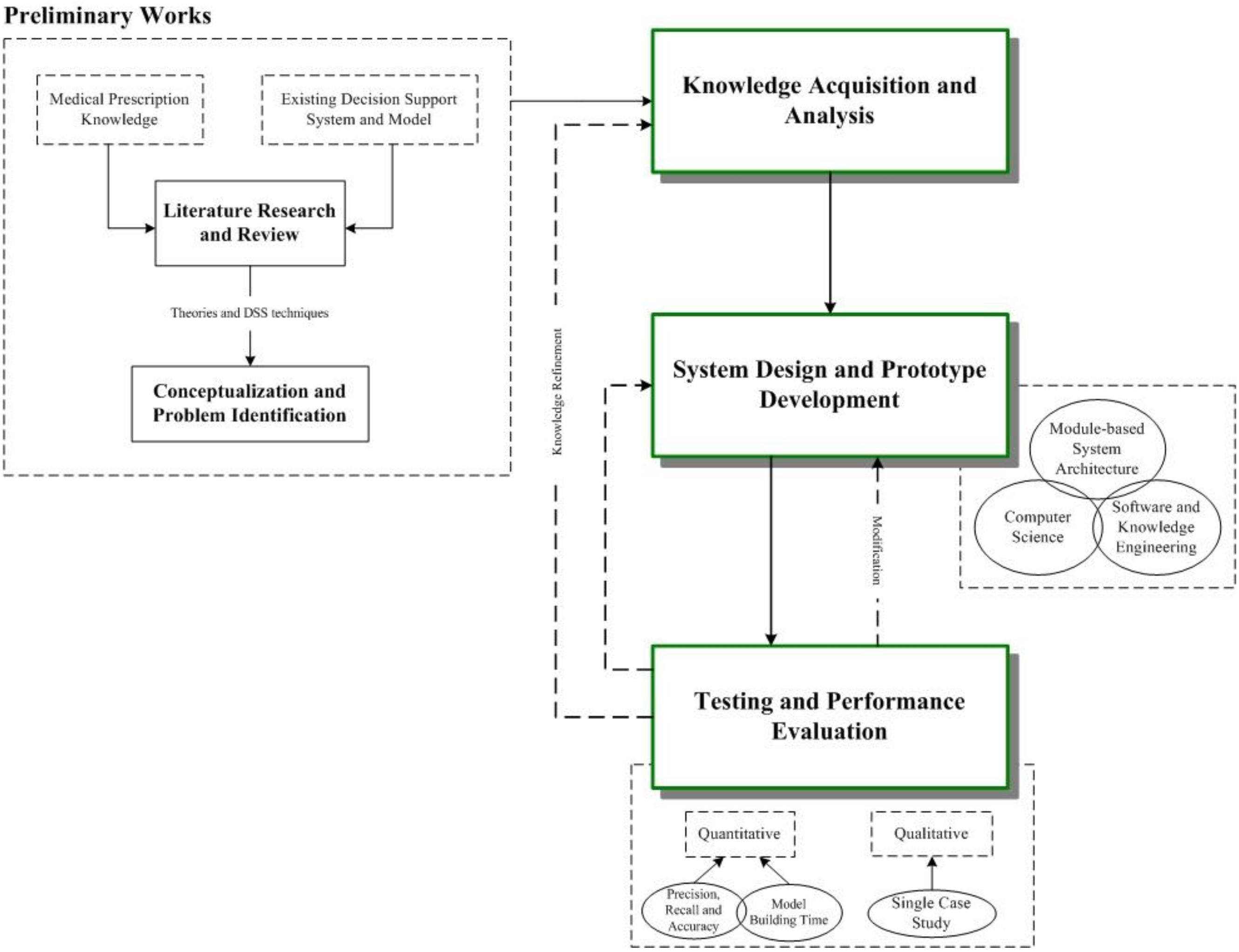


Figure 1.1 – Research Procedure

With all the elements being designed, a prototype is developed to represent the final product of the proposed approach. Various functions are built based on the insights and understandings of the problems derived and system's requirements. System testing is then conducted to ensure the integration of every part of the DSS being worked properly. The system evaluation is done by both quantitative and qualitative approaches. The quantitative measure included precision, recall, time required for building the model, and accuracy of the prescription selection, which are the common measurement for the information retrieval (Ellis, 1996; Tague-Sutcliffe, 1996). On the other hand, the qualitative measure, in form of a case study, addressed the feasibility of system adoption in real situation in form of the criteria of frequency of use, system user-friendliness, performance, and maintenance. As discussed by Naoum (1998), a case study is an in-depth data analysis approach focusing on a set of decisions for the specific problem presented. According to Yin (2009), when forming a single case research design, it is very important to choose a unique and representative case on this specific topic. In this study, one single case (i.e. a Hong Kong medical organization which has not a DSS) is used to study the prescription process. Although Yin (2009) argues that multiple case study approaches are preferred over single case designs in order to produce robust results, it is appropriate for this study to use single case. As one of the focus areas of this study is to acquire the prescription behavior of the physicians and hence model the collected pattern to serve as decision support approach, the case company (which is a medical organization with four clinics located in different areas of Hong Kong) has a number of physicians and there are at least forty visits per day; is suitable for use as a reference site. Therefore, using a single case is a time and cost saving method in this research since the study only focuses on modeling the prescription decision for the general diseases. The results can also provide a strong indication that adoption of DSS in the medical prescription would be useful in similar business contexts. Further details of the case study are explained in Chapter 4 and Chapter 5.

1.5 Thesis Outline

This thesis is composed of seven chapters. Chapter 1 provides the background and motivation of this research. An academic review which identifies the level of understanding and degree of application of knowledge acquisition and modeling in medical prescription decision support are discussed in Chapter 2. Furthermore, the existing practices in medical prescription, and technologies and techniques for such decision support are also discussed.

In Chapter 3, the research methodology for the MedicPDSS is presented. Hence, the theoretical basis and concept of MedicPDSS development are discussed afterwards. The MedicPDSS consists of an EMRS with specific templates for transforming the tacit knowledge of physicians into explicit forms. The template-based EMRS and the three computational intelligence algorithms (the automatic knowledge elicitation, the rule-associated case-based reasoning, and the drug information extraction) that support the

MedicPDSS are described in details in this chapter.

Chapter 4 and Chapter 5 focus on the case study of the proposed methodology and the experimental evaluation of the intelligence algorithms, respectively. Through a prototype implementation of MedicPDSS in a selected reference site, the feasibility and practicability of the system in a real-life environment have been studied in Chapter 4. The results of the system implementation in the case company and the performance of the computational intelligent algorithms for the MedicPDSS are discussed in Chapter 5. Finally, a conclusion is drawn in Chapter 6 to summarize this research study and some suggestions for further study are also discussed in this chapter.

Chapter 2 Literature Review

2.1 Introduction

This chapter provides a review of literature relevant to the study. The basic concept of medical prescription is discussed in the beginning of this chapter. A review on the relationship of medical practices among physicians, nurses, and patients, and related studies on medical prescription knowledge acquisition is discussed in the section devoted to overview of medical prescription. Then, an academic study of information systems for medical prescription knowledge acquisition and modeling is provided. After that, Electronic Medical Records System (EMRS) and Computer-assisted Decision Support Systems (DSS) for medical prescription are discussed. Then a literature review is conducted on existing Artificial Intelligence (AI) and Data Mining (DM) techniques for medical prescription knowledge acquisition and modeling.

The author classifies the existing techniques for knowledge acquisition and modeling into two domains. One is the experience-based approach that utilizes the previously experienced and concrete problem situation to construct the modeling algorithms; whereas another one is the collective-based approach that learns from all the possible solution for generating an aggregated solution. Related techniques applied in medical domain are reviewed. Furthermore, the limitations of these approaches are also discussed so as to highlight the research opportunity.

2.2 Overview of Medical Prescription

A medical prescription is a medication order form written by a qualified medical professional (Riou et al., 1999). It is noted that a series of procedures (such as registration, physical examination, drug dispensing) are carried out when a patient visits the physicians. Amongst these procedures, prescription is the most critical one since it can affect patients' life. The prescription process, as shown in Figure 2.1, involves deciding what therapies are appropriate for patients after diagnosis made physicians.

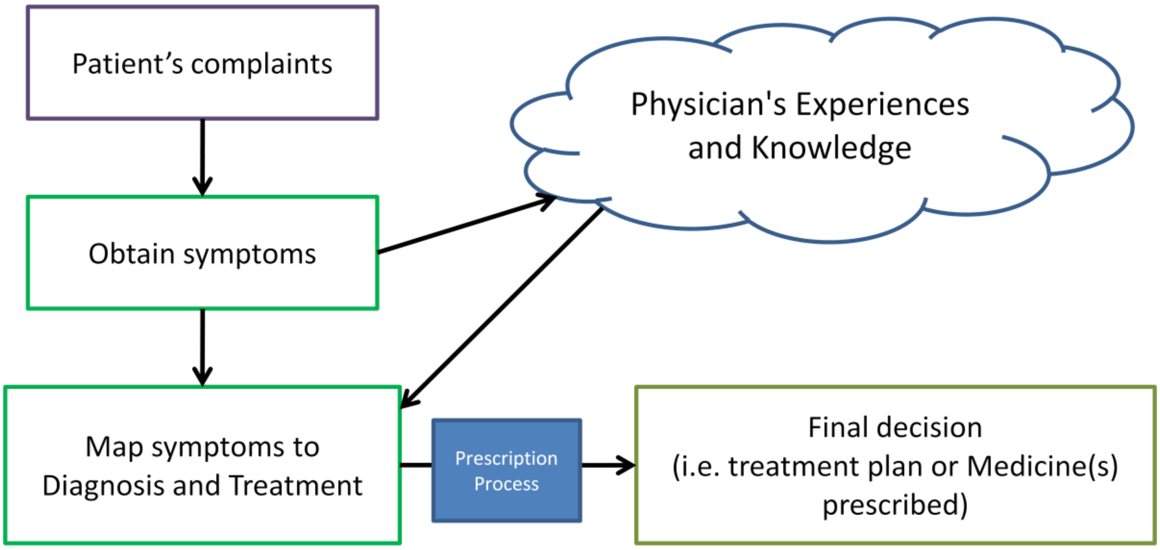


Figure 2.1 – Medical Prescription Process

The writing of prescription is one of the most frequent and significant therapeutic interactions between physicians and patients. Normally, physicians will give more than one drug to patients according to the physicians' diagnosis. Multiple drugs possibly increase the risks of drug-drug interactions and other adverse cases. Moreover, it would be harmful if the patient does not have a medical need or have medical conditions or allergies that could interact dangerously with the drugs (Majkowski et al., 2002). In-person consultation with a physician is significant for the physicians to consider a patient's medical history and any core conditions when deciding the appropriate drug to prescribe. Because patients do not understand well about drugs properties, all patients tend to trust the physicians and follow instructions from the prescription. If the prescription is not suitable for the patients or contains negligence, the patients suffer pains and even die after taking the drugs. In order to protect the benefits of patient, the demands imposed upon the physician from the prescribing process are significant (Majkowski et al., 2002).

According to the research from Federal Interagency Forum in 2000, the increasing number of patients has caused confused prescriptions or inappropriate medication (Classen et al., 2005). Furthermore, as stated in the study of Department of Health in Hong Kong in 2006, 442 medical errors were identified within two months (Tam et al., 2008). More than half of these cases were related to prescription errors. The other two hundreds cases involved side effects of the drugs such as allergies. These prescription errors were mainly attributed to mistakes in choosing drugs by physicians. Moreover, physicians often do not have the patient's medication history because physicians cannot read the illegible handwriting of other physicians. No medical knowledge supports physicians to make decisions. All of these factors endanger prescription safety.

2.2.1 Relationship Among Healthcare Professions and Patients

A medical prescription serves as a medium of communication between the physician and the pharmacist/nurses to ensure that the right medication is delivered to the patient. Figure 2.2 depicts the medical prescription practices among physicians, nurses, pharmacists and patients. However, with voluminous drug information (i.e. more than 240,000 prescription drugs on the market), it is not easy for medical experts to be knowledgeable and familiar with the use of different drugs and with dosage instructions. Even with the same diagnosis, the medical prescription may differ from one patient to another as the patient's age and physical condition must also be taken into consideration in the prescription. This is especially the case for General Practitioners (GPs) as they are primarily responsible for providing comprehensive health care to individuals seeking medical care, and for making arrangements for other health care personnel to provide specialist services when necessary (Philips and Haynes, 2001). Thus, learning about new drug information, and remembering the appropriateness and possible contradictions of a large number of drugs remain open challenges for GPs (Westberg and Miller, 1999; Naylor, 2002).

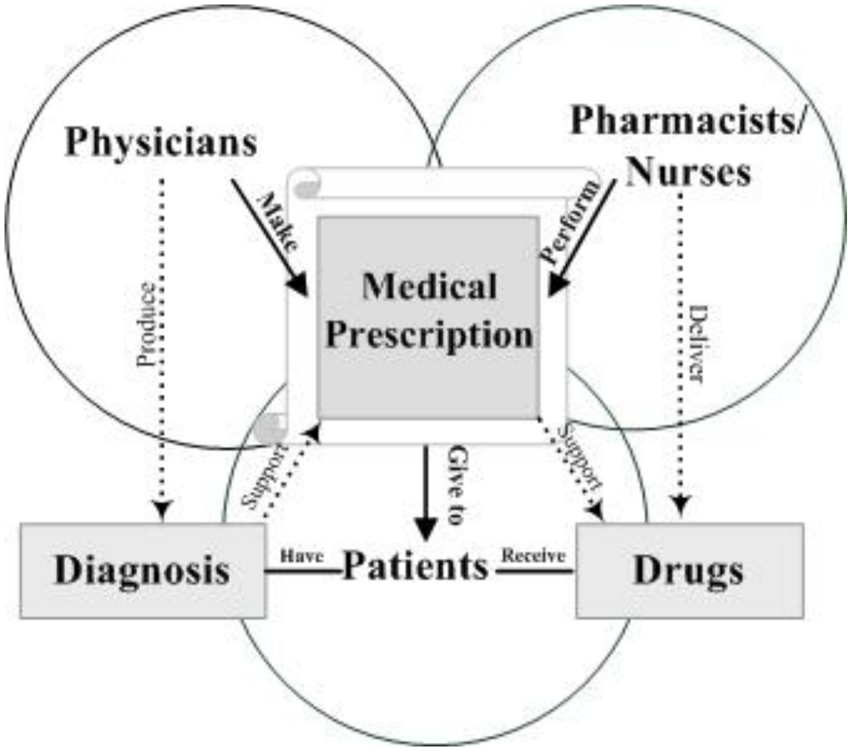


Figure 2.2 - Relationships between Physicians, Pharmacists and Patients in General Medical Prescription Practice

2.2.2 Related Studies on Medical Prescription Knowledge Acquisition

2.2.2.1 Direct Sharing Among Physicians

In the medical domain, knowledge is largely confined to experts who have worked in specific specialties for a long period of time. Their knowledge is usually in the form of tacit knowledge which is hard to extract (Ford and Sterman, 1998). When junior physicians have queries about a specific medical treatment, they usually consult senior colleagues or search for relevant information from books, journals and compendiums (Thompson, 1997; Dawes and Sampson, 2003; D'Alessandro et al., 2004; Grefsheim, 2007; Gonzalez-Gonzalez et al., 2007; Payne et al, 2007). However, these approaches to knowledge sharing and communication require plenty of time and human efforts in searching, filtering and interpreting information. Figure 2.3 shows the current approach to knowledge sharing among physicians which has the shortcomings described above. To address this problem, it is desirable to develop systems which can automatically acquire knowledge and experience of practitioners from medical databases to be deployed for decision support (Klein et al., 1989; Cooke, 1999; Kaur and Wasan, 2010).

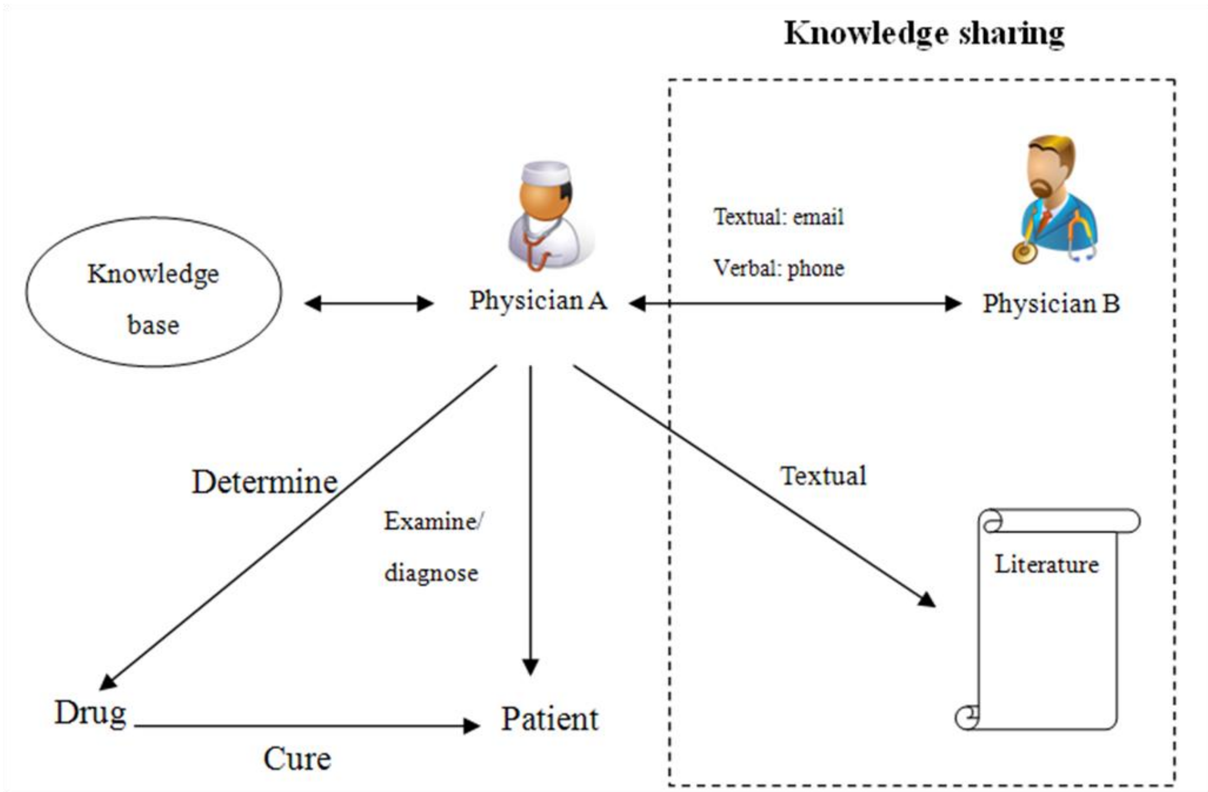


Figure 2.3 - Current Approach to Knowledge Sharing Among Physicians

Each diagnostic or medical decision involves selection amongst alternatives, each of which has a probability of correctness inferred from the experience of the expert. Experiential inferences and decisions, the implication of valuable tacit knowledge, are extracted by using knowledge elicitation techniques which can be classified into direct elicitation techniques and indirect elicitation techniques (Hudlicka, 1996). Direct methods elicit knowledge from experts without analysis. They include techniques such as story-telling, case-study and interview. The efficiency and effectiveness of these direct elicitation techniques largely depend on the ability of experts to articulate their implicit knowledge. In some cases, it may be difficult to verbalize or to introspect the knowledge being elicited (Cooke, 1994). On the other hand, eliciting knowledge using indirect methods requires human intervention, such as observation. These methods involve knowledge engineers to analyze the elicited information so as to generate the internal mental structure or models of the expert's knowledge (Hudlicka, 1996; Cooke, 1999).

2.2.2.2 Computer-assisted Learning

Due to the frequent occurrences of medication errors, educating physicians on medical prescriptions has become a widely discussed issue in recent years. According to the findings of Boreham et al. (2000), medical practitioners were of the opinion that prescription errors can be avoided when a knowledge base integrating scientific knowledge with clinical know-how is available to support their practice. In response to this need, the clinical section of the British Pharmacological Society developed a curriculum that covers the basic principles of medical prescription that prescribers should follow (Maxwell and Walley, 2003). Under the umbrella of the curriculum, Computer-Assisted Learning (CAL) may also have a role in the education of good prescriptions (Aronson, 2006). For example, interactive case-based and evidence-based learning modules on prescription are introduced into Australian medical education programs (Smith and Tasioulas, 2002). More details in CAL for education of good prescription practices can be found in the study of Herzig et al. (2002). Apart from the medical school students, the target audience of e-prescribing systems can be extended to practicing physicians.

In recent years, e-prescribing has emerged as a new trend in medical continuing education as it improves physicians' prescription practices at the point of care through the use of Information and Communication Technology (ICT) and e-learning interventions (Ruiz and Hagenlocker, 2006). It provides a platform for physicians to update their pharmacological knowledge on a need basis even though they may not have enough time to keep abreast with the latest knowledge in the domain. By integrating with a range of electronic resources (such as medical databases, electronic journals, and scientific drug information), e-prescribing can provide just-in-time prescribing training in the clinical environment (Ruiz and Hagenlocker, 2006). With the sheer volume of reputable sources, the information extracted from such systems has promised that medical prescriptions can become safer and more effective without forced disruptions to the physician’s workflow (Schiff et al., 1998; Rosenbloom et al., 2005).

2.2.2.3 Limitations of Existing Approaches on Medical Prescription Knowledge Acquisition

Prescribing medicines is a complex decision that a small change may result in serious consequences. Up to now, the therapy and medicines to be administered to a patient rely heavily on the knowledge of physicians. In other words, physician’s cognition and understanding of a drug’s properties (like dosage instructions) are critical and important for determining accurate prescription in a timely manner. Although numerous sharing platforms have been proposed for physicians to obtain relevant information, physicians often do not have time to keep abreast with the growing body of knowledge due to heavy demand for their services (Jabr, 2007). Furthermore, a high degree of direct involvement of experts is one of the major limitations of these knowledge acquisition techniques. The domain experts may not be available due to lack of time or they are unwilling to share knowledge. Furthermore, knowledge of medical experts is usually tacit, making it difficult to acquire. The effectiveness of information transfer through verbal exchanges remains a questionable issue (Herschel et al., 2001). The usability of the knowledge is highly dependent on the verbal communication skills of the knowledge provider. If a prescribed medication is not appropriately matched to the pathophysiology of the disease, safety and health of the patient will be jeopardized (Aronson, 2006). Furthermore, patients may take multiple medicines in a serving (Susan and Warren, 2000). In such situations, interactions between drugs (so-called drug-drug interactions) may cause harm to the patient.

Although the current knowledge sharing approaches and CAL system can help to reduce medication errors significantly, they are primarily static and rule-based such that they are ill-equipped to take into consideration dynamic clinical situations. To be an effective prescriber, he must (i) understand the basic pathophysiology of diseases, (ii) assess the balance between benefits and harms of a particular form of treatment, (iii) pay attention to the practical matters related to the choice of drugs, and (iv) discuss with the patient the proposed treatment and its potential effects, before deciding on any prescription (Aronson, 2006). Attempted to address these issues, numerous researchers have suggested that applying technology in medical practices can help GPs to stay informed about the latest development of drugs and thus can help to reduce medical errors and improve patient safety. To support the decision making process of the medical experts, Electronic Medical Records (EMR) systems have been introduced to transform the traditional handwritten medical records into digital ones. Rector et al. (1993) present a model for an electronic medical record system which provides a permanent, complete record of patient care and the medical decisions made.

Kohane et al. (1996) applied client-server technology of the World Wide Web (WWW) to design national EMRS. Hammond et al.'s study (1990) has demonstrated that using EMR not only can improve the quality of patient care and decrease medical errors, but also can result in a positive financial return on investment. With such a sound financial achievement of EMR, many researchers are focusing on how to inform the clinical decisions from the stored medical data. Shiffman et al. (1999) and Linnarsson (1993) claimed that integration of EMR with a Decision Support System (DSS) can enhance effectiveness in ensuring patient safety. The benefits of current DSSs used in general practice include assisting doctors in performing diagnosis, enhancing decision making quality in the primary care consultation and in selecting appropriate dosage (Thornett, 2001). All these are in line with the results of Wang et al.'s 5-year study (Wang et al., 2003).

2.3 Academic Study of Medical Informatics for Knowledge Acquisition and Modeling in Medical Prescription Decision Support System

Medical informatics is the intersection of the computer science and art of processing medical information (Sarbadhikari, 1995; Shortliffe et al., 2000; Sarbadhikari and Pal, 2002). Today, a tremendous amount of medical information has been accumulated; and confusion is easily created when there is an improper access to the information. In response to this challenge, numerous researchers advocate adoption of medical informatics that employs computers, computational intelligence techniques, clinical guidelines, and formal medical terminologies to deals with the resources, devices, and methods required to optimize the acquisition, storage, retrieval, and use of medical information.

Despite growing concerns about the quality and accuracy of prescription drug information, Kozier et al. (2004) discusses that there are eight factors needing to be considered during the prescription process:

(i) Developmental factor

(ii) Demography factor

(iii) Cultural, ethnic, and genetic factor

(iv) Eating habit factor

(v) Environmental factor

(vi) Psychological factor

(vii) Illness and disease factor

(viii) Time of administration factor

Apart from the above mentioned factors, adverse drug events (e.g. drug-drug interactions), and unexpected and mutated diseases are also critical in driving the prescription being succeeded or failed. This complex knowledge of prescription practices will make physicians hard to select the appropriate drugs to their patients. Unlike other industry, one mistake in healthcare industry may result in a serious consequence like death. Thus, it is necessary for physicians to have a better method of acquiring others' experience and hence alert intelligently if there is any abnormal processes in prescription.

2.3.1 Electronic Medical Records System

Traditionally, patient information and prescriptions are recorded on paper. Obviously, such recording systems present storage, readability and retrieval problems. Thus, there is a recent trend of migrating from paper-based to computer-based processing and storage of medical records. This change facilitates the use of patient data and medical knowledge; it also offers more functionalities that support medical decisions (Haux, 2006).

Concurrently, medical informatics has developed significantly, particularly in the improvement of quality and effectiveness (Downing et al., 2009). Patient data can be used as a valuable resource for creating a physician's medical knowledge that supports the specific clinical advice given to the patient (Kawamoto et al., 2009). Various methods such as data mining techniques can also be used to extract IF-THEN diagnostic rules from the available data,. Knowledge can then be discovered from medical data sets effectively. Kaur and Wasan (2010) suggest that the association-rule mining technique can discover the association and connection relationships within huge medical data sets. Once discovered, these association relationships will provide valuable support for medical care planning and prediction of patient condition and recovery. These rules or relationships constitute a physician's tacit knowledge. Knowledge can be classified into tacit and explicit knowledge. Domain experts usually possess tacit knowledge, which is personal and unarticulated, making it difficult to elicit and manage (Barbosa et al., 2009).

EMRS is gaining acceptance as tools used in medical informatics for computerizing medical records (Hersh, 2009). A comprehensive EMRS can provide its users a variety of functions to handle routine operations in a healthcare organization or manage patient information (Lai et al., 2009). Essentially, EMR is a repository of health-related information created by physicians. Such information comprises prescriptions for medications and results of diagnostic tests. Some EMRSs even provide decision support information such as drug-to-drug interactions and recommendations of care practices (The National Alliance for Health Information Technology, 2008). EMR is mainly adopted to enhance clinicians' ability to discover and provide effective treatment of patients in their care (The National Alliance for Health Information Technology, 2008). EMR, as a large repository of electronic patient medical records, can be used to identify physicians' experiential therapeutic decisions by employing data mining and knowledge discovery techniques for generation of guidelines and alternatives for clinical practices (Toussi et al., 2009; Kaur and Wasan, 2010). EMR can be used as a platform for converting clinician’s tacit knowledge to explicit form to facilitate deployment of tacit knowledge (Herschel et al., 2001). For example, EMR can be integrated with Case-Based Reasoning (CBR) technology that advances patient-centered medicine through individual knowledge processing (Pantazi, 2004). It can also facilitate sharing of patient information for improving quality of patient care and detection of medical errors (Zhang, 2002; Wang et al., 2003; Sequist et al., 2007).

2.3.2 Computer-assisted Decision Support System

Turban and Aronson (2001) define DSS as “an interactive, flexible, and adaptable computer-based information system, especially developed for supporting the solution of a non-structured management problem for improved decision making. It utilizes data, provides an easy-to-use interface, and allows for the decision maker's own insights”. Computer-assisted DSS is mostly developed to provide physicians with advices on either diagnosis or treatment by the means of Artificial Intelligence (AI) (Delaney et al., 1999). Often, a simple diagnostic procedure or test is overlooked and the disease eludes diagnosis, the DSS is therefore conducted a comprehensive history taking to facilitate the solutions made in each diagnostic process. Table 2.1 summarizes the DSS employed in medical aspects.

Table 2.1 - Common DSS Employed in Medical Aspects (Modified from Kaplan, 2001)

Authors	**Description**
Kaplan et al., 1997	Design suggestions and user acceptance issues were identified.
Bates et al., 1998	POE decreased rate of medication errors.
Bouaud et al., 1998	Clinicians agreed with 96% of the recommendations and followed one of the recommendations in 65% of cases.
Monane et al., 1998	A computerized drug utilization review database linked to a telepharmacy intervention improved prescribing patterns.
Raschke et al., 1998	Physicians changed orders as a result of being notified by the pharmacist or radiology technician who screened the alerts.
Friedman et al., 1999	DSS consultation modestly enhanced subjects' diagnostic reasoning.
Kuperman et al., 1999	The automatic alerting system reduced the time until treatment was ordered.

Supported by the reasoning and modeling of human brains, DSS can give advices and explanations like a human consultant, if necessary, it can also give the logic behind the advices. Therefore, there is a strong possibility that DSS demonstrates a great potential in the area of medical prescription for handling the complex drug information. However, in existing literatures, only a few publications have discussed this issue. One of the publications as proposed by Warren et al. (1998) can reduce the drug choices after specifying the diagnosis; it lacks the consideration of physicians' prescription behaviors and the patients' particular details. Therefore, an intelligent and dynamic medical prescription support approach is required to ensure the right medications at the right amount to the right patient as reported by Chan (2006).

2.3.3 Computerized Drug Order Entry

Generally, a registered physician needs to write an average of more than 2000 prescriptions every year. Stanton et al. (1994), and Nelson and Talbert (1996) report that modern potent drugs are the cause of hospitalization in 10–16% of internal medicine cases. In response to this worrying issue, Kohn et al. (2000) state that Computer-based Drug Order Entry (CDOE) systems have been recommended by the Institute of Medicine as a means of enhancing patient safety. In the existing literatures, CDOE allows physicians to use a computer to enter medical orders directly into a clinical information system, potentially saving time and reducing medical errors (Sitting and Stead, 1994). Oliven et al. (2005) propose a computerized, on-line surveillance method to prevent prescription errors of CDOE by connecting the drug database, and patients' database with the hospital's administrative and laboratories database. All the adverse drug events, such as real-time drug-allergy (including cross-allergy to similar drug class or compounds), drug-disease and drug-laboratory contraindications, are detected automatically after the physicians chose the drug be prescribed to patient. Even though CDOE can drive the physicians into less prescription errors, a study conducted by Lee et al. (1996)

exhibits that screen design and options in order entry are common recommendations in changing the CDOE (Figure 2.4). This aligns with the results of the research discussed by Shepherd (2007). Thus, it is suggested to introduce the elements of computational intelligence techniques in the existing CDOE in order to enhance the quality of services.

Responses to the Open-ended Question: "If There Were One Thing You Could Change about Order Entry, What Would It Be?"

Physicians

*Most common responses**

- Speed, slow response time
- Too many screens—too many steps between logon and order entry

Common responses

- Cannot cancel duplicate orders without canceling all orders
- Not user-friendly
- Nurses not accepting verbal/telephone orders
- Allowing laboratories for coverage list
- More help/support
- More flexibility—being able to take orders "verbatim"
- Too many options in order entry
- Asks for too many variables

*The "most common" responses were cited by more than half of the respondents, while the other responses were cited by more than one respondent.

Figure 2.4 - Physicians' Answers to the Change of CDOE (Lee et al., 1996)

2.3.4 Knowledge-based System

A Knowledge-Based System (KBS) aims to understand and initiate human knowledge in the computer system, thereby making expertise available for decision making, and information sharing, when and where needed (Wiig, 1994). Concerning the tremendous amount of data stored within the organization, the classical KBS needs to combine machine learning and structured background knowledge representation, such as ontology, and causal representations and reasoning, to obtain better performance. In additions, information sharing is concerned with creating collaborative knowledge environments for sharing and disseminating information.

In recent years, KBS have gained increased attention both in healthcare knowledge management and in medical prescription. Most KBSs employ AI techniques to develop a knowledge-centric healthcare system for gathering prescriptions in a knowledge repository and disseminating the knowledge to all parties for reuse and problem solving (Schmidt et al., 2001; Sim et al., 2001; Van Hyfte et al., 2001). CBR is one of the most prevalent knowledge extraction methods used in developing KBSs because it has a stronger explanation capability than other techniques like neural networks (Baesens et al., 2003). Related work on using KBS enables physicians to share past experiences stored in the knowledge base to encounter new situations. Generally, physicians have developed their own prescription style and behaviors based on their knowledge and experience. In this situation, the problem solving is presented in a single looping process that generates a solution prescribed by the physician himself/herself previously.

2.3.5 Data Mining

Data Mining (DM) is the process of finding the patterns, associations or relationships among data using different analytical techniques involving the creation of a model and the concluded result will become useful information or knowledge. DM can also be expressed as:

- Nontrivial extraction of implicit, previously unknown, and potentially useful information from data (Frawley et al., 1992); and
- Making sense of large amounts of mostly unsupervised data in some domain (Cios et al., 2007).

It is an interdisciplinary subject that lies at the intersect of pattern recognition and database systems and emerges the techniques from the mathematics and statistical disciplines as well as from the artificial intelligence and machine leaning communities. It has a great deal in common with statistics but on the other hand, there are differences. Unlike statistics, data mining can be due with heterogeneous data fields.

Very often, the term knowledge discovery is used together with Data Mining. Knowledge discovery, also known as Knowledge Discovery in Database (KDD), is the process that seeks new knowledge in some application domain. DM is one of the steps in the knowledge discovery process. Figure 2.5 is an outline of the six step hybrid KDD model developed by

Cios et al. (2007).

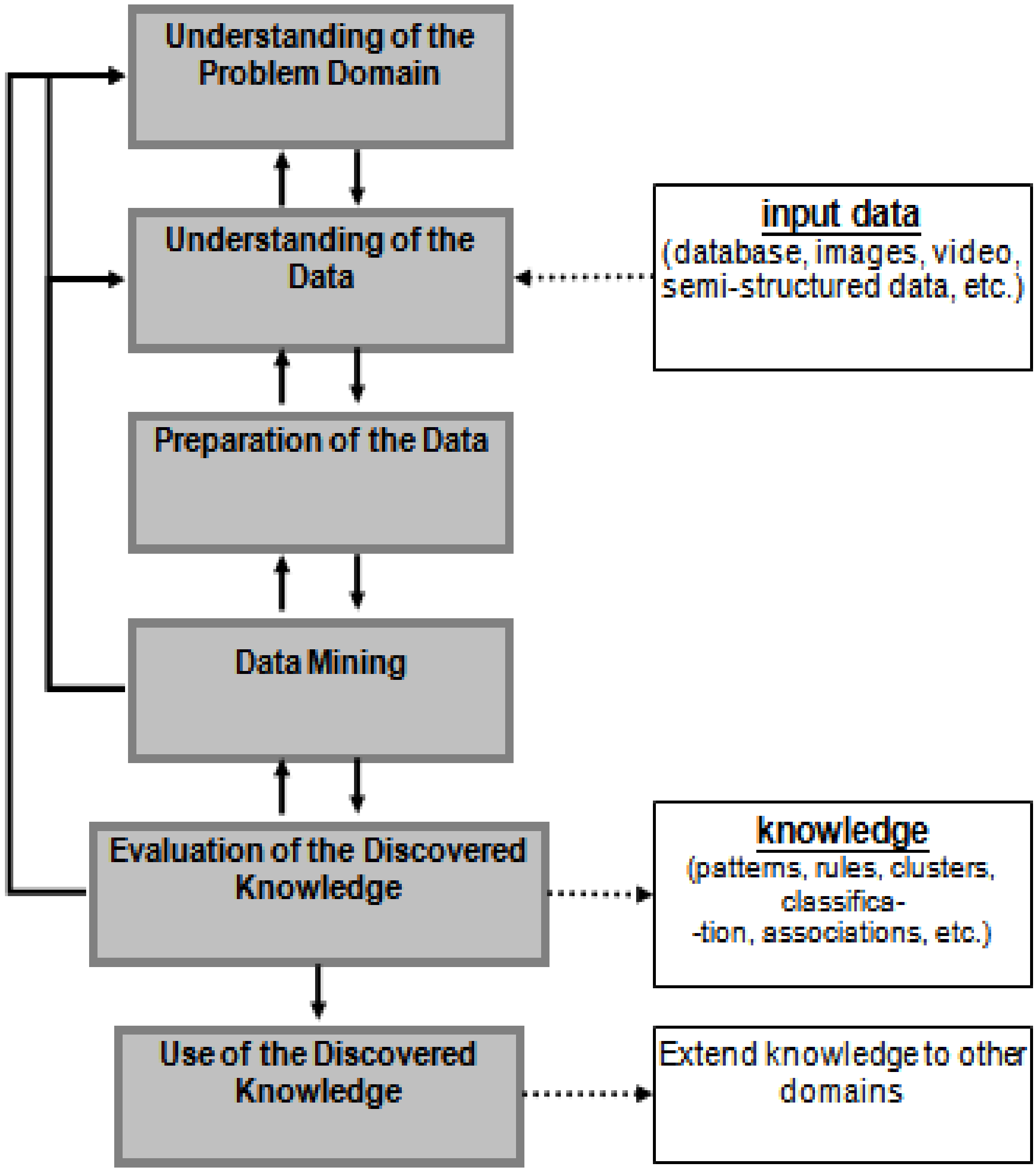


Figure 2.5 – Six-step Hybrid KDD Model (Cios et al., 2007)

The initial step of understanding the problem domain involves working closely with domain experts to define the problem and determine the project goals, and learning about current solutions to the problem. A description of the problem, including its restrictions, is prepared. The DM tool to be used in the later stage is selected. Next, we need to understand the data which includes collecting sample data and deciding which data, including format and size, will be needed. Data are checked for completeness, redundancy, missing values, plausibility of attribute values, etc. Preparation of data decides which data will be used as input for DM methods in the subsequent step. It involves sampling, running correlation and significance tests, and data cleaning. Data miner then uses various DM methods to derive knowledge from preprocessed data. Evaluation includes understanding and checking if the result is novel. Finally, we will decide how to use and deploy the discovered knowledge.

The typical data mining process involves transferring data originally collected in production systems (such as electronic medical records) into data warehouse, cleaning or scrubbing the data to remove errors and check for format consistency, and then searching the data using statistical model, artificial intelligence (such as neural networks), and other machine learning methods (Krivda, 1995). Prather et al. (1997) employs the KDD for identifying the factors that will improve the quality and cost effectiveness of perinatal care in an extensive clinical database of obstetrical patients. Given the data warehouse of diabetic patients, Breault et al. (2002) employ the CART to investigate the factors affecting the occurrence of diabetics. They are surprisingly discovered that younger age predicts bad diabetic control, in which explore a new area to manage the diabetic control in younger age. Similar applications of data mining can also be found in Table 2.2.

Apart from the diagnostic prediction, the knowledge discovery ability in data mining also demonstrated a good detector in Adverse Drug Events (ADE). Wilson et al. (2004) utilize the KDD techniques in pharmacovigilance for detecting signals earlier than using existing methods. Lian et al. (2003) has pointed out that the prescription is specified by a preference function based on the user's preference in prior clinical experience. Thus, they propose a dose optimization framework based on probability theory. Susan and Warren (2000) have demonstrated that the Conditional Probability (CP) model is superior in optimizing the drug lists over the multiple linear regression and discriminant analysis models. Concerning the strong relationship between the diagnosis and medication, it formulates a posterior probability (what medication is needed) based on a priori probability (what diagnosis has been made). This approach aligns with the Mediface as purposed by Warren et al. (1998).

2.4 Existing AI and DM Techniques Used in Medical Prescription Decision Support System

AI and DM are emerging research areas promoting and stimulating in medical informatics that better deals with clinical health-related information, its structure, acquisition and use. In general, healthcare organizations are heavily depended on knowledge in terms of patient medical history, drug prescription procedure, hazard reports and medical expertise; whereas they are uncertain and complex in nature. To cope with this challenge, AI and DM methods, which are classified into experience-based and collective-based approaches, increase the accuracy and consistency of complex medical processes and problems.

2.4.1 Experience-based Approaches

Medical knowledge and information representation and visualization are significant in the design of medical informatics. To enhance the efficiency of knowledge and information processing and representation in computers, various techniques such as semantics and reasoning theory are commonly applied. Semantics specifies a commitment to how a symbol in the language relate to the task domain (Poole et al., 1997). With this semantic commitment in natural and intuitive characteristics, the result of a computation can be easily interpreted to represent the knowledge. Concerning the reasoning technique, reasoning theory or proof procedure is a possibly nondeterministic specification of how an answer can be derived from the knowledge base. In general, it is common to synthesize the reasoning theory with semantics techniques to ensure the soundness and completeness of answers obtaining (Poole et al., 1997).

Table 2.2 - Recent Applications of Data Mining

Authors	**Description**
Megalooikonomou et al. (2000)	They introduce statistical methods that aid the discovery of interesting associations and patterns between brain images and other clinical data
Brossette et al. (2000)	They design a Data Mining Surveillance System (DMSS) that uses novel data mining techniques to discover unsuspected, useful patterns of nosocomial infections and antimicrobial resistance from the analysis of hospital laboratory data
Antonie et al. (2001)	They investigate the use of different data mining techniques for anomaly detection and classification of medical images
Coulter et al. (2001)	They examine the relation between antipsychotic drugs and myocarditis and cardiomyopathy
Li et al. (2004)	They explore a novel analytic cancer detection method with different feature selection methods and to compare the results obtained on different datasets and that reported by Petricoin et al. in terms of detection performance and selected proteomic patterns
Delen et al. (2005)	They use two popular data mining algorithms (artificial neural networks and decision trees) along with a most commonly used statistical method (logistic regression) to develop the prediction models on breast cancer using a large dataset.
Su et al. (2006)	They use four different data mining approaches to select the relevant features from the data to predict diabetes
Phillips-Wren et al. (2008)	They assess the utilization of healthcare resources by lung cancer patients related to their demographic characteristics, socioeconomic markers, ethnic backgrounds, medical histories, and access to healthcare resources in order to guide medical decision making and public policy

2.4.1.1 Case-based Reasoning

CBR is used to solve a new problem by remembering a previous similar situation and by reusing information and knowledge of that situation (Aamodt and Plaza, 1994). Similar to human problem-solving process, CBR requires a knowledge-based learning mechanism to learn from old cases and reuse the most specific case or set of cases to explain the new situations (Hammond, 1989). Compared with other problem-solving techniques (such as Bayesian networks and neural networks), CBR does not have the tendency to over-generalize (Mitchell, 1997), and thus CBR can achieve excellent accuracy provided that it generates the solutions from the memorized cases (Bichindaritz and Marling, 2006). Aamodt and Plaza (1994) stated that CBR cycle comprising of four "Re"s (Figure 2.6). They are:

(i) Retrieve the most similar case(s);

(ii) Reuse the case(s) to attempt to solve the problem;

(iii) Revise the proposed case if necessary; and

(iv) Retain the case as part of a new solution.

CBR is argued to be very effective in the medical domain. Bichindaritz and Marling (2006) stated that CBR is an essential tool in decision support in the health sciences because reasoning from historical examples is natural for healthcare professionals and case histories have long been used in the training of health care professionals. Table 2.3 presents how CBR is applied in the medical area. Huang et al. (2007) explained that the favor of CBR adoption in medicines are due to its cognitive adequateness, explicit experience, duality of objective and subjective knowledge, automatic acquisition, and system integration. Dussart et al. (2008) also argued that CBR is an effective reasoning strategy for optimizing clinical practice in which it learns through experiences and matches the natural reasoning model of human.

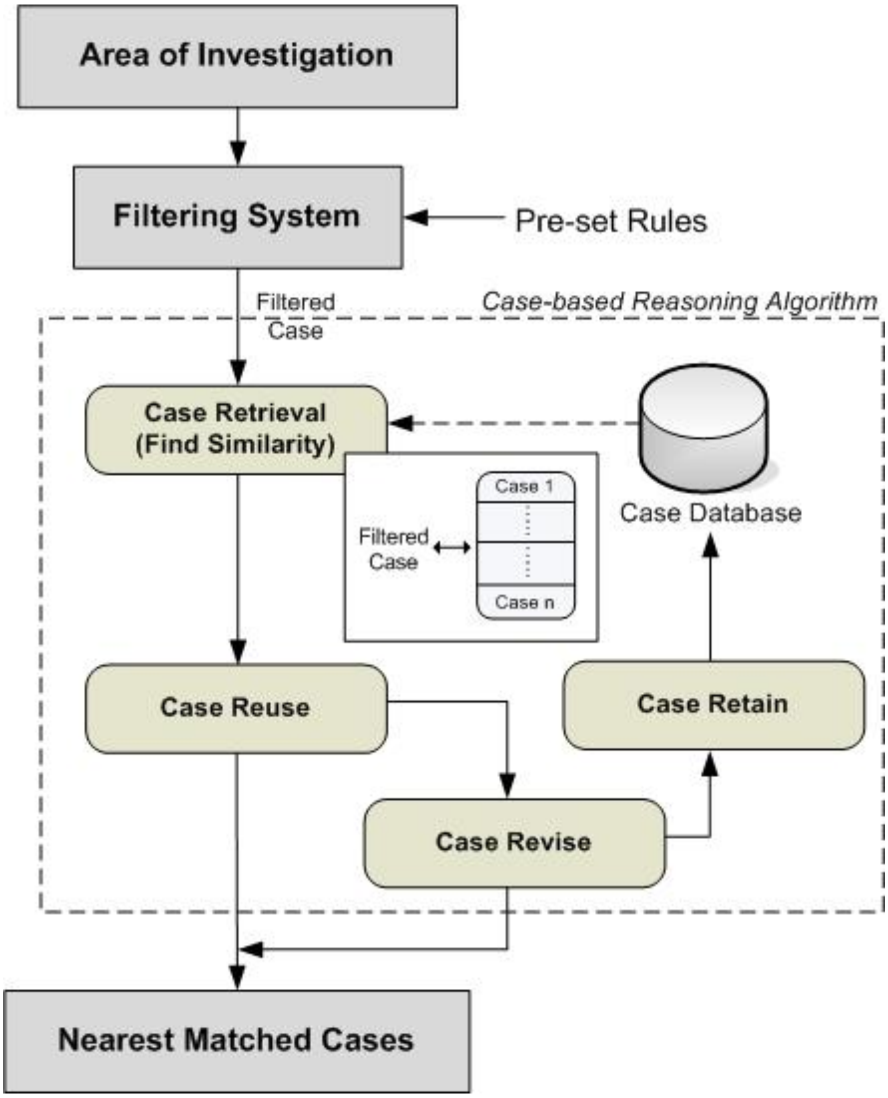


Figure 2.6 – Four "Re"s in CBR Cycle

Table 2.3 - Summary of CBR Adoption in Medical Area

Author(s)	**Description**
Turner, 1988	Diagnose pulmonology
Koton, 1988	Diagnose heart failure
Marling and Whitehouse, 2001	Enhance care of Alzheimer’s disease patients
Schmidt et al., 2001	Develop antibiotics therapy advice system
Montani et al., 2003	Integrate case-based reasoning with rules and models for diabetes mellitus management
Khan and Hoffmann, 2003	Provide diet recommendation
Bichindaritz and Potter, 2004	Classify and analyze medical image

Over the decades, CBR has widely applied in medical domain ranged from supporting diagnosis, prescription to treatment planning. For example, Marling and Whitehouse (2001) developed AUGUSTE to support treatment planning in Alzheimer’s disease by using CBR to determine if a neuroleptic medicine should be prescribed and hence select the approved medicines for a patient via a rule-based mechanism. Hartge et al. (2006) proposed a similarity measurement algorithm for a CBR system to help minimizing inappropriate selection of medicines that will cause adverse drug-drug interaction. Hartge et al. (2006) modeled the patient treatments as groups of vectors representing discrete time intervals to explore similarity in treatment of different patients. With the enhancement in the flexibility and speed of wireless computing, O'Sullivan et al. (2007) proposed that caregivers can input patients' symptoms to a mobile device for quickly retrieving similar profiles in supporting effective diagnoses and prognoses by comparing symptoms, treatments, diagnosis, test results and other patient information. In-depth review of applying CBR in medical domain can be found in Schmidt et al. (2001), and Yusof and Buckingham (2009).

Despite numerous researches showing CBR is effective in problem-solving in medical domain, several researchers argued that the chance of reusing a case from CBR is not high in some areas (Atzmueller et al., 2003), such as insurance claims prediction (Daengdej et al., 1999) and multiple medical disorder cases (Shi and Barnden, 2005). According to the study of Atzmueller et al. (2003), CBR can only solve about 3% of the cases on their real world dataset. It limits the power to explain and address the new problems. This is also true in the domain of prescription support. Since the solution of a prescription case typically involves multiple medicines (usually 5 to 7 medicines), not all the medicines are effective in addressing the problem in the new case. Thus, further modification of CBR is required to improve the accuracy of selecting the appropriate set of medicines in prescription support.

2.4.1.2 Concept Mapping Techniques

Concept maps, which are graphs that represent the relationships amongst concepts for a certain domain, are an example of knowledge visualization tools. Under the umbrella of concept mapping, related tasks performed by expert can be gathered, transcribed, analyzed, summarized, interpreted and transformed the acquired knowledge into a machine readable form (Klein et al., 1989; Wilson and Corlett, 2005). As attempts to construct reliable and knowledgeable concept maps, numerous research works that focus on the use of inductive learning methods to extract the decision rules from experts in an automated manner have been reported (Han et al., 1996; Kirkwood et al., 1998; Shaw and Gentry, 1998; Hung and Liang, 2001). These learning methods are efficient and useful because they operate from a set of training examples taken from previous decisions. Given the training examples, inductive learning can generalize and identify a set of rules to express the knowledge in a specific application domain (McKee, 1995). These methods have been successfully applied in a number of medical applications as these methods can help to induce the informal and implicit medical knowledge that is difficult to elicit. discuss the use of inductive learning methods to Examples of these applications include generation of rules for detection of honeycombing pattern from high-resolution computed tomography images (Zrimec and Wong, 2007) and the idea of using the patient's blood pressure, his general information and basic biochemical data as bases for classifying hypertension in a diagnosis (Wozniak, 2006), and prediction of lung cancer survival through integrated analyses of clinical and transcriptional data (Berrar et al., 2003).

2.4.2 Collective-based Approaches

2.4.2.1 Artificial Neural Networks

Artificial neural networks (Duda et al., 2001) are signal processing systems that try to emulate the behavior of human brain by providing a mathematical model of combination of numerous neurons connected in a network (Figure 2.7). It learns through examples and discriminate the characteristics among various pattern classes by reducing the error and automatically discovering inherent relationships in a data-rich environment. No rules or programmed information is need beforehand. It composes of many elements, called nodes which are connected in between. The connection between two nodes is weighted and by the adjustment of these weights, the training of the network is performed. The weights are network parameters and their values are obtained after the training procedure. There are usually several layers of nodes. During the training procedure, the inputs are directed in the input layer with the desirable output values as targets. A comparison mechanism will operates between the out and the target value and the weights are adjusted in order to reduce error. The procedure is repeated until the network output matches the targets. There are many advantages of neural networks like adaptive learning ability, self-organization, real-time operation and insensitivity to noise. However, it also has a huge disadvantage that it is highly dependence on the training data and it does not provide an explanation for the decisions they make, just like working in the ‘black box’.

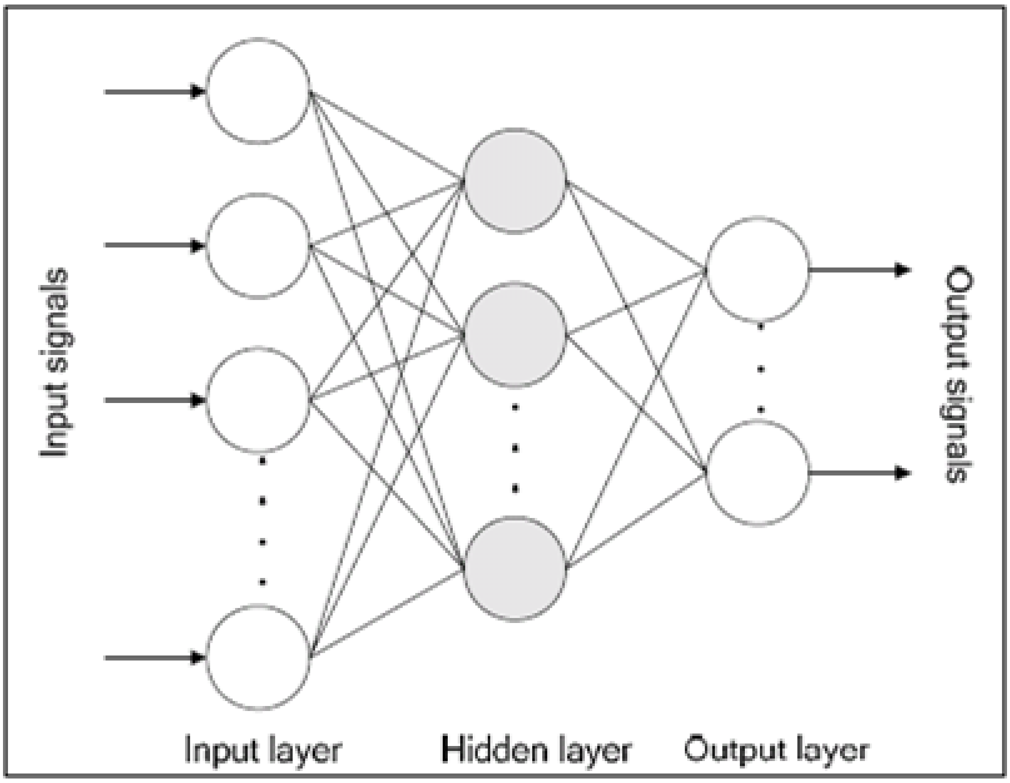


Figure 2.7 – Example of Artificial Neural Networks (ANNs)

2.4.2.2 Bayesian Theorem

Knowledge modeling is a popular research topic in medical informatics. It is “the concept of representing information and the logic of putting it to use in a digitally reusable format for purpose of capturing, sharing and processing knowledge to simulate intelligence”. With the aid of computerization of medical records, all the individual diagnosis transactions are collected and stored; thus forming a data warehouse that stores the collective behaviors of the medical practices within the organization. In recent years, many researchers find that it is effective to employ probability theory as a knowledge modeling model in medical prescription. For example, Lian et al. (2003) has pointed out that the prescription is specified by a preference function based on the user's preference in prior clinical experience. Thus, they propose a dose optimization framework based on probability theory. Susan and Warren (2000) have demonstrated that the CP model is superior in optimizing the drug lists over the multiple linear regression and discriminant analysis models. Concerning the strong relationship between the diagnosis and medication, it formulates a posterior probability (what medication is needed) based on a priori probability (what diagnosis has been made). This approach aligns with the Mediface as purposed by Warren et al. (1998).

With the growth of data in medical databases, modeling knowledge from data is growing in popularity. Due to the traditional analytical approaches are not enough to analyze such large data sets comprehensively; Bayesian Network (BN), which is a state-of-the-art representation of probabilistic knowledge by a graphical diagram, has emerged in recent years as essential for pattern recognition and classification in the healthcare field (Lee and Abbott, 2003). In general, BN describes the probability distribution over a set of variables by specifying a set of conditional independence assumptions along with a set of conditional probabilities (Jensen, 1996). Moreover, it is a Directed Acyclic Graph (DAG) with conditional probabilities for each node (Jensen, 2001). Figure 2.8 illustrates an example of a BN.

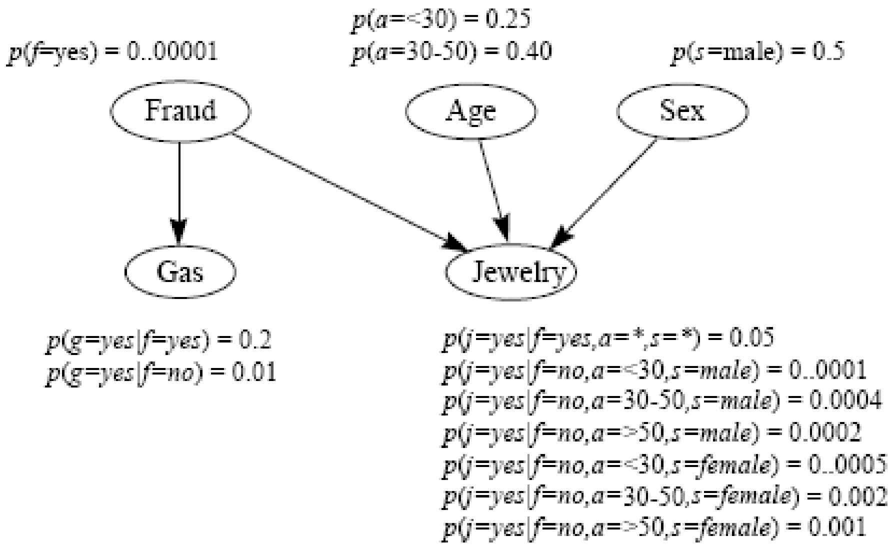


Figure 2.8 – Example of a Bayesian Network (Heckerman, 1995)

BN depends heavily on the conditional probability table that contains probabilities of the node being a specific value given the values of its parents. Conditional probability is a popular statistical modeling technique in BN in which CP is the probability that one event happens, given that some other event happens. In the medical domain, CP is particularly useful in prescription decision support, because it can quickly determine the probability of the drug required if a diagnosis has been made. Spenceley et al. (1997) argue that their model based on the conditional probability can reduce prescription choices by more than a half when compared with the conventional model. Nevertheless, it relies heavily on diagnosis classification, and there could be some problems such as the failure to take into consideration of physician's behaviors and patient's details (such as medicine allergy).

Discovery science (also known as discovery-based science) is a scientific methodology which emphasizes analysis of large volumes of experimental data or text data with the goal of finding new patterns or correlations, leading to hypothesis formation and other scientific methodologies (Fayyad et al., 1996; Fayyad and Uthurusamy, 1996). Knowledge discovering in databases is a kind of discovery science that looks for associations or relationships in operational or transactional data whereas text mining and information extraction is a new discovery technique in looking for concepts and their associations or relationships in natural language text (Lau, 2003). Supported by the discovery science, we can automatically capture the up-to-dated medical information from all the available sources (such as from Internet and literatures).

2.4.2.3 Association Rules Mining

Association rules mining is another common techniques used in DSS (Lee et al., 2001; Cho et al., 2002; Chien and Chen, 2008; García et al., 2008). It aims to extracts interesting correlations, frequent patterns, associations or casual structures among sets of items in databases (Kotsiantis and Kanellopoulos, 2006). A famous example of applying association rules is the market basket analysis. Agrawal and Srikant (1996) introduced the Apriori algorithm for discovering regularities between products in large scale transaction data recorded by point-of-sale systems. The rules can be expressed as "{X, Y} $\Rightarrow$ {Z} [support: 60% and confidence: 80%]" meaning that X, Y and Z occur in 60% of all transactions (i.e. support) and 80% of the transactions containing X and Y contains Z (i.e. confidence). In general, a rule regards as interesting if it satisfies both the minimal support and confidence thresholds that pre-defined by experienced users or domain experts.

Similar to CBR, association rules mining is also widely applied in medical domain. The main reasons is due to its ability in uncovering new information and relationships embedded in the large databases, and generating new patterns and relationships throughout learning mechanism (Abidi, 2001; Hung et al., 2006). Kuo et al. (2006) employed clustering techniques to cluster the medical database into several groups, and hence apply the association rules mining algorithm to discover the hidden relation in the groups easier. Jiang and Gruenwald (2004) proposed to use association rules to mine the association relationships among different genes under the same experimental conditions. Shan et al. (2008) presented an application of association rule mining to detect fraud and inappropriate practice in the health service management domain.

All in all, association rules mining discovers important rules which provide useful references for GPs in making decisions. However, it is always difficult to handle the redundant frequent association when the database is large. Particularly in the prescription domain, the complex nature and large variety of medicines make the association rules mining difficult to identify useful and meaningful rules, and hence limits its ability in deriving solutions accurately. Furthermore, the threshold values of minimum support and confidence are difficult to be determined. It requires rigorous training of the parameters before real life applications.

2.4.2.4 Web and Text Mining

Regarding there is a lot of information presented in text or document databases, in form of electronic books, research articles, digital libraries, medical dictionaries, etc., several researchers developed a novel data mining approach in extracting useful knowledge from textual data or documents, so called the text mining (Hearst, 1999; Chen, 2001). For example, we can employs text mining techniques to extract the information of protein-protein interaction within three different documents.

In addition to the traditional data mining techniques, text mining uses techniques from many multidisciplinary scientific fields (e.g. text analysis techniques) to gain insight and automatically reveal useful information to the human users. Cohen and Hunter (2008) describe text mining is “the use of automated methods for exploiting the enormous amount of knowledge available in the biomedical literature”. One of the examples of text mining is to manage the health information in Internet and response the needs for those who have health information inquiry in HIV/AIDS (Ku et al., 2008). Another common application of text mining is used to extract the information of protein-protein interaction. When given the unstructured text, Zhou et al. employ the semantic parsing and hidden vector state model to mine the knowledge within the text (Zhou et al., 2006). By setting the annotation PROTEIN_NAME(ACTIVATE(PROTEIN_NAME), the system will automatically generate the result as shown in Figure 2.9.

Furthermore, Internet is growing at a tremendous speed. WWW becomes the largest database that ever existed. In particular, many medical literatures are written in electronic format which are widely available and accessible in the Internet nowadays. Therefore, the capability of knowledge discovery and retrieving information from WWW is important to physicians. But, the complexity of web pages and the dynamic nature of data stored in the Internet make adoption of data mining techniques difficult. Web mining is the use of data mining techniques to automatically retrieve, extract and evaluate information for knowledge discovery from the Internet (Mitra and Acharya, 2003). With its exploratory of hidden information ability, Yu and Jonnalagadda (2006) present an approach regarding Semantic Web and mining that can improve the quality of Web mining results and enhance the functions and services and the interoperability of medical information systems and standards in the healthcare field.

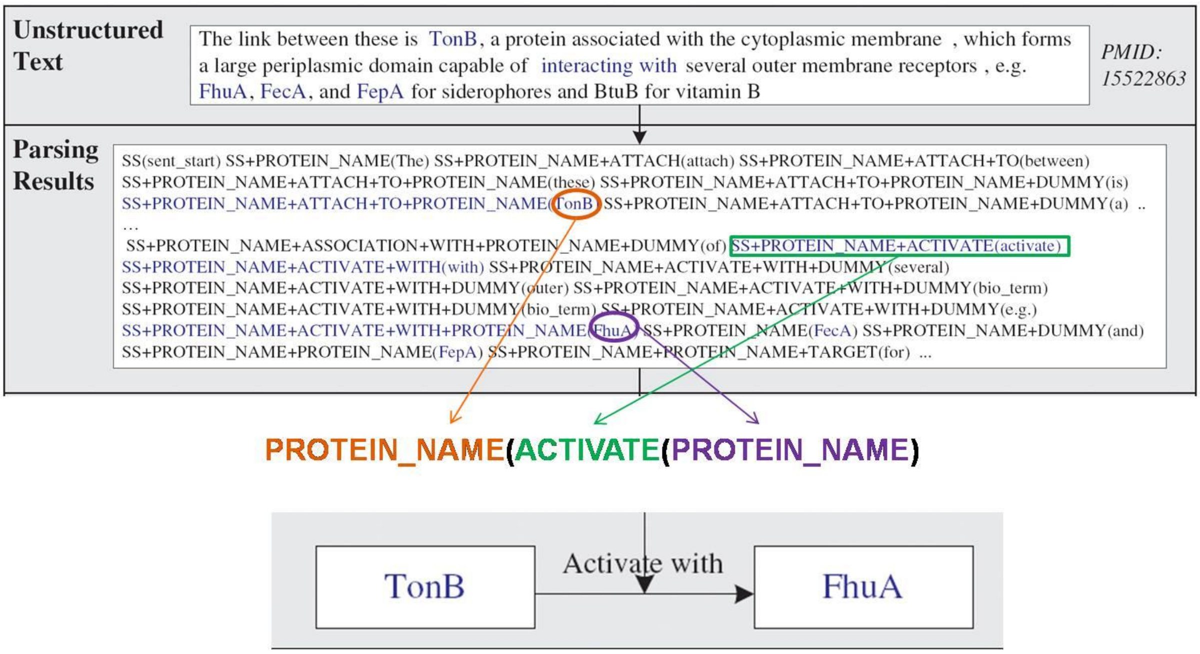


Figure 2.9 – Semantic Parsing Employed in Protein Documents (Zhou et al., 2006)

2.4.3 Hybrid Approaches

In recent years, numerous researchers intend to integrate several data mining and artificial intelligence techniques together to enhance the mining result and support decision making. For example, Kuo et al. (2007) integrate the clustering analysis and association rules mining technique to cluster the health insurance database and hence discover the useful rules for each group. Zhuang et al. (2009) combine the data mining and CBR methodologies to provide intelligent decision support for pathology test ordering by GPs. They guarantee the integrated system can enhance the testing ordering in term of evidence based, situational relevance, flexibility and interactivity. Huang et al. (2007) propose a model of a Chronic Diseases

Prognosis and Diagnosis (CDPD) system by integrating data mining and CBR to support the chronic disease treatment. Compared with traditional Coronary Artery Diseases (CAD) diagnostic methodologies, Tsipouras et al. (2008) integrate the decision trees and fuzzy modeling to form a fuzzy rule-based decision support system that obtain a significant improvement compared with artificial neural networks and adaptive neuro-fuzzy inference system. Example of such integration can be found in Figure 2.10.

The quality of a DSS is highly depended on its inference mechanism. Among the numerous inference techniques, CBR and association rules mining are the two common techniques used in the medical industry. CBR utilizes the specific knowledge of previously experienced and concrete problem situations (cases), while association rules mining relies on general knowledge of a problem domain and making associations along generalized relationships between problem descriptors and conclusions (Zhuang, et al., 2009). They are two distinct techniques that consist of their own strengths and limitations. One of the possible options is to integrate with rules derived by the domain experts or generated by mining from the databases as such integration can achieve new synergies and significantly improve the problem-solving capabilities in CBR (Kumar et al., 2009). Several researches have been done by integrating the two approaches. Montani et al. (2003) integrated CBR, Rule-Based Reasoning (RBR) and model-based reasoning to provide physicians with a reliable decision support tool in the context of type 1 diabetes mellitus management. Apart from approaches integration, Rossille et al. (2005) proposed a multi-modal reasoning decision-support system based on the RBR-first CBR-last approach to automatically compare the patient's case to the corresponding guideline, then to other cases, and retrieve similar cases for breast cancer. In recent years, Park et al. (2009) integrated CBR with Rule Induction (RI) techniques for case filtering. They applied their method to three medical diagnosis datasets and their findings demonstrated that the hybrid approach significantly outperforms the results in either CBR or RI. Important contributions have been made by integrating CBR and rules in numerous applications. However, lack of researches and empirical investigations have been done for the prescription related topics.

2.5 Summary of Literature Review

In this chapter, the state of the art of medical prescription and decision support method is reviewed and studied. It is clear that there is an increasing emphasis on medical informatics, computational intelligence, and knowledge acquisition and modeling to respond rapidly, effectively and efficiently to changes in the uncertain healthcare industry. There is a demand for a tool or methodology to support the drug selection and detect drug contraindication from ongoing drug information which captures and recognizes the complexity of the medical prescription.

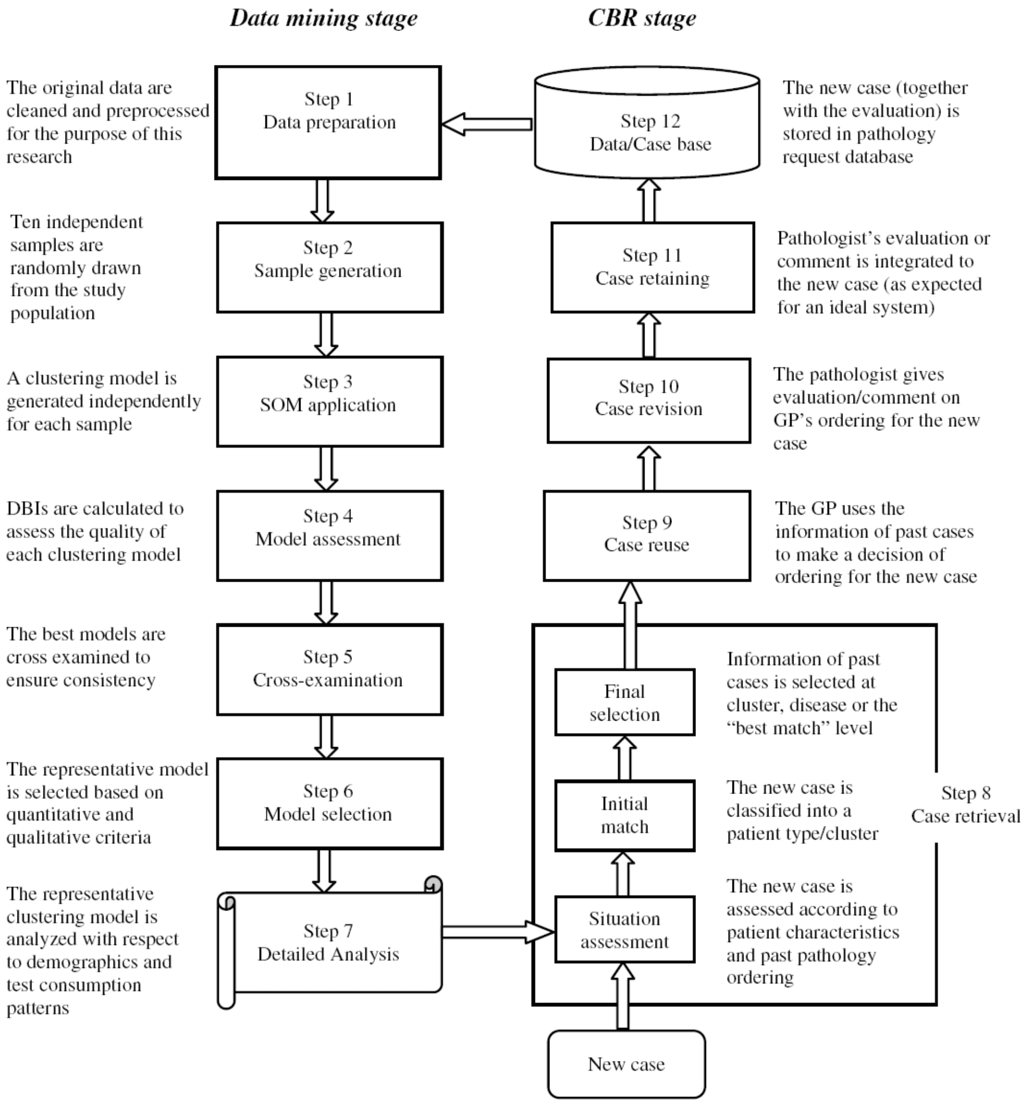


Figure 2.10 – Framework for the Integrated Approach (Zhuang et al., 2009)

It is interesting to note that the application of DSS in medical domain is mostly developed to provide physicians with advice on either diagnosis or treatment by means of AI and DM techniques (Delaney et al., 1999). Because of the complexity of drug information, DSS demonstrates great potential in the area of medical prescription. However, only a few publications have discussed this issue. One of the publications, proposed by Warren et al. (1998) describes how drug choices can be reduced after specifying the diagnosis; but it lacks consideration of physicians' prescription pattern and the patients' clinical background information. In this case, it cannot satisfy with the 'five rights' (i.e. the right drug, the right dose, the right time, the right route and the right patient) of medication administration (Chan, 2006) which is a standard of care. Therefore, a more comprehensive medical prescription support approach is required to ensure the right medications of the right amount are administered to the right patient and this needs to be addressed in this research.

To cope with the dynamic changes in the medical prescription and fluctuations in drug information (e.g. new dug launched in the market, new rules and policies set by the government), modeling the physician's prescription logic and acquiring ongoing prescription knowledge are required. From the review of the literature it is clear that there is a need for the development of a knowledge acquisition and modeling system for the medical prescription, as this will promote the acquisition, sharing and dissemination of knowledge from knowledge workers among the healthcare domain. EMR is an emerging tool used in medical informatics to computerize medical records and to establish a knowledge sharing platform among physicians (Hersh, 2009; Herschel et al., 2001). Since EMR stores various items of important medical data, it is argued that these data items can be turned into knowledge that is valuable to inform clinical decisions. Furthermore, EMR is an explicit medical record that stores the physician's tacit knowledge being deployed in each diagnostic process (Herschel et al., 2001). With the help of EMR, physicians' prescription behavior can be acquired and stored for further modeling. However, the adopting of EMR is always considered as an ill-defined problem. Better design with users' preferences in EMR development is useful for transforming the captured knowledge into explicit form that is easy for computation modeling.

On the whole, the results of literature review reveal that there is a need for the establishment of a methodology which enables knowledge acquisition, sharing, and modeling of physicians' prescription knowledge, as well as identifying the most appropriate medication list for the physicians after they determine the diagnosis of the patients. Table 2.4 describes the comparison between the knowledge acquisition, modeling, and support for the medical prescription between the existing EMRS, DSS, and the proposed system. It is claimed that the research methodology needs to incorporate the EMRS, AI, and DM technologies for providing the above knowledge inference process.

Table 2.4 – Comparison between Existing EMRS, DSS, and the Proposed System

	Existing EMRS[1]	**Existing DSS[2]**	**Proposed System**
Visualization and Record of Physician's Prescription Knowledge	✓	×	✓
Automatic Knowledge Elicitation of Physician's Prescription Knowledge	×	×	✓
Provision of Specific Diagnostic Template	×	×	✓
Sharing of Others' Prescription Knowledge	×	✓ (Based on pre-defined guidelines, which is not dynamic)	✓
Capability of Learning from Others' Prescription Knowledge	×	×	✓
Support of New Drug Selection	×	×	✓
integration of Individual and Collective Prescription Knowledge to Support the Prescription	×	×	✓
Detection of Drug-drug Interaction	×	✓ (Based on pre-defined guidelines, which cannot suit for new situation)	✓
Modeling of Multi-class Prescription Solution	×	× (Based on retrieving the whole set of solution)	✓

Available: ✓; Not Available: ×; [1]: Existing EMRS (Iakovidis, 1998; Levy et al, 2008; Lai et al., 2009); [2]: Existing DSS (Monane et al., 1998; Warren et al., 1998)

Chapter 3

Design of The Medical Prescription Decision Support System (MedicPDSS)

3.1 Introduction

An overview of research methodology is provided at the beginning of this chapter. Hence the architecture of the Medical Prescription Decision Support System (MedicPDSS) is described. MedicPDSS consists of a Template-based Electronic Medical Records System (TEMRS) and five different modules for supporting the prescription knowledge acquisition and modeling, knowledge sharing, and knowledge application. The TEMRS and five modules are supported by a Knowledge Repository (KR) and three computational intelligence algorithms (i.e. the automatic knowledge elicitation, the rule-associated case-based reasoning, and the drug information extraction).

The research methodology is shown in Figure 3.1. A literature review is firstly conducted which focuses on medical prescription, knowledge acquisition and modeling, knowledge-based system, data mining, etc. Hence, a MedicPDSS is designed which specifies the operation, and modules integration, information and knowledge flow. Then, a prototype of MedicPDSS is built. The algorithms of the system are evaluated so as to ensure the performance of each algorithm developed for MedicPDSS. The evaluations are carried out by comparing them with different well-known reasoning methodologies in forms of the precision and recall analysis. Thus, a trial implementation is carried out in a selected reference site to apply the whole system. Results of the trial implementation are evaluated through a case study.

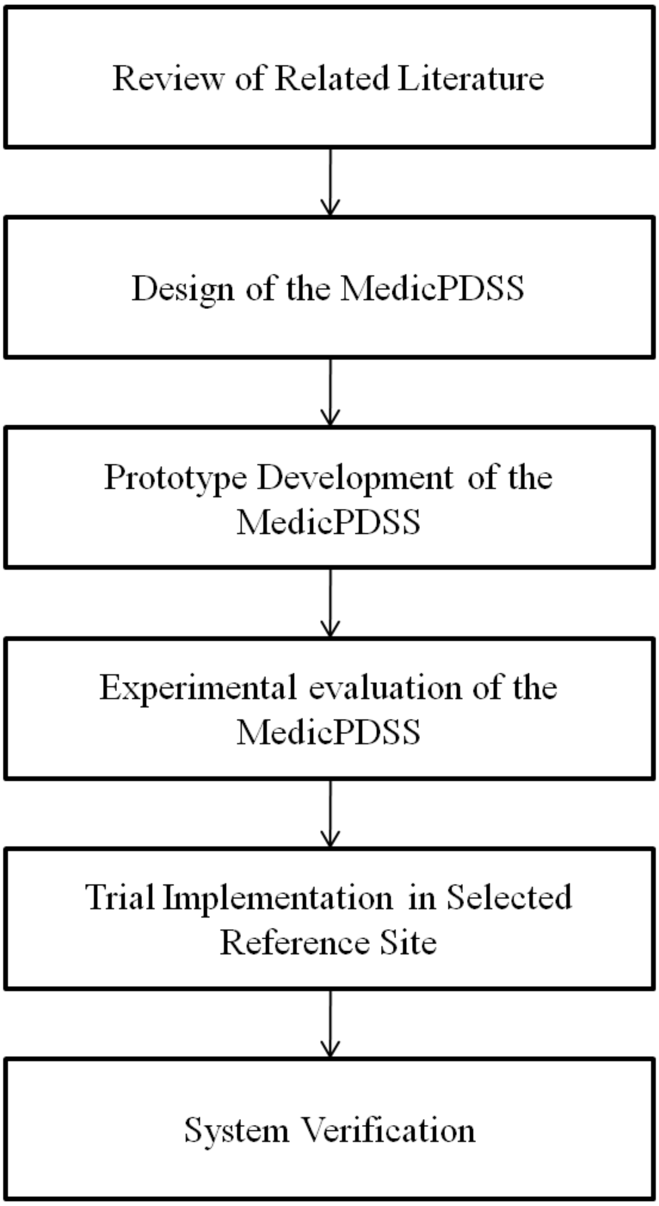


Figure 3.1 – Research Methodology

3.2 Medical Prescription Decision Support Approach

With the increased complexity and uncertainty in drug information, issuing medical prescriptions has become a vexing issue. As many as 240,000 medicines are available on the market, the approach supporting the medical prescription should provide GPs with medication advice and suggest a range of medicines for specific medical conditions by taking into consideration the collective physicians' prescription decisions as well as the latest drug-drug interaction rules from the Internet. It is aimed at acquiring the physicians' prescription behaviors through EMRS, and hence representing the captured knowledge into various concepts for visualization and analysis. Regarding the tacit prescription knowledge is presented explicitly; a computational intelligence approach is then applied to model the prescription experience of all physicians and thus recommend a range of medicines along with informed evidential decisions and latest drug interaction rules. Figure 3.2 shows the four fundamental elements of the approach, which are:

(i) Knowledge acquisition of prescription pattern and updated drug interaction;

(ii) Prescription concept formalization;

(iii) Individual and collective prescription behaviors modeling; and

(iv) Prescription decision support.

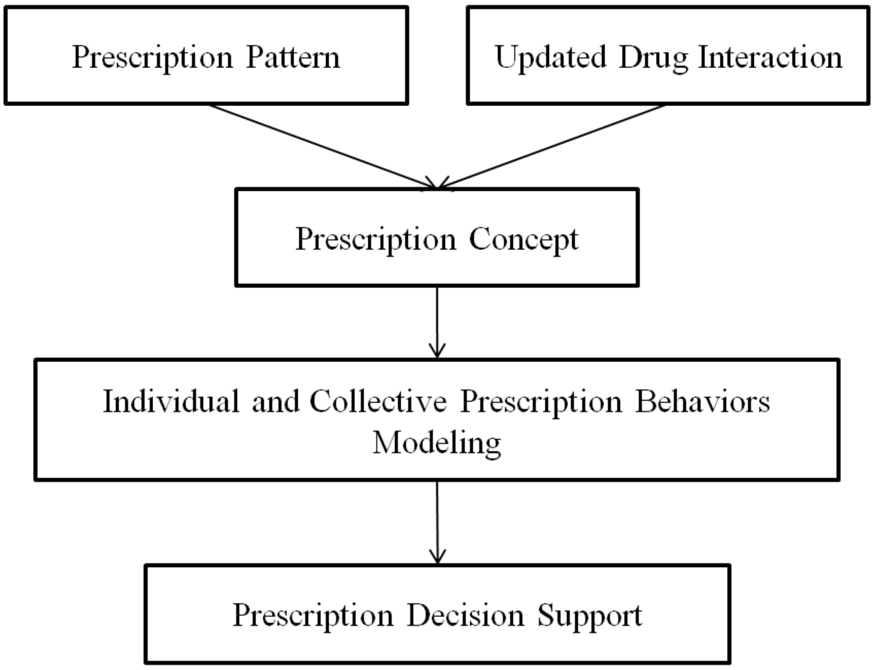


Figure 3.2 – Elements of Supporting the Medical Prescription

3.3 Architecture of MedicPDSS

To realize the concept of the prescription decision support approach, a Medical Prescription Decision Support System (MedicPDSS) is designed and built with the following objectives:

(i) To demonstrate the applicability of the prescription decision support approach in current studies as a proof of concept;

(ii) To develop a platform for acquisition, sharing, and reusing of the prescription knowledge; and

(iii) To develop a set of algorithms to support the modeling of prescription knowledge.

Figure 3.3 shows the system architecture of the MedicPDSS. The system consists of a TEMRS and five modules namely, Automatic Knowledge Elicitation Module, Medical Diagnosis Template Module, Prescription Modeling Module, Risk Surveillance Module and Information Services Module. The philosophy behind this breakdown is to separate the major activities of system applications into logical sections of display, processing logic and data services.

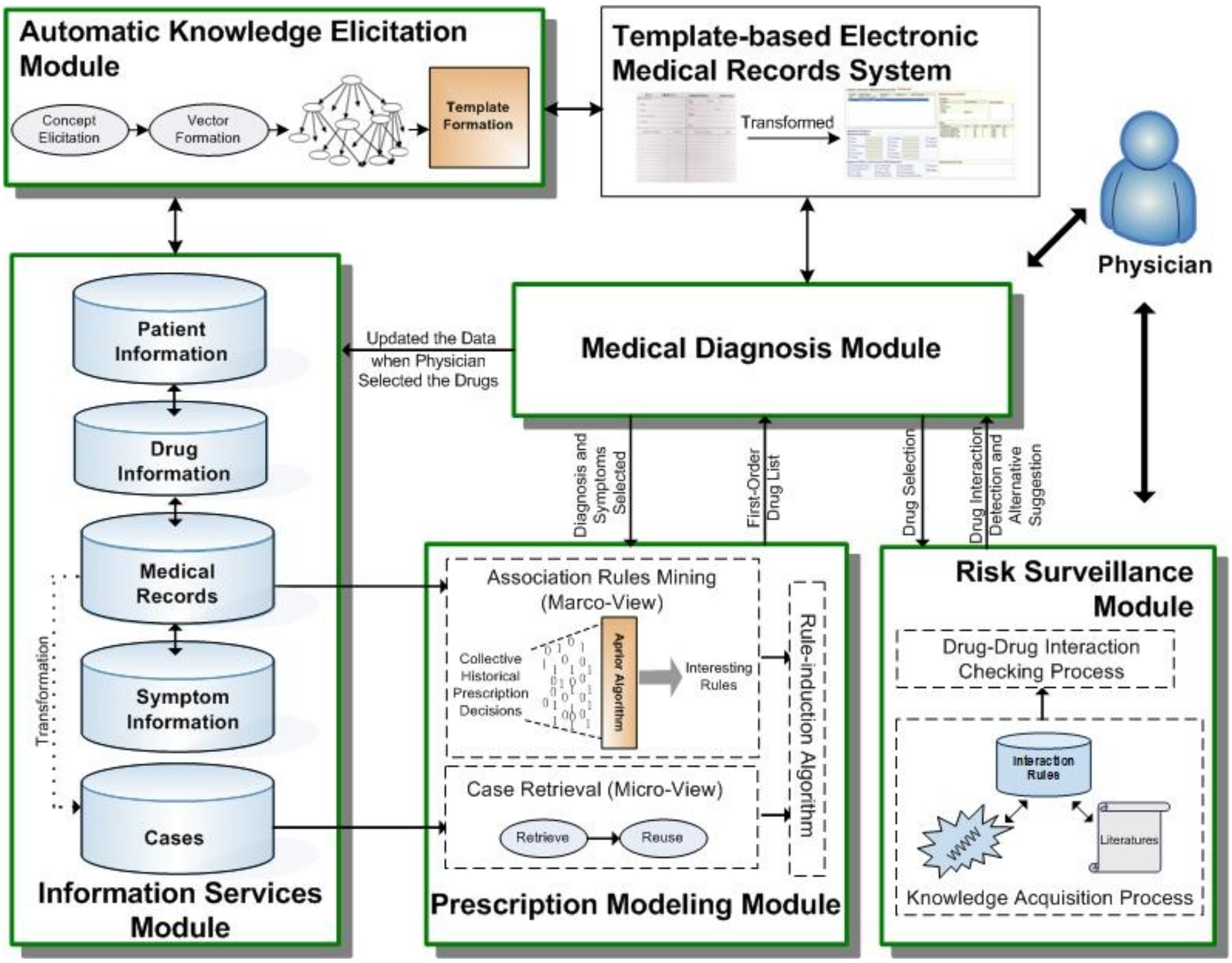


Figure 3.3 – System Architecture of the MedicPDSS

3.3.1 Template-based Electronic Medical Record System (TEMRS)

MedicPDSS leverages the clinical information stored in the medical records. In order to facilitate the decision support to the prescription, a machine-readable format for representing the medical knowledge is required. Thus, an EMRS with a simple interface is built for physicians to input data. However, EMRS has long been introduced into healthcare practice and have viewed to have the potential to augment patient care and clinical services (Brender et al., 2000; McInnes et al., 2006), and yet the implementation failure is high with a low penetration rate (Keshavjee et al., 2006; DesRoches et al., 2008). According to the literature, initial cost, time cost, physician perceptions and incentives, lack of standardization/interoperability of the EMRS, lack of technical support, data privacy and complexity of the user interface are some of the factors that have been identified (Iakovidis, 1998; Holbrook et al., 2003; Leung et al., 2003; Miller and Sim, 2004; Valdes et al., 2004; Anderson. 2007; Ludwick and Doucette, 2009). As EMRS is multifunctional, it is necessary for them to incorporate different information about the patient (i.e. demographic data, diagnosis, medications, radiological result, etc.). It is reflected by the physicians that 'industry-leading EMRS to be challenging to use because of the multiplicity of screens, options, and navigational aids' (Miller and Sim, 2004). This extends the time required for the physicians to learn to use the EMRS, and this can be a significant barrier that refrain physicians from adopting the EMRS in their practice. Moreover, human-machine interface flaws may lead to increase in incident of adverse patient events, rather than reducing it (Ash et al., 2004; Koppel et al., 2005). As a user's experience on the system is based on error frequency, learnability, and memorability (Rose et al., 2005), therefore, a system with high usability is essential in enabling it success in implementation.

As a result, a more user friendly and flexible user interface of the electronic medical record system is thus a vital factor to determine its success (Matsumura et al., 2007). It was showed that integrating a template into traditional EMRS, namely Template-based EMRS (TEMRS), can provide physicians with a flexible system in which their dynamic control can be exercised on. In order to build the template, it is important to acquire the diagnostic and prescription behaviors for each physician. Since the traditional EMRS is capable for managing the medical information of the patients, therefore in general the related information about patients, such as the patient's personal information, physical examination results, symptoms observed, diagnosis and treatment employed, is recorded in the system and thus used to be analyzed for formulating the template. Figure 3.4 shows the typical items of information stored in the EMRS.

Compared with the traditional EMRS, TEMRS provides various clinical selection interfaces (i.e. templates) for diagnostic and prescription process. In reducing the time in data entry, Automatic Knowledge Elicitation Module is employed to support the construction of the template for the entire common and critical symptom for particular diagnosis. For example, after studying the diagnostic decision stored in the traditional EMRS, it is found that fever and running noses are two of the common symptoms in the diagnosis Upper Respiratory Tract Infections (U.R.T.I.). Thus, number of templates regarding to the physician's preferences are set in the system. Details of Automatic Knowledge Elicitation Module are further described in the Section 3.3.2.

Furthermore, an administrative function is employed to govern the information accessibility of the healthcare professionals. It is used to improve the robustness of functions performed in the system and information stored in the databases. For example, users can configure the basic system settings as to define the number and font sizes of all tags, enter system communication server, and so on. Furthermore, various types of searching queries and user authentication are built to supporting data management (including patient data, medical case data and drug data) and query of users. All the searching queries are used to extract only those records that fulfill a specified criterion. On the other hand, user authentication is used to manage the permission of system access, thereby ensuring the data privacy of patient. Furthermore, when the users input incorrect information in the system mistakenly, this module allows the authorized users to manage, update or delete those wrong records. Since the stored information is critical for the use in other modules, details of all the modifications are stored in a log file to ensure data validity and traceability; as a result, this comprehensive audit trail allow users to know when a change was made to a patient record, what was changed and who changed it.

All in all, TEMRS is designed that physicians can easily communicate with the MedicPDSS to input the diagnosis and therapeutic decisions in each patient visit. With the user-friendly and flexible template for each diagnosis, physicians can interact and query the system effortlessly. Furthermore, this module can help the medical organization in three different aspects:

(i) To retrieve the patient's medical information stored in the TEMRS;

(ii) To systematize the medical diagnostic process; and

(iii) To heighten the medical prescription efficiency.

3.3.2 Automatic Knowledge Elicitation Module

Knowledge is key to improving quality of medical judgment of physicians. However, researchers and practitioners are still striving for more effective ways to capture tacit knowledge and transform it into a machine readable form so as to enhance knowledge sharing. To address this knowledge gap, the Automatic Knowledge Elicitation Module is proposed to capture the medical knowledge embedded in the TEMRS. The Automatic Knowledge Elicitation Module implements two processes. First is the knowledge acquisition process that elicits knowledge automatically from the TEMRS, capturing the physician's decision logic of medical treatments. This method of automatic knowledge acquisition is an indirect approach that does not require any human intervention. Furthermore, information on the medical cases stored in EMR are hence transformed into a uniform structure, i.e., in the Extensible Markup Language (XML) format. In the second process – knowledge representation – individual physicians' decision logics (that have been explicitly represented in the XML format) are visualized by the concept mapping technique, which presents information graphically making it easy to understand.

Figure 3.5 shows the architectural framework of Automatic Knowledge Elicitation Module. This module consists of modules for vector formation, case tokenization, rule construction and map construction. Generally, patients' personal and medical information, as well information entered by physicians during consultations is captured in the database through a TEMRS interface. All these information can then be filtered and classified to discover critical cases that are useful for identifying decision rules. After extraction and tokenization of useful cases, rules employed by individual physicians for diagnosis of each type of disease is then formulated. An example of these rules is "Physician A diagnosed the disease U.R.T.I. with the concepts Fever, Headache, Cough, and Stuffy Nose". These decision rules represent knowledge of a specific physician, and the relationships between the concepts used and the diagnosis made. To characterize the importance of various concepts used in diagnosing a particular disease, related rules are collected and hence used to construct concept maps that associate with diagnosis of that disease. As a result, common concepts used by individual physicians can be analyzed and identified. For example, it may discover that Cough is a common factor (symptom) of U.R.T.I. because around 95% of U.R.T.I. cases involve Cough as reflected in the records of diagnoses made by a pool of physicians.

3.3.2.1 Vector Formation

During each visit, the physician uses his knowledge and experience to make medical judgments on the patient. Each keyword that exists in the medical record represents an essential concept contributing to the diagnosis and treatment (i.e. medicines being selected) recorded. A patient's basic personal information is recorded at his first visit. The other pieces of information are collected by physicians and stored as a case at each visit of the patient. Consider the case of an eight-year old patient who has a fever, headache, cough, and is prescribed the medicine paracetamol. The physician diagnosed that this patient has the disease Influenza. Then, the concepts of this case are "Infant", "Age = 8", "Fever", "Headache", "Cough", and "Paracetamol". Thus, a vector that represents a specific diagnosis as determined by a physician is formulated as follows:

$$P_a = \{Concept_1, Concept_2, Concept_3, \ldots, Concept_n\} \rightarrow D_j \qquad (3.1)$$

where

P_a is physician *a*

D_j is diagnosis *j*

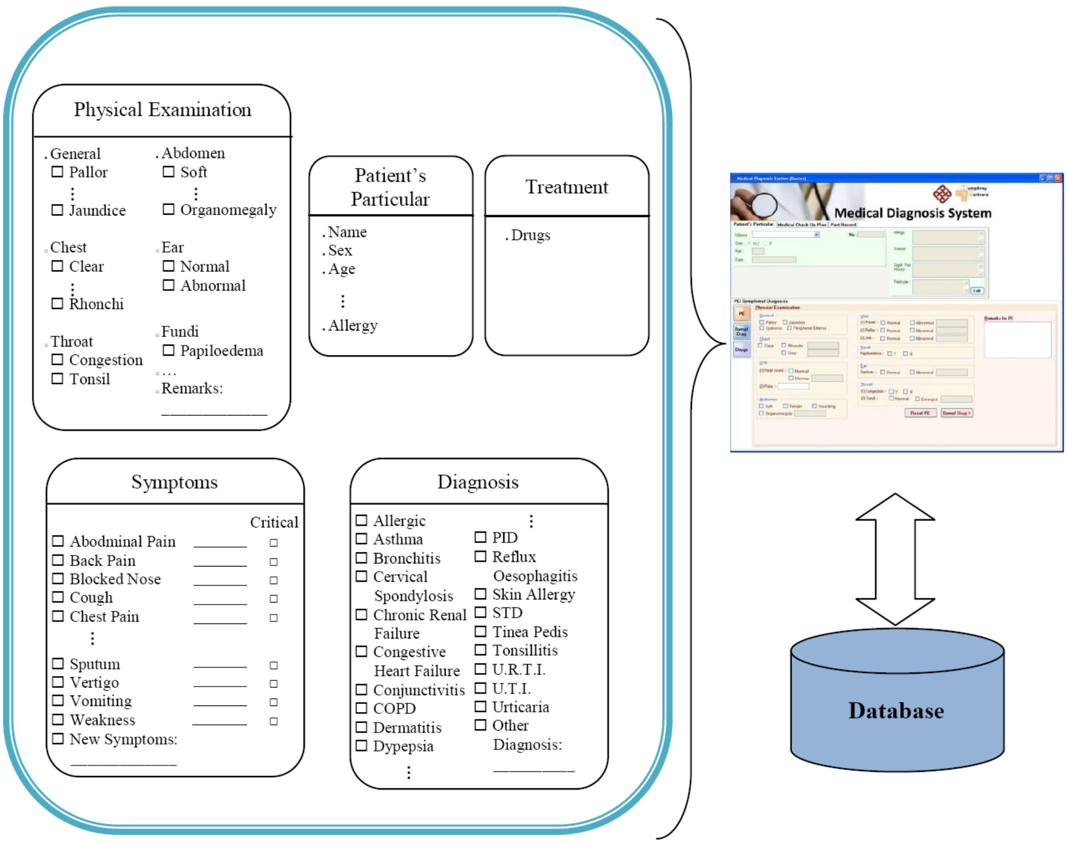


Figure 3.4 – Elements Stored in the EMR

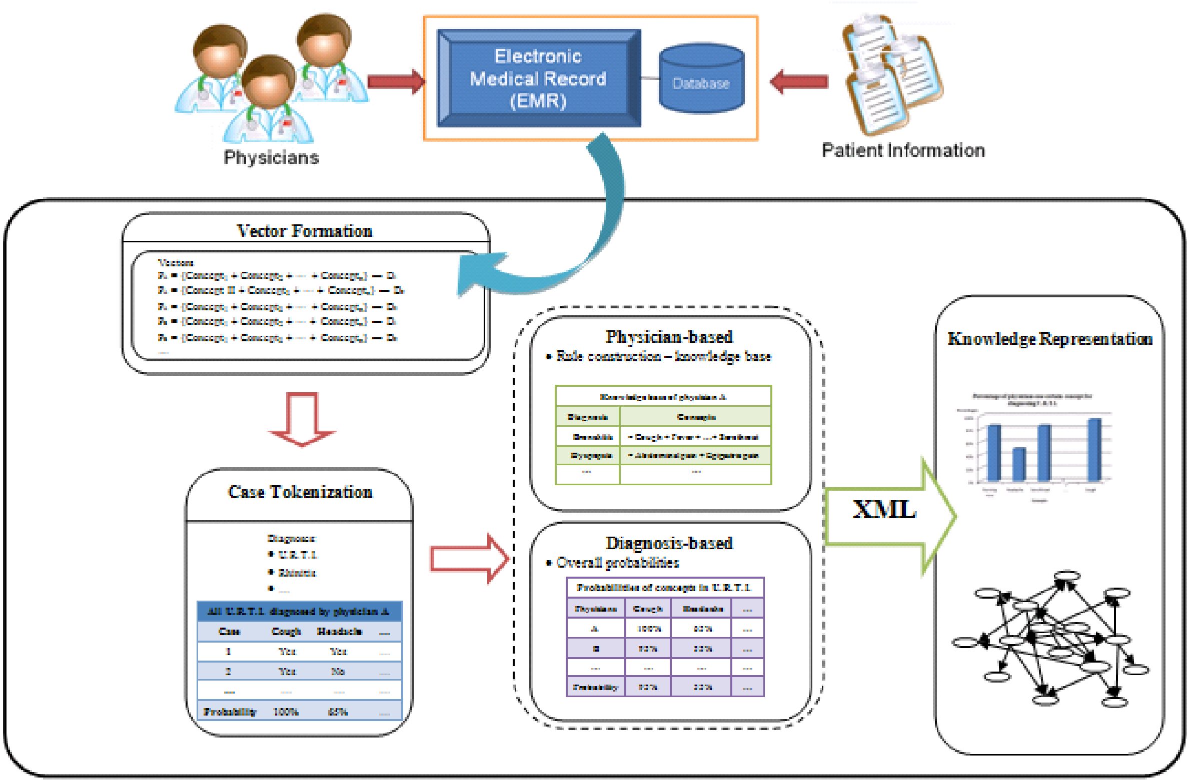


Figure 3.5 – Framework of the Automatic Knowledge Elicitation Module

In the above example, P_A represents a physician, Dr. A, while D_j represents a specific diagnosis, such as Influenza. $Concept_1$, $Concept_2$, $Concept_3$, …, $Concept_n$ represent the set of concepts Dr. A used to make the diagnostic or treatment decision. Each vector represents one diagnosis made by one physician. Therefore, a physician will have a number of vectors as he may have made numerous diagnoses in his practice. It is also noted that different physicians may use different sets of concepts to make the same diagnostic decision, because the choice of concepts employed in the deliberation varies with the physician's expertise and clinical experience. Thus, a diagnosis typically has multiple vectors. Several vector examples can be found in Table 3.1.

Table 3.1 – Examples of Case Formation

Vector Examples	**Description**
P_A $=\{Concept_1, Concept_2, Concept_3\} \rightarrow D_1$ $=\{Concept_1, Concept_2, Concept_3, Concept_4\} \rightarrow D_1$ $=\{Concept_2, Concept_3, Concept_4\} \rightarrow D_2$ $=\{Concept_5, Concept_4, Concept_7\} \rightarrow D_3$	Physician A determines diagnosis 1, 2, 3 based on different concepts. The set of concepts used a physician to infer a given diagnosis may vary from one case to another
$P_A = \{Concept_1, Concept_2, Concept_3\} \rightarrow D_1$ $P_B = \{Concept_1, Concept_3, Concept_5\} \rightarrow D_1$ $P_C = \{Concept_1, Concept_2, Concept_3, Concept_5\} \rightarrow D_1$	Diagnosis 1 is determined by physicians A, B and C from different sets of concepts. Concept 1 always exists, whereas other concepts may not apply in some cases, like Concepts 2 and 5 in this example

3.3.2.2 Case Tokenization

Voluminous vectors are formulated from the previous step. Vectors of a certain diagnosis are further tokenized and placed in a table. Such table shows the relationship between the physician and concepts used to diagnose a specific disease. Each physician has a number of tables for different diagnoses. These tables provide a clearer presentation of concepts used on various vectors. Table 3.2 illustrates the sets of concepts used to diagnose - U.R.T.I.

Table 3.2 – The Sets of Concepts Used to Diagnose U.R.T.I.

All U.R.T.I. Cases Diagnosed From Physician A						
Concepts / U.R.T.I.		**Running nose**	**Headache**	**Sore throat**	**…**	**Paracetamol**
Case	**1**	Yes	Yes	No	…	Yes
	2	Yes	No	Yes	…	Yes
	⋮	⋮	⋮	⋮	⋮	⋮
	N	No	Yes	Yes	…	Yes

As shown in Table 3.2, all cases diagnosed as U.R.T.I. by physician A are tabulated and the existence of concepts are shown in binary, either “Yes” or “No”. All value cases of a specific diagnosis made by a physician are listed in various tables similar to Table 3.2. Using the data in each table, the percentages of existence of specific concepts are normalized by counting their frequencies over the total number of cases of such diagnosis to discover their importance by means of probabilistic approach, as computed by Eq. (3.2).

$$P(Concept|Diagnosis) = \frac{P(Concept)P(Diagnosis|Concept)}{P(Diagnosis)} \tag{3.2}$$

where

$Concept$ is the particular concept applied by the physician

$Diagnosis$ is the clinical decision determined by the physician after investigating the patient's complaint

$P(Concept|Diagnosis)$ is the probability that the $Concept$ is involved in making the $Diagnosis$, as determined from all the cases that are related to $Diagnosis$

$P(Concept)$ is the probability that the $Concept$ is applied under all possible hypotheses

$P(Diagnosis|Concept)$ is the probability that the $Diagnosis$ is made given that the $Concept$ is applied

$P(Diagnosis)$ is the probability that the $Diagnosis$ is made under all possible hypotheses. That is,

$$P(Diagnosis) = \sum_{j=1}^{n} P\left(Concept_j\right) P\left(Diagnosis\middle|Concept_j\right) \tag{3.3}$$

Table 3.3 illustrates the determination of the probabilities of the concepts for diagnosing U.R.T.I. (i.e. *P(Concept|U.R.T.I.)*). Iit is noted that physician A frequently uses cough as one of the concepts for diagnosing U.R.T.I. In this way, $P(Cough|U.R.T.I.)$ = 1 (100%). Compared to other concepts, headache is less frequently used by physician A, i.e. 0.85 (85%).

3.3.2.3 Rule Construction

All probabilities are found in the case tokenization process. They are then used to calculate diagnosis-based overall probabilities. Concept maps are thus built based on the rules that physicians used to make diagnostic decisions. These rules are composed of the concepts stored in each vector.

Table 3.3 – Existence of Concepts in U.R.T.I. Cases Expressed as Probabilities

All U.R.T.I. cases diagnosed from physician A						
Concepts / U.R.T.I.		**Running nose**	**Headache**	**Sore throat**	**…**	**Paracetamol**
Case	**1**	Yes	Yes	No	…	Yes
	2	Yes	No	Yes	…	Yes
	⋮	⋮	⋮	⋮	⋮	⋮
	N	No	Yes	Yes	…	Yes
Probability		90%	65%	90%	…	100%

Diagnosis-based Overall Probability

The probabilities of specific concepts used by different physicians in making a given responses are grouped together with the purpose of determining their respective overall probabilities. An example of this calculation is shown in Table 3.4.

Physician-based Rule Construction

Rules that physicians used during clinical decision making are vital in constructing the concept maps and developing the knowledge elicitation system. Each rule involves a specific diagnosis and the set of concepts used to determine that diagnosis. These rules are represented in the form of “Diagnosis = {$Concept_1$, $Concept_2$, $Concept_3$, …, $Concept_n$}”. Some examples of these rules are shown in Table 3.5.

Table 3.4 – Overall Probabilities of Specific Concepts Used in Diagnosing U.R.T.I. Across Multiple Physicians

Probability of Concepts in U.R.T.I. cases					
Concepts / Physicians	**Running nose**	**Headache**	**Sore throat**	**…**	**Paracetamol**
A	90%	65%	90%	…	100%
B	85%	55%	75%	…	95%
⋮	⋮	⋮	⋮	⋮	⋮
X	80%	50%	80%	…	95%
Overall (Column Average)	**85%**	**55%**	**85%**	**…**	**95%**

Physician’s Experience and Area of Expertise

A physician’s level of expertise and the area of his specialty are also considered as parameters to be weighted for the relevance of that physician’s specific diagnostic decision. As stated by Meltzer et al. (2002), the more cases a physician encounters, more proper clinical decisions can be made. Furthermore, the closer the relationship between the area of the physician's specialty and the patient's complaint, the more appropriate medical services and treatment are likely to be provided. A weighting method is suggested to be used in measuring the level of expertise and the specialty of physicians. Experience value is calculated by the Eq. (3.4).

$$Exp(P) = \sum_{n=1}^{k} {Case^{Diagnosis}}_{n} \tag{3.4}$$

where

P is a specific physician

$Diagnosis$ is the diagnosis

$Exp(P)$ is the experience value of the physicians in treating the particular diagnosis

Sternberg and Horvath (1999) suggested that expertise can be divided into six levels, which are layperson, beginner, novice, intermediate, subexpert and expert. In this paper, only four levels of expertise are considered to be appropriate (i.e. junior, senior, specialist and professor) for a healthcare organization. The levels of expertise are determined by the following steps:

Step 1: Identify the maximum number of cases encountered among all the physicians, which is denoted as *Max(No_of_Case)*

Step 2: Represent the lower-bound value of each level of expertise by dividing the *Max(No_of_Case)* incrementally (i.e. junior = *Max(No_of_Case)*/4, senior =

Max(No_of_Case)/3, specialist = *Max(No_of_Case)*/2, and professor = *Max(No_of_Case)*)

Step 3: Formulate an "IF-THEN-ELSE" rule to satisfy the expertise classification, which is shown as below:

IF the number of cases encountered by the physician ≤ *Max(No_of_Case)*/4, THEN it is classified as junior

ELSE IF *Max(No_of_Case)*/4 < the number of cases encountered by the physician ≤ *Max(No_of_Case)*/3, THEN it is classified as senior

ELSE IF *Max(No_of_Case)*/3 < the number of cases encountered by the physician ≤ *Max(No_of_Case)*/2, THEN it is classified as specialist

ELSE IF *Max(No_of_Case)*/2 < the number of cases encountered by the physician ≤ *Max(No_of_Case)*, THEN it is classified as professor

Furthermore, rules are set to measure the relevance of physicians' areas of expertise to a specific diagnosis. For instance, some concepts are highly related to certain expertise areas, such as "Pregnant" to "Obstetrics & Gynaecology" and "Infant" to "Pediatrics". In order to link up these important concepts in cases with physicians' expertise areas, they are extracted and used to form "IF-THEN" rules, which are further used to determine the level of importance of a particular case. Some examples of these "IF-THEN" rules are given in Table 3.6.

Table 3.5 – Examples of Physician-based Rules

	Knowledge base of physician A	
	Diagnosis	**Concepts**
Rules	Bronchitis	= {Cough, Fever, Running nose, …, Sore throat}
	Dyspepsia	= {Abdominal pain, Epigastric pain}
	Gastroenteritis	= {Diarrhoea, Fever, Running nose, …, Vomiting}
	Rhinitis	= {Headache, Running nose, Sputum, …, Sneeze}
	⋮	
	Tonsillitis	= {Cough, Fever, Sore throat, …, Sputum}
	U.R.T.I.	= {Cough, Headache, Running nose, …, Sore throat}

These rules are used to match the relevance of the specialties of physicians with cases, so as to assign appropriate weights to the relationship between the concepts used by a physician in making a diagnosis. As shown in Table 6, if the patient is an infant, a pediatrics specialist is preferred to a physician having other specialties; thus his concept used in U.R.T.I. is much more important than others. Table 3.7 shows the importance ratings of cases by taking into consideration both the physician's level of expertise and the relevance of his specialty.

Table 3.6 – Examples of "IF-THEN" Rules for Specialties

Diagnosis	**IF (concept) involved**	**THEN (1st Related Specialty)**	**(2nd Related Specialty)**
U.R.T.I.	(None of Below)	Respiratory Medicine	
	Pregnant	Obstetrics & Gynaecology	Respiratory Medicine
	Children	Pediatrics	Respiratory Medicine
	⋮	⋮	⋮
⋮	⋮	⋮	⋮

Table 3.7 – Importance of Case

Level of Expertise / Relevance of Expertise Area	**Junior**	**Senior**	**Specialist**	**Professor**
Relevant	Medium	Medium	High	High
Normal	Low	Medium	Medium	Medium

3.3.2.4 Map Construction

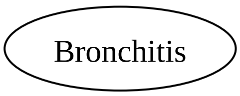
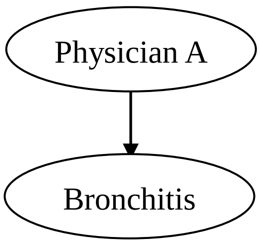
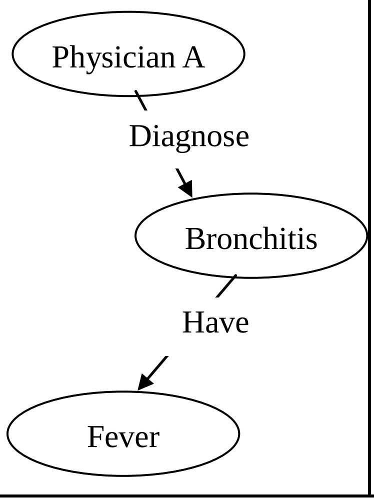
The physician-based rules are presented in concept maps by using XML. First, there is a need to convert the information stored in each case of EMR into specific XML trees automatically.

Each XML tree, as shown in Figure 3.6, represents the knowledge base of a physician, whereas Table 3.8 provides the descriptive information of such trees. In this example, “PA” is the id of the node of physician A, “D1” is the id of the bronchitis node and “S230” is the id of the fever node. The “edge fromID=“PA” toID= “D1”” represents the link between nodes, whereas “level” refers to the number of cases having the fromID and toID. For example, this XML tree shows how physician A diagnoses Bronchitis by taking fever as one of the concepts.

```
<nodes>
  <node id=“PA” level=“1”>
    <locationSgln code=“PA” sgln= “PA”>
      <description>Physician A</description>
    </locationSgln>
  </node>
  <node id=“D1” level=“1”>
    <locationSgln code=“PA” sgln= “PA”>
      <description>Bronchitis</description>
    </locationSgln>
  </node>
      ⋮
  <node id= “S230” level=“1”>
      <locationSgln code=“S230” sgln= “S230”>
        <description>Fever</description>
      </locationSgln>
  </node>
</nodes>
<edges>
  <edge fromId=“PA” level= “2143” toId=“D1”>
  <edge fromId=“D1” level= “31” toId=“S010”>
  <edge fromId=“D1” level= “1” toId=“S060”>
  <edge fromId=“D2” level= “55” toId=“S050”>
  <edge fromId=“D2” level= “1” toId=“S080”>
      ⋮
</edges>
  <nodes />
  <edges />
```

Figure 3.6 – An Example of XML Tree

Table 3.8 – Visual Representation, Concept Map Language, and XML

Visual representation	**Concept Map language**	**XML with description**
	Node	<node id="D1" level="1"> <locationSgln code="D1" sgln= "D1"> <description>Bronchitis</description> </locationSgln> </node> (Bronchitis is noded as an id of "D1")
	Link	<edge fromId="PA" level= "2143" toId="D1"> (Link from Physician A to Bronchitis)
	Relationship	(Physician A applies the Fever concept to determine the diagnosis Bronchitis)

A framework for the overall map construction is depicted in Figure 3.7. Each physician and his diagnosis has a map similar to Figure 3.8, which shows all diagnoses made by physician A and the concepts he used for these diagnoses. These concept maps are named as individual maps. Furthermore, these maps can be combined to produce an aggregated map that includes all physicians, diagnoses and concepts found from the TEMRS database.

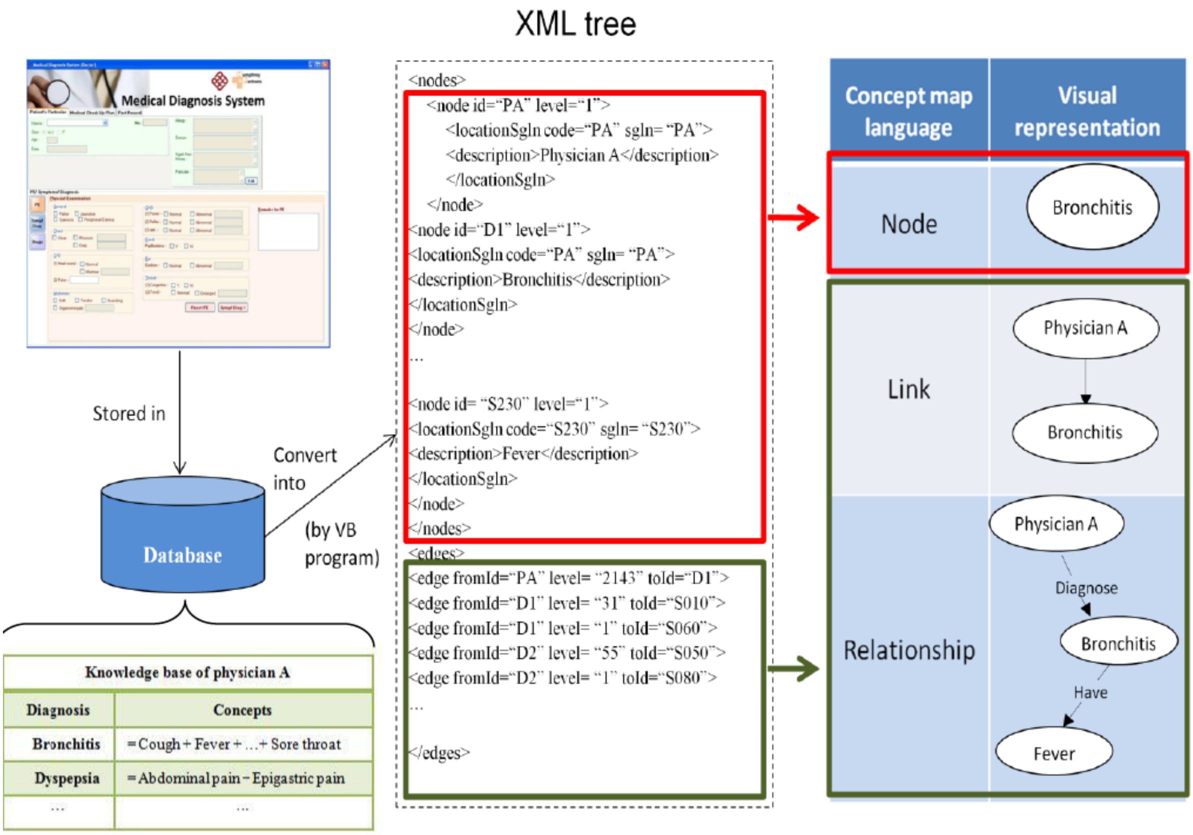


Figure 3.7 – Conversion from TEMRS to Concept Maps via XML Trees

3.3.3 Medical Diagnosis Module

With the generalization of the concept map in the Automatic Knowledge Elicitation Module, the Medical Diagnosis Module is so designed to covert each map into specific diagnostic template for physicians to communicate with the MedicPDSS via the interactive interface. With the user-friendly and easy-to-use graphical user interface, even the physicians, who are not skillful in using commuter, can query the system effortlessly. Furthermore, this module can help the medical organization in three different aspects:

(i) To examine the medical information stored in the TEMRS;

(ii) To streamline the medical diagnostic process; and

(iii) To facilitate the medical prescription efficiency (supported by the Prescription Modeling Module and Risk Surveillance Module).

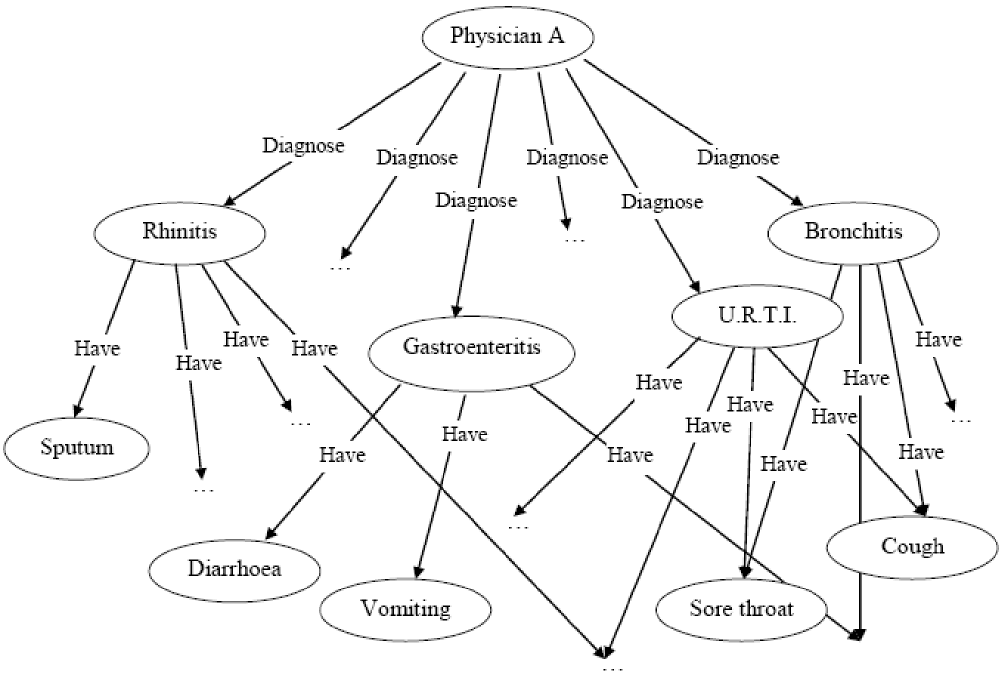


Figure 3.8 – An Example of a Concept Map that Represents All the Concepts Used by Physician A in Making Various Diagnoses

3.3.4 Prescription Modeling Module

3.3.4.1 Concept of 'Micro-view' and 'Macro-view'

In each diagnostic process, the physician may reuse previous solutions in relevant situations to address the new problem. Therefore, we apply the Case-based Reasoning (CBR) approach proposed by Aamodt and Plaza (1994) to specifically retrieve previously experienced cases with information on concrete problem situations and their solutions. As each retrieved case represents a particular patient's medical history on the basis of a physician's specific knowledge of the prescription practices, the solution obtained in the CBR process relates a specific patient to the physician (i.e. patient-centric). When a patient has consulted several physicians in the past, more knowledge in diagnostic and prescription decisions related to that particular patient will have been acquired. The associated network, that formulates a patient-physicians relationship, represents a 'micro-view' in the medical data (Figure 3.9a).

On the other hand, when applying Association Rules Mining (ARM), the prescription patterns of the diagnostic experiences within the organization can be captured and characterized through a quantitative measure. Such statistical approximation expresses the knowledge that is accumulated from all the physicians, thus the solution obtained in ARM depicts a peer-based relationship among the physicians. The associated network, at this time, centers on the characteristics of the whole organization. The physician-physicians (within the organization) relationship, thereby forms a 'macro-view' in the medical data (Figure 3.9b).

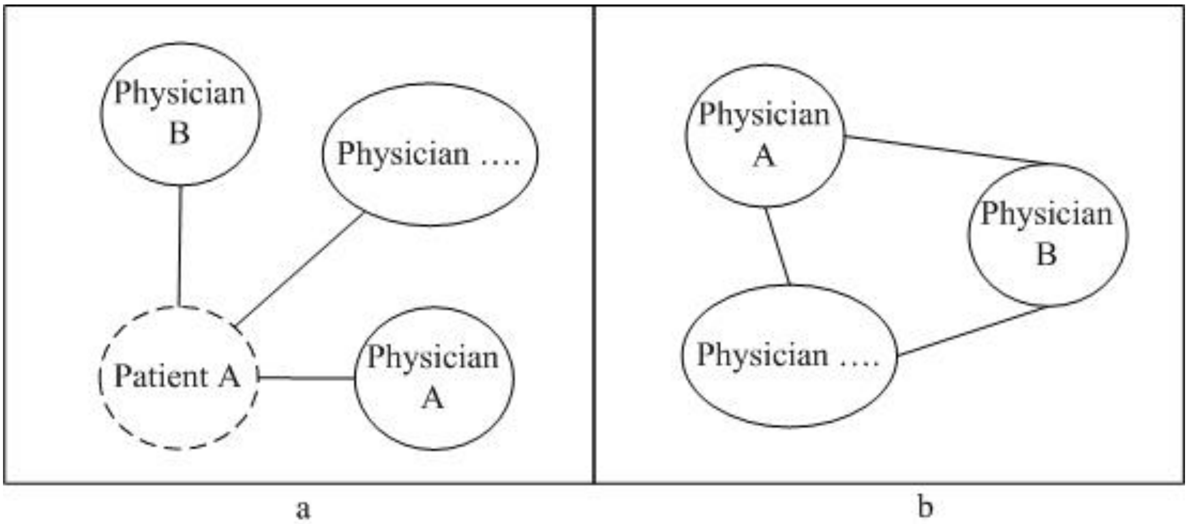


Figure 3.9 – Assuming that a Patient A Has Visited and Diagnosed by Various Physicians, (a) Shows the Patient-physician Relationship in 'Micro-view' and (b) Shows the Physician-physician Relationship in 'Macro-view'

3.3.4.2 Rule-Associated CasE-based Reasoning (RACER) Algorithm

Rule-Associated CasE-based Reasoning (RACER) algorithm is used to support the decision support process in the Prescription Modeling Module. The RACER methodology is mainly composed of three parts: cases retrieval, association rules mining, and suggestions combination. RACER starts from the point where the GP interprets the diagnosis of patient. As shown in Figure 3.10, a new case (the diagnosis) is firstly codified based on the predefined TEMRS. The medical data recorded in TEMRS consists of all the examination data and patient particular information which is voluminous and heterogeneous. It is important to preprocess the data by selecting attributes or features which are useful for prescription making. The codified new case is then processed by comparing with the previous cases retained in the knowledge base. ARM and case retrieval are then applied. ARM is used to extract the most interesting association rules based on support and confidence measure. Weightings are then assigned to the associated medicines. Simultaneously, the most similar cases are extracted based on a similarity measure for cases retrieval. Weightings are then assigned to the retrieved medicines. Then, the weightings of the associated medicines and the retrieved medicines are combined based on a simple rule of combination which is adapted from the Dempster's rule of combination (Dempster, 1968). Based on the combination, a consolidated medicine list is provided as suggestion for the new case. The suggestion is then reviewed and revised by the GP. When the case and results are verified, they are then retained to the knowledge base for future reuse.

Cases Retrieval in RACER

In general, CBR consists of case retrieval, adaption, reuse, and retain. However, Schmidt et al. (2001) discussed that adoption of complete CBR cycle are rather exceptional in the medical field. Zhuang et al. (2009) also mentioned that it is almost impossible to generate adaptation rules to consider all possible important differences between current and former similar cases in medical application. Since the adaptation knowledge is difficult to be acquired, the present study is focused only on the retrieval of similar cases.

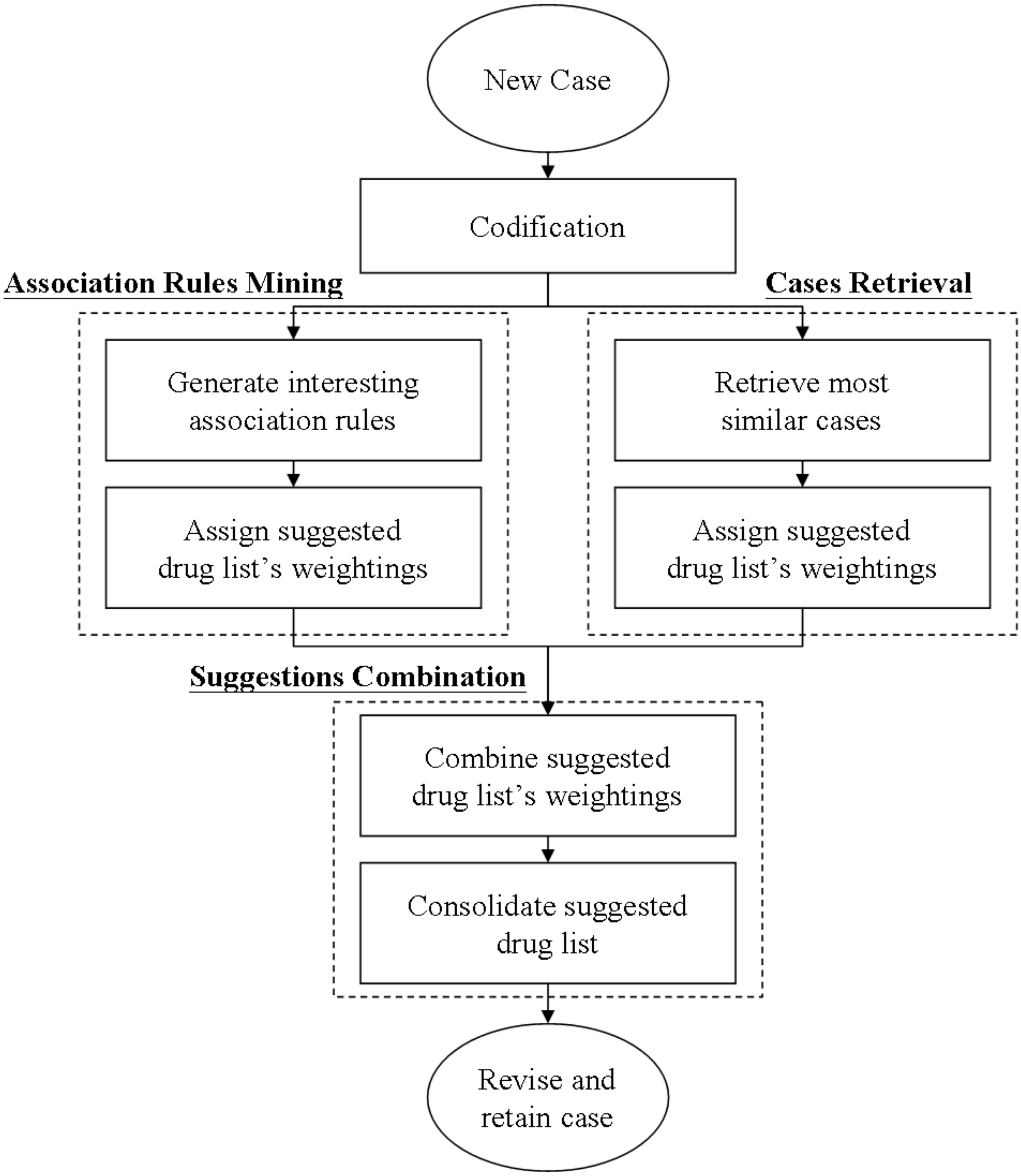


Figure 3.10 – Architecture of RACER Algorithm

A case consists of features for describing the problem and solution. In the present study, the information of diagnosis and patient particular information (such as age and gender) are representing the problem features of a case, while the medicines to be prescribed are described as the solution of the case. Mathematically, each case is represented in the following notation:

Let Ω_E be the set of all cases, Ω_A be the set of all problem features, and Ω_D be the set of all medicines (i.e. the solution attribute), where each case $c \in \Omega_E$, each attribute $a \in \Omega_A$, and each medicine $d \in \Omega_D$. Thus,

$$c = (A_c, D_c) \tag{3.5}$$

where

$A_c \subseteq \Omega_A$ is the set of problem features observed in the case c.

The set $D_c \subseteq \Omega_D$ is the set of medicines to be prescribed for this case.

Medical records are codified and stored as cases in a knowledge base for case retrieval. The present study employs a similarity measure approach, Nearest-Neighbor Retrieval (NNR), for determining the degree of similarity between the new case and old case. This method is used due to its simplicity and good performance in case indexing (Sun and Finnie, 2004). During comparison, the features of a new case are matched to their corresponding features of all cases stored in the knowledge base.

The algorithm of cases retrieval is shown in Figure 3.11. A threshold γ is set for determining the maximum number of similar cases being retrieved. The contributions of the retrieved cases are weighted by their corresponding similarity, so that the similar cases contribute more to the average than the less similar ones. It is accomplished by the following steps:

(i) n most similar cases are retrieved based on Eq. (3.6) and (3.7);

(ii) A list of unique medicines is extracted from the retrieved cases;

(iii) Weightings of the unique medicines are determined based on the occurrence of the corresponding medicines prescribed in each retrieved case, which is shown in equations (3.8) and (3.9)

The similarity for each case is calculated by Eq. (3.6) and (3.7):

$$similarity(c^I, c^R) = \frac{\sum_{i=1}^{n} w_i \times sim(a_i^I, a_i^R)}{\sum_{i=1}^{n} w_i} \tag{3.6}$$

$$sim(a_i^I, a_i^R) = \begin{cases} 0 & \text{if } a_i^I \text{ and } a_i^R \text{ are categorical and } a_i^I \neq a_i^R \\ 1 & \text{if } a_i^I \text{ and } a_i^R \text{ are categorical and } a_i^I = a_i^R \\ \dfrac{a_i^R}{a_i^I} & \text{if } a_i^I \text{ and } a_i^R \text{ are numerical and } a_i^I > a_i^R \\ \dfrac{a_i^I}{a_i^R} & \text{if } a_i^I \text{ and } a_i^R \text{ are numerical and } a_i^I < a_i^R \end{cases} \tag{3.7}$$

where

c^I and c^R represent the new case and the old case respectively

a_i^I and a_i^R represent the *i*-th feature value of the new case and the old case respectively

the similarity function $sim(a_i^I, a_i^R)$ computes the similarity between a_i^I and a_i^R, and w_i represents the feature weighting for each *i*-th feature.

Input: Examination data of disease determined by the GP

Output: A set of medicines in a ranking list

Preprocessing

Set the threshold γ as the maximum number of cases retrieved

Set the weightings w_i for each i th feature

Case retrieval algorithm

Do while (a new case is ready)

  Trigger Similarity Analysis

    Compute similarity for each cases in the knowledge base

  End Trigger

  Sort the cases by their similarities in descending order

  Extract the first γ most similar cases

  Extract the unique medicines list from the retrieved cases

  Trigger Weighting Assignment

    Compute the weights for each medicine in the unique medicines list

  End Trigger

  Sort the unique medicines list by their weights in descending order

End Do

Report the results

Figure 3.11 – Algorithm of Cases Retrieval in RACER

The weightings of the unique medicines are calculated by equations (3.8) and (3.9):

$$W_j^{cbr} = \frac{\sum_{i=1}^{n} similarity(c_i^I, c^R) \times s_{i,j}}{\sum_{i=1}^{n} similarity(c_i^I, c^R)} \tag{3.8}$$

where W_j^{cbr} is the weighting of medicine j

n is the number of retrieved cases

$similarity(c_i^I, c^R)$ is the similarity between the i-th retrieved case and the new case

$s_{i,j}$ is the occurrence of medicine d_j being prescribed in i-th case c_i. $s_{i,j}$ is determined by the following equation:

$$s_{i,j} = \begin{cases} 1 & \text{if } d_j \text{ is prescribed in } c_i \\ 0 & \text{if otherwise} \end{cases} \tag{3.9}$$

As a result, a unique medicines list with weightings is generated as the suggestions of CBR.

Association Rules Mining in RACER

A standard association rule consisting of an antecedent (i.e. X) and consequent (i.e. Y) is implicated as follow:

$X \Rightarrow Y$ where $X, Y \subset I$ is a itemset (3.10)

In the present study, the ARM approach aims to discover interesting association rules between the medicines and the problem features of the new case by analyzing the previous cases stored in the knowledge base. Thus, X is the set of problem features of the new case, and Y is the set suggested medicines. The interestingness of a rule is measured by its Support (i.e. the probability that the antecedent and consequent occur among cases in the knowledge base) and its Confidence (i.e. the conditional probability that the consequent occurs given the occurrence of the antecedent). A rule is considered as interesting when it satisfies both the minimum thresholds of support and confidence. Support and confidence are determined by Eq. (3.11) and (3.12), respectively.

$$Support(X \Rightarrow Y) = \frac{\text{Number of cases containing both } X \text{ and } Y}{\text{Total number of cases}} \tag{3.11}$$

$$Confidence(X \Rightarrow Y) = \frac{\text{Number of cases containing both } X \text{ and } Y}{\text{Number of cases containing } X} \tag{3.12}$$

The algorithm of association rules mining is shown in Figure 3.12. Apriori algorithm (Agrawal and Srikant, 1996) is applied to identify the associations. It is the best-known algorithm to mine association rules. It uses a breadth-first search strategy to counting the support of rules and uses a candidate generation function which exploits the downward closure property of support. It is applied in the present study for speeding up the mining process. Similar to the consolidated of similar cases in CBR, the mined rules are consolidated to extract a list of unique medicines. The weightings of the medicines in the list are determined by the maximum confidence of the rules associated with the corresponding medicines, which is shown in Eq. (3.13):

$$W_j^{arm} = Max\{Confidence(a_1 \Rightarrow d_j), Confidence(a_2 \Rightarrow d_j), \ldots, Confidence(a_m \Rightarrow d_j)\} \quad (3.13)$$

where

W_j^{arm} is the weighting of medicine *j*

a_i is the *i*-th problem feature

d_j is the *j*-th medicine

m is the number of problem features.

Suggestions Combination

The algorithm of combing the suggestions of CBR and that of association rules mining is shown in Figure 3.13. Before combining the suggestions, the weightings of the suggestions are needed to be normalized by Eq. (3.14):

$$N_j = \frac{W_j}{\sum_{i=1}^{n} W_i} \quad (3.14)$$

where

N_j and W_j are the normalized weighting and suggested weighting of medicine *j* of CBR (or ARM) respectively

n is the number of medicines in the suggested medicines list of CBR or ARM

Input: Examination data of disease determined by the GP

Output: A set of medicines in a ranking list

Preprocessing

Set the minimal support α and minimal confidence β

Association rules mining algorithm

Do while (a new case is ready)

    Trigger Apriori algorithm

        Measure the support of the features of the new case

        Remove the features that do not satisfy α

        Measure the support of the medicines

        Remove the medicines that do not satisfy α

        Trigger Rule Extraction

            Associate filtered medicines with the filtered features

            Measure the support and confidence of the association rules

            Remove the rules that do not satisfy α and β

        End Trigger

    End Trigger

    Extract the unique medicines list from the associated rules

    Trigger Weighting Assignment

        Compute the weights for each medicine in the unique medicines list

    End Trigger

    Sort the unique medicines list by their weights in descending order

End Do

Report the results

Figure 3.12 – Algorithm of Association Rules Mining in RACER

A simple rule of combination is proposed to integrate the normalized weightings of CBR and

ARM into one single solution. The combination method is adapted from the Dempster's rule of combination (Dempster, 1968), which compensates the missing medicines in the solutions of CBR or that of ARM, and updates the weightings of the medicines when new evidences are available. The combination weights of the medicines are calculated from the aggregation of normalized weightings of CBR and ARM as shown in Eq. (3.15):

$$N_i^{com} = \frac{w_i^{cbr} N_i^{cbr} + w_i^{arm} N_i^{arm}}{w_i^{cbr} + w_i^{arm}} \tag{3.15}$$

where

N_i^{com}, N_i^{cbr}, and N_i^{arm} are the combined weighting of medicine *i* normalized weighting of CBR of medicine *i*, and normalized weighting of ARM of medicine *i*, respectively

w_i^{cbr} and w_i^{arm} are weighting of CBR and ARM for combination of medicine *i*.

The final solution is then sorted by the combined weightings of the medicines in descending order.

Rule-based Results Aggregator

The objective of the rule-based results aggregator is to match the results between CBR and ARM. In the matching algorithm, the ranking of drugs is represented in the form of three different 'IF-THEN' statements as shown in Figure 3.14.

Input: A set of medicines in a ranking list from CBR and a set of medicines in a ranking list from ARM

Output: A set of medicines in a ranking list

Preprocessing

Set the threshold γ as the maximum number of medicines of the output medicines list

Suggestions combination algorithm

Do while (the input is ready)

Normalize the weighting of medicines list of CBR

Normalize the weighting of medicines list of ARM

Combine the weighting of medicines lists of CBR and ARM

Sort the unique medicines list by their weights in descending order

Extract the first γ medicines

End Do

Report the results

Figure3.13 – Algorithm of Suggestion Combination in RACER

The first statement classifies the drugs which appear in both CBR and ARM, into Rank A, which is the top ranking recommended list, for the physician's consideration. However, if the drugs do not match any instances (neither in CBR nor in ARM), they will be classified as Rank C. For the remaining prescribed instances (the drugs appear either in CBR or ARM), they will be grouped into Rank B. With each medicine has a weight calculated in the Suggestion Combination, therefore the medicine in each rank will be ranked according to the weight in descending order. An example of such illustration can be found in Figure 3.15. Furthermore, the prescribing pattern of the physician can even be visualized and compared with the pool of prescriptions of many physicians. The physician can learn from this comparison. The entire rule-based results aggregator is repeated until all the drugs are categorized into corresponding areas. Thus, the final solution in the combined medication list is represented as follow:

$$Finalsolution = \begin{cases} drug_{RankA} & if \quad (drug_{RankA} \in cbrsolution) \text{ and } (drug_{RankA} \in armnsolution) \\ drug_{RankB} & if \quad \text{otherwise} \\ drug_{RankC} & if \quad (drug_{RankA} \notin cbrsolution) \text{ and } (drug_{RankA} \notin armsolution) \end{cases} \tag{3.16}$$

Input: Results in CBR and ARM

Output: A set of medicines in three different ranking list

Rule-induction matching algorithm

For Each (drug name)

    If (the drug name in both CBR and ARM) ***Then***

        (put the drug name into Rank A List)

    Else If (the drug name in either CBR or ARM) ***Then***

        (put the drug name into Rank B List)

    Else If (the drug name in not in either CBR or ARM) ***Then***

        (put the drug name into Rank C List)

   End If

End For

Report the results

Figure 3.14 – Algorithm of Rule-based Results Aggregator in RACER

In Prescription Modeling Module, the appropriate drug choice is optimized concurrently with the matching algorithm and illustrated as a ranking list to promote the flexibility and possibility of considering both individual behavior (i.e. micro-view) and collective behavior (i.e. macro-view). Because of the complex nature of prescribing, the recommended medicine selection list serves only as a reference for physicians which they can use for quick identification of the relevant medicines from past experience. The physician can deviate from the recommendations at any time as they have complete autonomy; thus the final decision still rests with the individual physician.

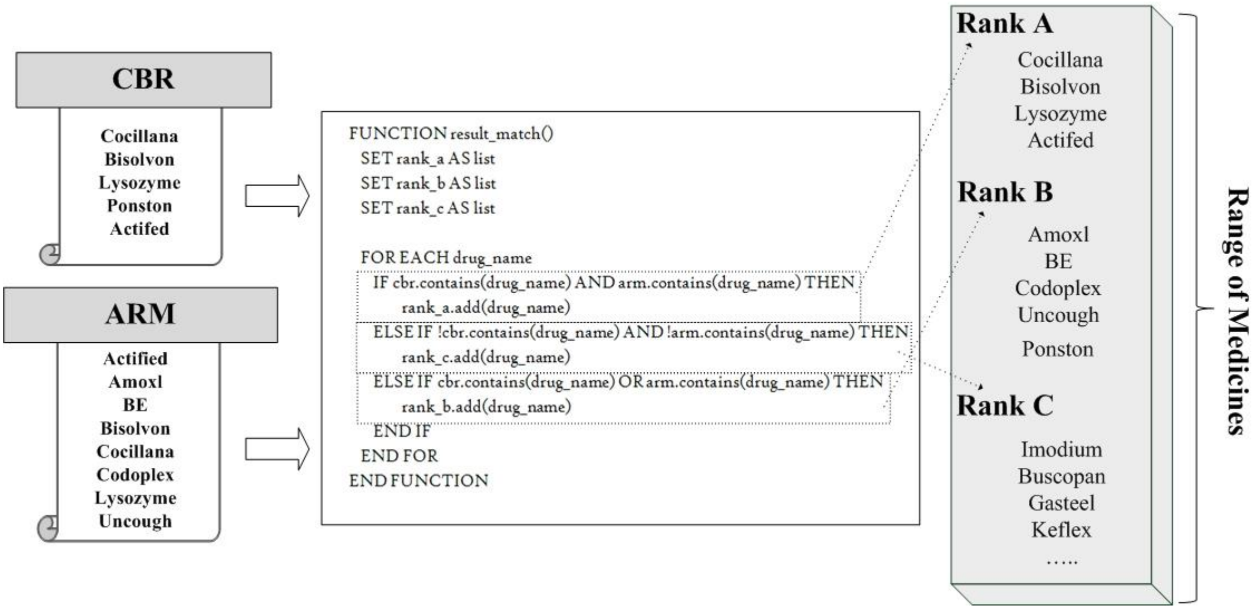


Figure 3.15 – Rule-based Results Aggregator

3.3.5 Risk Surveillance Module

After the decision choices made by the physicians, the Risk Surveillance Module is used to detect the drug-drug interaction within the selected medicines, so as to ensure the medicines prescribed are correct and safe to the patient. According to Bell and Sethi (2001), physicians claimed that the online medical journal articles offer them with a channel for quick and 24-hour access to drug information. Therefore, a web information retrieval approach aims to provide medical information from within the collection that are relevant to an arbitrary user information need, communicated to the system by means of a one-off, user-initiated query (Manning et al., 2008). Contrasting to manually web searching of drug interaction, this module can automatically discover and identify updated drug information from various trusted medical databases (such as MEDLINE).

The Risk Surveillance Module makes use of text and web mining techniques to identify the interaction rules from the literature. It is used to support Medical Diagnosis Module by raising alerts of drug contraindication (such as drug-drug interaction and food-drug interaction) and, if necessary, suggesting alternative drugs for physician's consideration. All related drug information is captured from various trusted medical databases and then automatically decoded as sets of interaction rules that are stored in the system's database. These rules are validated by authoritative medical practitioners before they are adopted as decision support information within the module. An example of such interaction rules is 'IF $Drug_1$ = Epiklor and $Drug_2$ = Lomotil THEN Interaction=YES'. Supported by these drug interaction rules, alerts of drug contraindication are generated upon completion of the selection of medicines. If an interaction exists, information, such as degree of interaction (i.e.

major, moderate, and minor), cause of contraindication, and mitigation of the risk, are displayed. Otherwise, the message ‘No Interaction is Found!’ is shown.

This module is supported by a Drug Information Extraction algorithm and the overall architecture and features of this module, are constructed by, IDEF0, and are presented in Figure 3.16. In this module, two processes are involved, including the Keyword Extraction Process (KEP) and Drug Information Classification Process (DICP). KEP helps to retrieve, prepare and preprocess the web information collected from the online journal articles for further processing and analyzing in the latter modules. In DICF, the Naïve Bayes Classifier (Lewis, 1998; McCallum and Nigam, 1998) is adopted to help identifying and classifying the retrieved drug information into different categories (i.e. interaction related or non-interaction related). Furthermore, a medical database, which is a medical library or medical dictionary, is presented to support the entire functioning of the system. There are two purposes for its existence. First, it is used to check up medical jargons that appeared in the retrieved information. Second, it is used to link up medical synonym, for example the word “fever” and “hot”, in order to help increase the system’s overall performance and accuracy.

3.3.5.1 Keyword Extraction Process (KEP)

KEP provides the methodology to retrieve web information from online medical articles, then process the information into readable and suitable format for later classification and analysis.

A0 Risk Surveillance Module
- **A1** Keyword Extraction
  - **A11** Web Crawling
  - **A12** Web Content Preprocessing
    - **A121** HTML Tag Removal
    - **A122** Stop Words Removal
    - **A123** Stemming
  - **A13** Term Frequency-Inverse Document Frequency (TF-IDF) Weight Measurement
- **A2** Drug Information Classification
  - **A21** Naïve Bayes Classifier Building
  - **A22** Interaction Rules Building

Figure 3.16 – IDEF0 Architecture of Risk Surveillance Module

Web Crawling

Web crawling is the process by which pages from the web are gathered and hence, indexed to support a search engine (Cothey, 2004). It is also referred as a spider for bulk downloading of web pages (Olston and Najork, 2010). The objective of web crawling is to quickly and efficiently gather as many useful web pages as possible, together with the link structure that interconnects them (Manning et al., 2008). In KEP, the web crawler is responsible for fetching useful web information that could highly match with the health professionals' query need. The web crawler obtains suitable web pages and web contents from time to time, within several web-based medical databases. The time period for the web crawler to gather web information, for example, operating once a week; and also the medical database sources are defined and decided by the physicians. Finally, the fetched materials are stored into the information repository and brought to the next step for further processing. Figure 3.17 illustrates an overview of the web crawling process.

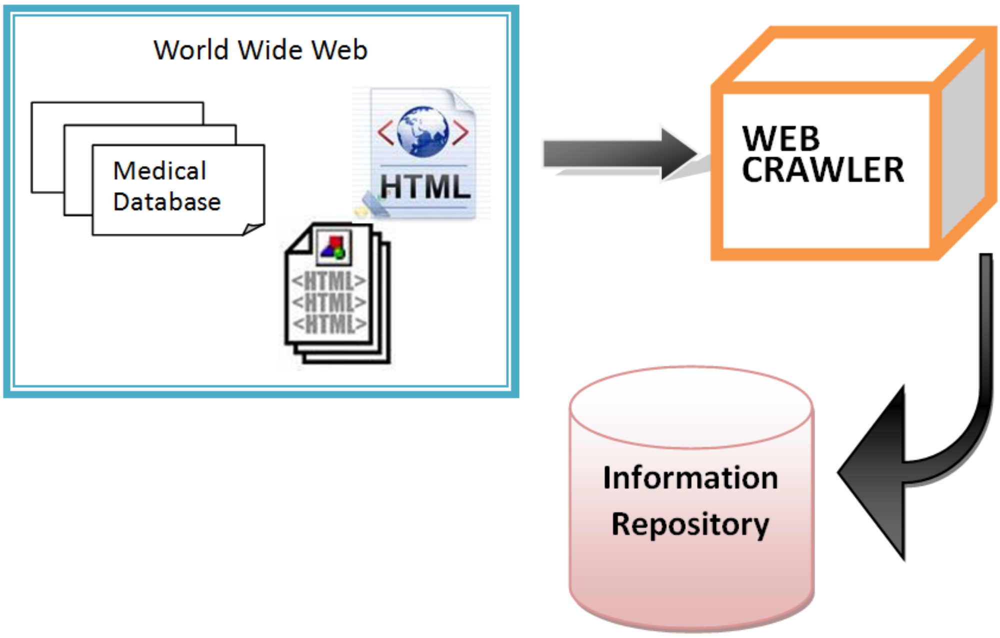


Figure 3.17 – Web Crawling Process

Web Content Preprocessing

After retrieving the useful web information from the web crawling process, several preprocessing steps are conducted to these web documents. All the figures and tables appearing in the web documents are first removed; then three preprocessing elements, including HyperText Markup Language (HTML) tag removal, stop words removal and words stemming, are then conducted. In order to improve the performance of the text retrieval and classification, the preprocessing steps are critically important, as demonstrated in the study of Yang and Chute (1994). In this study, three preprocessing processes are used:

(i) HTML Tag Removal

Since the web document displayed in form of HTML, therefore the disposal of some standard web pages components are conducted. The most common components found are the HTML tags. By interpreting the source code of the web page, all HTML tags (such as <html>, <body>, <p>, <b>, etc.) are being eliminated. The texts wrapped by the HTML tag are taken out, and then generated into plain text file format and being brought to the next preprocessing element. Figure 3.18 illustrates the procedures done in content extraction.

(ii) Stop Words Removal

The second element is stop words removal which is applied to reduce the noisy information and to improve text processing accuracy (Chakrabarti et al., 2003). Stop words are words that rarely contribute useful information in terms of document relevance. They are functional words that do not carry any meaning, including articles, prepositions, conjunctions, and some other high frequency words. Examples of these stop words are the, a, in, of, and, it and this. The assumption of stop word removal is that by ignoring the non-informative functional words, assessment of contents of natural language can be facilitated since meaning can be conveyed more clearly, or interpreted more easily (Patwardhan and Pedersen, 2006).

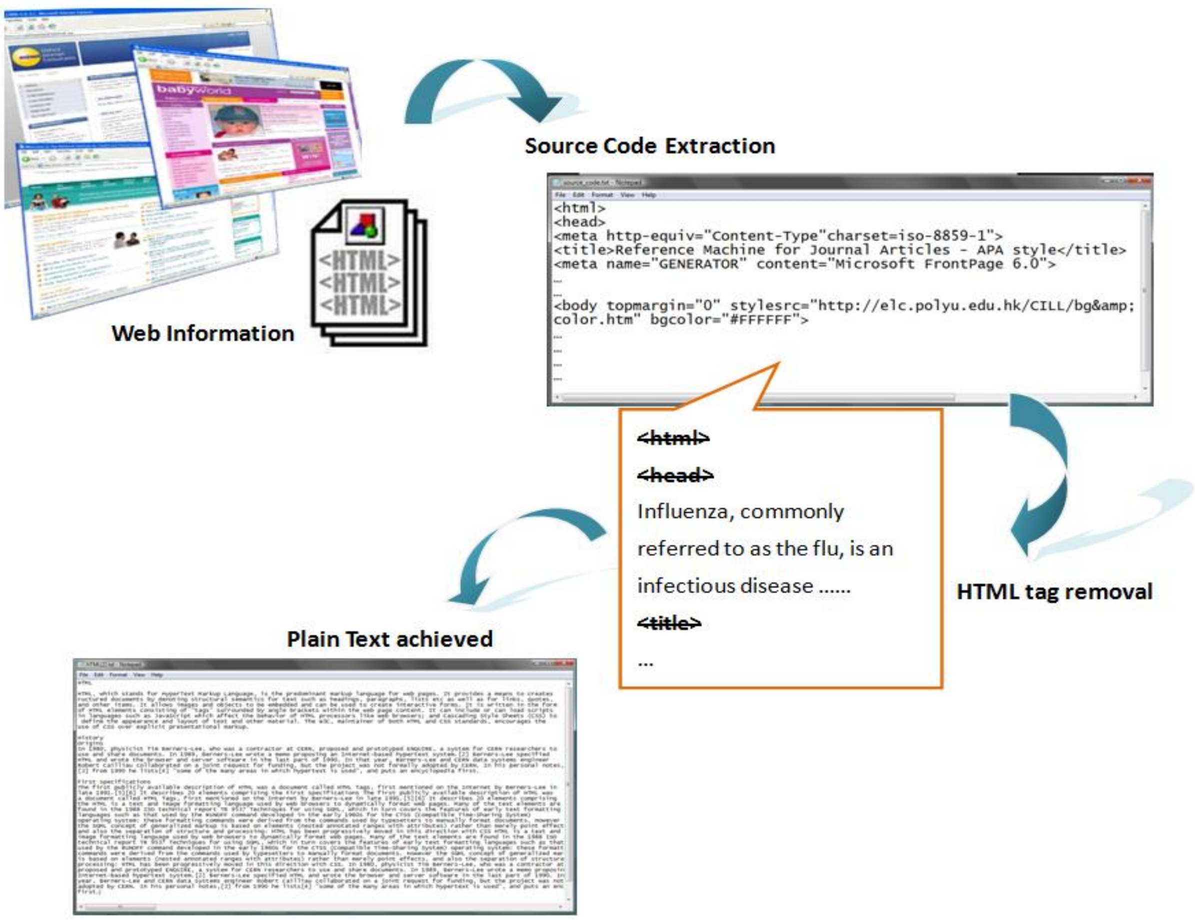


Figure 3.18 – Procedures Done in Content Extraction

(iii) Stemming

The last preprocessing element is words stemming. As it is necessary to avoid the influence of syntactical features and tenses of the English language when identifying and extracting keywords, word stemming (Salton, 1989) is done to reduce inflected or derived words to their stem, base or root form. For example, a stemming algorithm for English should stem the words computation, computing, computes, computed, computational, computable, computationally and computers to the root word, compute. It is proven that words stemming has the capability to reduce the redundancy and dimension of the document space representation in an automatic text processing system (Zhan et al., 2009). Table 3.9 listed some examples of rules in words stemming:

Table 3.9 – Examples of Rules in Words Stemming

Rules	**Examples**
If the word ends in *ed*, remove the *ed*	vaccinated → vaccinat
If the word ends in *ing*, remove the *ing*	coughing → cough
If the word ends in *ly*, remove the *ly*	seriously → serious
If the word ends in *ious*, remove the *ious*	infectious → infect
If the word ends in *es*, remove the *es*	viruses → virus

TF-IDF Weight Measurement

The Term Frequency-Inverse Document Frequency (TF-IDF) weight calculation is a statistical measure used to evaluate and weight the importance of a term or a word to a document within the category collection. With higher TF-IDF weight, the more important a word is towards the document (Aizawa, 2003). The TF-IDF weight can be divided into two parts, the Term Frequency (TF) part and the Inverse Document Frequency (IDF) part.

The TF part of the weighting scheme indicates the number of frequency that a word occurs in a document; while the IDF part measures the percentage of all documents within the category collection that contain the given word, thus measures the general importance of the word (Radev et al., 2004). The TF-IDF is regarded as a more comprehensive and accurate weight used in text mining because the IDF factor is incorporated to diminish the weight of terms that occur very frequently in the collection, and at the same time to increases the weight of terms that occur rarely. As a result, a high weight in TF-IDF is reached by a high term frequency in the given document, whereas a low document frequency of the term in the whole category collection to avoid biased results (Wu et al., 2008).

In order to calculate the TF-IDF weight of each word in a document, first, each term appeared in the preprocessed document is extracted and viewed as a string. These strings are then inputted into a table to generate a "list of document terms". Finally, each term corresponding TF-IDF weight is calculated with the formula stated as Eq. (3.17):

$$(tfidf)_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \times log\frac{|D|}{|\{d: t_i \in d_j\}|} \qquad (3.17)$$

where

$n_{i,j}$ is the number of considered terms t_i appeared in the web documents d_j,

$|D|$ is the total number of web documents in the category collection,

$|\{d: t_i \in d_j\}|$ is the number of web documents where the term t_i appears.

To clearly illustrate the procedures for TF-IDF weight calculation, consider a medical document containing 1000 words wherein the word “swine” appears 10 times. Following the Eq. (3.17), the TF for “swine” is then (10 / 1,000) = 0.01. Now, assumed that there are 20 million documents and “swine” appears in four thousand of these. Then, the IDF is calculated as log(20,000,000 / 4,000) = 3.699. The TF-IDF weight of “swine” is hence 0.03699 (i.e. 0.01 x 3.699 = 0.03699).

By applying the TF-IDF weight measurement, each web document can be presented as a vector with one component corresponding to each term in the knowledge repository (i.e. other retrieved documents). A vector $\vec{V}$ for each web document d can be represented as:

$$\overrightarrow{V_d} = (tfidf_{1,d}, tfidf_{2,d}, tfidf_{3,d}, \dots, tfidf_{N,d}) \tag{3.18}$$

By summarizing all the keyword extracted in the repository, each value in vector corresponds to the computed TF-IDF value for the term in the document. For those terms that do not occur in a document, their weight is zero. As stated by the study of Golder and Huberman (2006), they argued that the position of a tag and its frequency are related. This implies that the frequently used tags will appear before less frequently used tags. Therefore, in this study, we extend the TF-IDF weight measurement by considering the tag structure in the web pages. In the science of document classification, a tag can be classified an important, a minor, an ignorable or a statistic tag. By assignment a specific weights regarding the tag type, the TF-IDF weight measurement can take the word marked as important tag to have a higher weighting. Table 3.10 summarizes the weightings of the frequently used tags. As a result, the TD-IDF weight measurement can be modified as follows:

$$(tfidf)_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \times log\frac{|D|}{|\{d: t_i \in d_j\}|} \times tagw_i \qquad (3.10)$$

where

$tagw_i$ is the tag weight of particular term *i*

Table 3.10 – Weights of Frequently Used HTML Tags

HTML Tag	**Weight**
<Title>	3
<H1> - <H4>	3
Font Size > 3	3
<B>, <I>,<U>,<Strong>,<Big>	2
<A Herf>	2

3.3.5.2 Drug Information Classification Process (DICP)

DICP initiates a document classifier to categorize the retrieved articles or documents, which are stored in the information repository, into predefined dimensions. The use of Naïve Bayes (NB) Classifier is proposed for document classification and its detailed usage is discussed in this section.

Naïve Bayes Classifier

NB Classifier is a probabilistic model based on Bayes Theorem to calculate the characteristics of a document using keyword and joint probability of a document category. According to Xhemali et al. (2009), NB models are popular in machine learning applications, due to their simplicity in allowing each attribute, i.e. string mentioned earlier, to contribute towards the final decision equally and independently from the other attributes, and this simplicity is equates to computational efficiency.

NB is a supervised leaning model which training is required in the building phase. During the training, a random data sample is used to test the proposed model, by comparing the predicted category with the real category to which the articles belong. Then, during classification stage, NB analyses the training set with the retrieved articles and compute probabilities about the matched category features found. The NB classification algorithm is developed based on the standard Bayes rule defined in equation (3.19):

$$\underset{n}{\operatorname{argmax}}\{P(C_n\,|d)\} = \frac{P(d|C_n) \times P(C_n)}{P(d)} \tag{3.19}$$

where

$P(C_n)$ = the prior probability of category n

d = the new document to be classified

$P(d|C_n)$ = the conditional probability of the test document, given category n

NB classifier is chosen to be adopted in the proposed system among the different text classification model because with reference to some research studies (Lin, 2009; Isa et al., 2009; Xhemali et al., 2009), NB strengths include to achieve satisfactory classification accuracy in a relatively short processing time; simplicity of the Bayes formula which then requires a relatively small number of training data and shorter training time; and the straightforward calculation and computation required in the building and classification process. Therefore, although NB model has been reported less accurate then Support Vector Machines (SVM), it is still be regarded as an “ideal” model.

Adoption of Naïve Bayes Classification

In order to implement NB classification technique to sort articles into distinct categories, two steps, training stage and classifying stage, are involved:

(i) Training Stage

Before utilizing a NB classifier for classification in the proposed system, it is required to build a NB model that can effectively analyze and match up the features found in the training collection set and those would appeared in the retrieved testing articles. To do so, training is needed for the NB classifier. In the training, a NB trainer is developed to analyze a set of articles that have been well organized and categorized into each of the defined category. It has to compare the contents and identify the keywords and features that appeared in each category, and to build a list of words with their occurrence for each category. Then, it would be able to match and compute probabilities about the feature-category pairs found in the newly retrieved articles to identify the right categories during classifying stage. Thus, the training results are crucial for the NB classifier to make intelligent decisions in classification stage. And in each training process, distinct keyword or features become more strongly associated with the different categories. The training is iterative and stops when the rate of correct classification is superior to a certain threshold.

(ii) Classifying Stage

The NB classifier performs the classification tasks starting with the first step, analyzing the text article by extracting keywords from the articles. This step has already been done in the previous TF-IDF weight calculation. Each individual word that appeared in the document was already extracted to generate “a list of document terms”. Besides, the words which have the highest TF-IDF weight, which means they are the important keywords that can highly represent the document, have already been identifies. Then, based on the list of document terms, the trained NB classifier calculates the probability of each word being annotated to a particular category using the following formula stated in Eq. (3.20):

$$\underset{n}{\operatorname{argmax}}\{P(C_n\,|d)\} \propto\; P(d|C_n) \times P(C_n) \tag{3.20}$$

A document is identified and classified to the right category according to the probability of occurrence of certain words in the document that match with the terms appeared in the list of word occurrence constructed for each category. As a result, the NB classifier may come up with a result that a particular document can fall into several categories. However, the Bayesian classification approach arrives at the correct classification as long as the correct category gives the highest probability value as compared to other categories.

Interaction Rules Building

After the generation of classified result, the documents related to drug interaction are identified. Within these documents, two different kind of information will be captured: drug-drug interactions and factors affecting the medication. A text analysis engine will use the drug name (like ‘Nembutal’ and ‘Lomotil’) to examine whether an interaction will occur or not. For example, in an online article (http://ohioline.osu.edu/ss-fact/0129.html), the engine finds that there is an interaction between these two drugs as stated:

“Mixing ***antidiarrheal medication*** *(e.g., Lomotil) and* ***tranquilizers*** *(e.g., Transxene, Valium),* ***sedatives*** *(e.g., Dalmane, Quaalude), or* ***sleeping pills*** *(e.g., Amytal, Nembutal, Seconal) can*

result in an increased effect of tranquilizers, sedatives, or sleeping pills." (Christine, 2001)

Therefore, the information extraction engine will transform these valuable information into some rules, like 'IF $Drug_1$= Nembutal and $Drug_2$=Lomotil THEN Interaction=YES'. Furthermore, we will adopt the scenario-based explanation as proposed by Druzdzel and Henrion (1990) to extract the information of factors affecting the medication. For example, a statement 'Older adults permit fat-soluble drugs (e.g. Pentothal) to move readily to the brain, often resulting in dizziness and confusion' is captured in the engine, a scenario can be set as 'AGE>60, FAT_SOLUBLE Yes, CONFUSION Yes, DIZZINESS Yes'. When several scenarios related to the same relationship are obtained, they can be accumulated together to form a new rule like 'IF AGE>60 AND Drug_Char=FAT_SOLUBLE THEN CONFUSION=Yes and DIZZINESS = Yes'. Upon the interaction is detected, all the information will store in the Information Services Module for further decision support.

Instead of warning the physicians about the drug-drug interaction, this module extends the function on providing alternative solution. By matching with drug description in the Drug Information database, it can filter all the appropriate replacement in which they match the diagnosis confirmed with no interaction being found. All the reporting information is displayed and presented to physicians by interfacing the Medical Diagnosis Module.

3.3.6 Information Services Module

The Information Services Module contains the system databases that maintain all the information of the MedicPDSS. It serves as the knowledge repository of the system. Physicians can retrieve patient data from this module and turn them into useful information. Five databases storing different type of information serve different purposes. Each record in the Patient Information database is associated with a registered patient with his/her unique identification number and personal information, for example patient name and contact information, whereas the Medical Records database stores information on the medical cases (such as the symptoms and diagnosis made, and the treatment and/or medicines prescribed) associated with individual patients. These two databases will provide input to the Automatic Knowledge Elicitation Module. The Drug Information database stores all the validated interaction rules to facilitate detection of drug contraindication. In order to help the physician to gain access to a wide variety of data in support of an investigation (i.e. in the hybrid reasoning and decision support process), the relevant data extracted from the Medical Records database can be transformed into the Cases database through instructions written in the Structured Query Language (SQL), which is a standard database language designed for managing data in database systems.

3.4 Summary

To realize the automatic knowledge elicitation for template building in the electronic medical records system and better modeling the prescription decision with latest drug-drug interaction rules discovery, a MedicPDSS is designed and developed based on module-based system architecture, computer science, as well as software and knowledge engineering techniques.

MedicPDSS consists of five modules and one sub-system. The sub-system, named TEMRS, aims to facilitate the knowledge acquisition process by representing the required medical knowledge via a machine-readable format. Compared with the traditional EMRS, TEMRS integrates a template concept (i.e. a tailored symptom list that is associated with a particular diagnosis) in designing the system interface so as to provide a better data input interface for physician to access the patient's information and input the patient's complaint. In order to build the template of TEMRS, Automatic Knowledge Elicitation Module is applied to capture the medical knowledge embedded in the TEMRS and hence transform it into an XML format and concept map diagram for better visualization. With the construction of map, Medical Diagnosis Module is employed to covert each map into specific diagnostic template for physicians to communicate with the MedicPDSS via the interactive interface.

As one of the objectives of this study is to enrich the decision support in the prescription process, therefore, the Prescription Modeling Module can automatically model the prescription decision stored in the Information Services Module and hence derive the appropriate prescription solutions for addressing the new complaint inputted by the physicians. An algorithm named RACER was designed to extract the retrieved prescription solution into a number of medicines by assigning weights to determine their appropriateness and hence consolidate the solution into three different rankings via the adaption of Dempster’s rule of combination.

To ensure the medicines prescribed are correct and safe to the patient, the Risk Surveillance Module (employed the text and web mining techniques to identify the interaction rules from the literature) can detect the drug-drug interaction within the selected medicines. With most of latest drug-drug interactions are available in the literature, this module first searches the relevant article and hence extracts the interaction information from those documents that contain the drug interaction rules.

In this chapter, the research methodology for the medical prescription support and the system architecture of MedicPDSS are presented. The theoretical base and working principles for a series of computational intelligent modules and algorithms are presented for supporting the MedicPDSS. These modules and algorithms are original and contribute significantly to advancement of technology for tacit knowledge acquisition, prescription modeling, and the detection of drug-drug interaction information. They attempts to address the limitations and deficiencies of existing algorithms as reviewed in Chapter 2. The performance of the key modules and algorithms of the MedicPDSS is evaluated in Chapter 4. The application of the system is demonstrated in Chapter 5 by a trial implementation in a selected reference site in the healthcare industry. A case study is used to measure the performance of the system.

Chapter 4 Implementation and Case Study

In order to demonstrate the medical prescription decision support methodology above, a Medical Prescription Decision Support System (MedicPDSS) was designed and then applied in a Hong Kong medical centre named Humphrey & Partners Medical Services Limited (HPMS). It was found during the study that by using the prescription decision support system, the medical prescription process was more effective and more accurate than the method used previously (see Chapter 5). The case study is described below.

4.1 Case Study Background

HPMS is one of the largest multi-disciplinary medical services providers in Hong Kong. It was founded by a team of dedicated medical practitioners, and consists of 4 core clinics located in different parts of the city and about twenty medical experts working on a rotational basis to provide various, high quality medical services to its patients. The general practice in a treatment consists of several steps, including patient registration, GP diagnosing, medical prescription and delivery of drugs. At HPMS, GPs find the current medical information system is not user friendly as they find it difficult to identify and choose the drugs (from two hundred drugs available in the clinic) required for the treatment; which makes the prescription process more complicated. Thus, MedicPDSS can support GPs to easily and quickly retrieve the patient information for the whole treatment process. The hybrid model can thus help the

GPs to look up and select the required drugs efficiently by ranking the drugs based on diagnosis and on the doctor's individual method of prescription.

4.2 Timeframe for the Study

With the approval from the Board of Directors (BoDs) of HPMS, interview and administrated questionnaires were conducted for examining the current practices, problems encountered as well as collecting the user feedback on designing the system. The entire project (design and development stage) lasted for 4 months (i.e. from 1 June to 30 September 2010). There were a total of six nurses, six administrative staff and eight doctors participated in the study.

Upon the pilot testing and fine-tuning of the system, we evaluated the effectiveness of the proposed approach in the period of 1 October 2010 to 28 February 2011 and collected the user feedback via questionnaires. The evaluation questionnaire was divided into 2 parts, in which part 1 was related to the operation of the system whereas another part was related to the functional features of the system. In general, the focus on the questionnaire is to understand the user's attitude towards the area in decision support function, reporting function, data input, information retrieval and system maintenance in the form of 5 satisfactory levels (i.e. 5-point Likert scale), namely Very Dissatisfied, Dissatisfied, Normal, Satisfied, Very Satisfied. Details of the survey and results are explained in Chapter 5.

4.3 Structural Framework for System Design and Development

The structural development framework of the MedicPDSS consists of nine stages, namely background analysis, project team formation, scope and goals identification, system design, hardware and software requirement, pilot testing, implementation, staff training, and system maintenance and monitoring. All these stages can provide a comprehensive thinking on necessary considerations in the system design and development. Furthermore, we divide this nine-stage structural framework into four different phases for simplicity. Figure 4.1 depicts the methodological approach for developing the MedicPDSS.

4.3.1 Phase 1: Preparation

The main purpose of preparation phase is to get the general picture of the whole development. Current workflow analysis (i.e. workflow across different departments) is first introduced for helping in understanding more about how the existing business process works. Then, problem identification is used to measure and analyze existing weakness of the circumstances, and to investigate room of improvement and possibility of applying ICT under the given situation. Scope and goals are also identified and determined for creating boundary of the project and target setting. The last stage in this phase is to form a project team to manage, control and facilitate the system development.

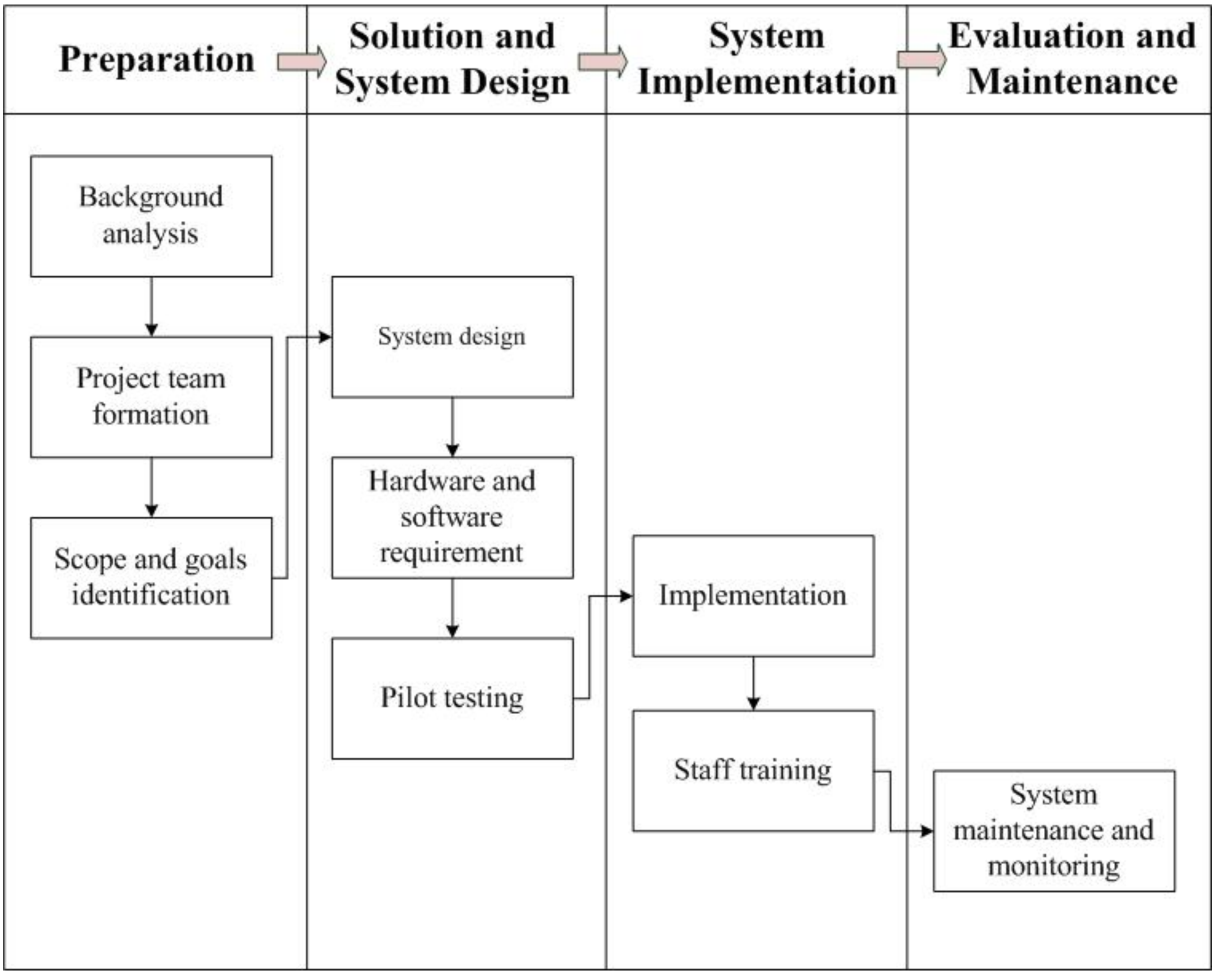


Figure 4.1 – Structural Framework for MedicPDSS Design and Development

4.3.1.1 Background Analysis

In this stage, the situation in HPMS is analyzed and an understanding of the situation emerges. It is noted that the purpose of ICT is to streamline the operation and optimize the workflow of the services, thus several steps in the existing workflow will be simplified and altered. Therefore, we first administrated questionnaires to all the staff in HPMS (including doctors, nurses and administrative officers) for understanding their attitudes towards to current situation and new technology adoption. Upon the completion of questionnaire collection, we then extracted and listed out all the critical and important opinions and comments from the survey and hence conducted interviews with 5 nurses and 5 doctors (who are randomly selected) for better recognizing the consequence for these concerned areas. For example, Table 4.1 lists out the top five challenges in the current situation of HPMS.

Table 4.1 – Top Five Challenges in HPMS

Challenge	**Response rate**
1. Poor handwriting	85%
2. Insufficient human workforce	70%
3. Loss of patient medical records	55%
4. Inconsistent medical terminology	48%
5. Wrong prescription	45%

4.3.1.2 Project Team Formation

In order to facilitate the project with support, a project team should be formed with executives on the managerial level and representatives from key departments. For example, in HPMS, head nurses, doctors in different medical professions and administrative officers are selected for committing the success of the system development. Their role is to monitor the project progress and make certain both milestones and schedules are being met. Representatives who selected from the departments require to familiar all the details of current operations in their own field. Furthermore, all the representatives have their authority to make key decisions related to the project and also have the target to attain the project goal.

4.3.1.3 Scope and Goals Identification

The project scope can serve as the boundary of the project that clearly defines which parts should be included in the project. It aims at preventing the project from growing out of control since it is common to find many projects grow gradually to encompass more and more business areas. Project goal is defined to specify the desired deliverables with considering the company goals for the future. It can guide the project team focusing on the tasks which are required to achieve in the project and should be acquired the top management support in term of both staffing and funding to ensure the project can be carried out smoothly.

4.3.2 Phase 2: Solution and System Design

This phase is aimed at recognizing the particular areas that need improvement of the current paper-based medical record system and prescription decision support. Documentations are written in detail to describe how the MedicPDSS works. Two stages are involved in this phase: one is to design the new solution for the MedicPDSS, whereas another is to specify the system hardware and software required for implementation of the new processes.

4.3.2.1 System Design

The overall system design is established before deployment, taking into account the functions and services that the system should provide to improve the current situation. Concerning user resistance is one of the critical obstacles in the system adoption, especially in the new technology introduction (Mohd and Mohamad, 2005), the results from the background analysis stage should also be considered in the system design, and thereby a more tailor-made system can be made. In this study, we discover that most of the doctors and nurses are afraid of using computer as they are not familiar with; so in this case, the system interface is proposed to design in a user-friendly, graphical and ease-to-use manner. Furthermore, the interface is designed as similar as the company's current paper-based medical record so as to make the staff more familiar in manipulating the system. In addition, we find that several doctors have tried to adopt EMR in their practices; but due to the complexity and difficulty of the system, they finally refuse to use. To cope with this issue, we propose to introduce a template-based EMR for them as it can base on the doctor's preferences to design what should be included in the interface. In general, Figure 4.2 shows the current paper-based records against the newly designed TEMRS.

Apart from the interface design, system functions are also one of important criteria should be taken into account. In this study, the MedicPDSS consists of seven subsystems: patient management system (PMS), visit scheduling system (VSS), diagnostic and treatment system (DTS), drug interaction detection system (DIDS), prescription dispensing system (PDS), billing system (BS), and reporting system (RS). All the interactions between the system and the users are shown in Figure 4.3 and corresponding system functions are explained in Table 4.2.

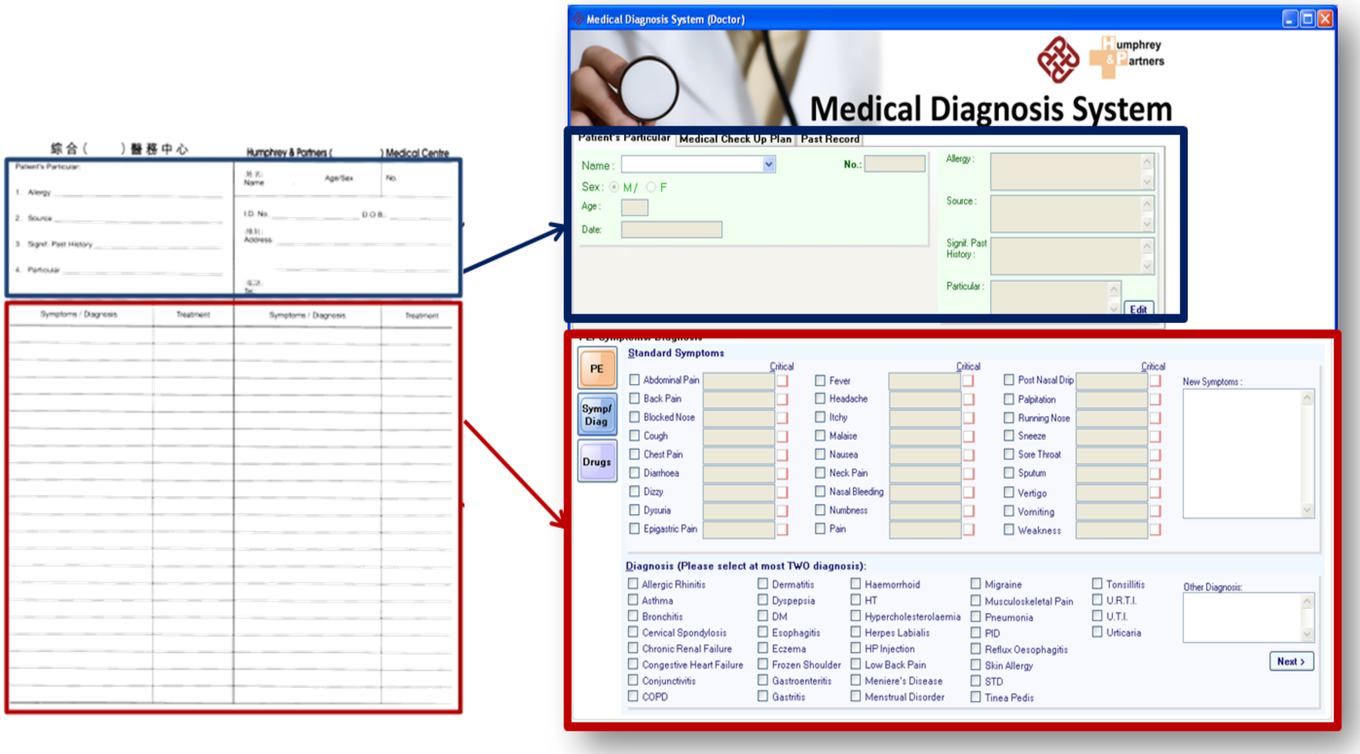


Figure 4.2 – Current Paper-based Medical Record and the Interface of the TEMRS

Once the system components have been designed and developed, it is essential to undertake the system integration. It aims to connect the subsystems together and ensure them functionally operate before the actual deployment. A system block diagram is used for indicating the data flow should be established first that can give a general overview of the connection of the system components and ensure a correct data flow to be generated after the integration.

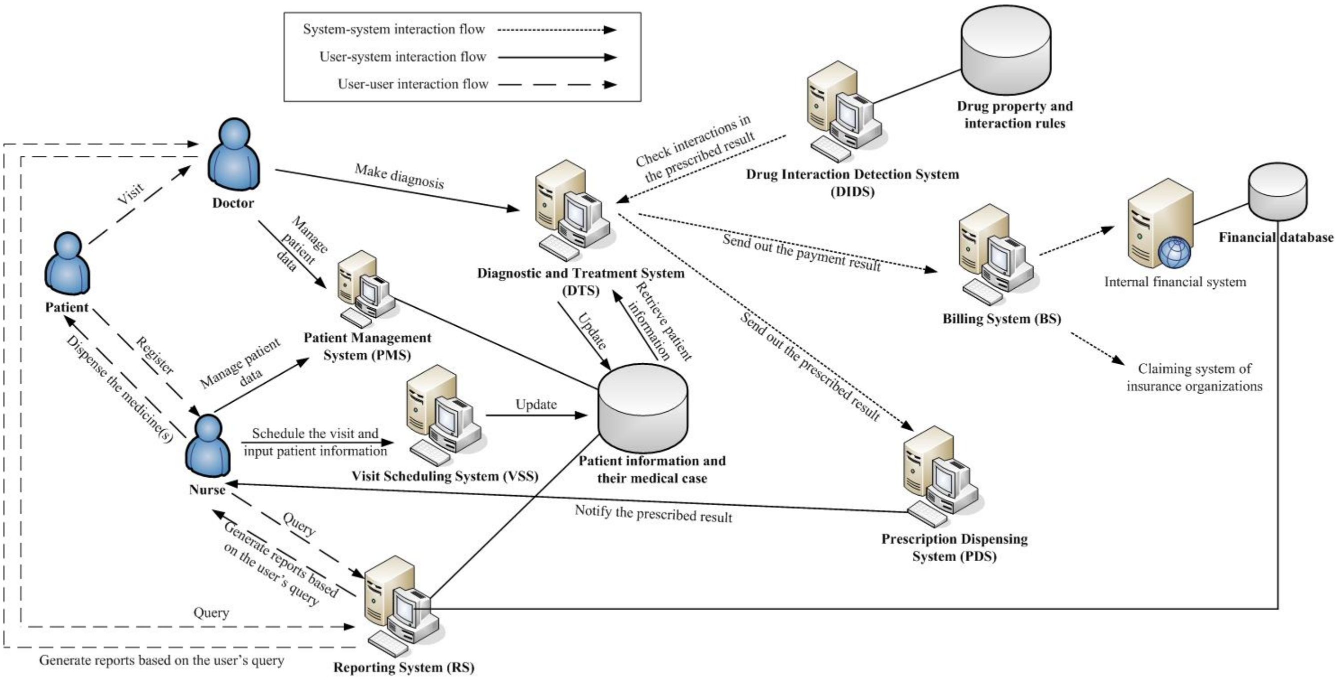


Figure 4.3 – Interaction between MedicPDSS and System Users

4.3.2.2 Hardware and Software Requirement

MedicPDSS must be supported by appropriate hardware that HPMS must find the suitable hardware for the implementation. Testing can ensure the hardware is reliable enough to operate as expected to support the desired implementation. The operating system, database and any communication links to legacy systems (such as accounting system) may also be considered. In addition, the hardware and operating system should be installed well in advance of the beginning of the implementation stage. In general, the system was composed of the following components: (i) web and application server, (ii) database server, (iii) wireless and Ethernet network, (iv) .NET framework 2.0, and (v) relational database management system (such as SQL Server 2005).

4.3.2.3 Pilot Testing

This step is to exercise the demo system and test the users' understanding of the system. The main aim of this demo testing is to identify the interrelationship between each functional area's actions and the problems encountered by HPMS. In this study, we carried out several trials in HPMS during its non-office hours to see if any errors would occur in the real situation. In general, testing can be categorized into three main types, hardware to hardware; software to software and software to hardware. The first one is to verify all required hardware and peripheral equipment have been installed properly in term of quantity and location. The second one is to ensure all software and system programs are successfully linked. The last one is to verify both software and hardware are fully integrated. Moreover, we marked all the errors on a chart and debugged the system afterwards. Although this was a demo testing phase, all users participated and provided feedback to us for tailor-making the system.

Table 4.2 – Core Subsystems in MedicPDSS

Subsystem	**Function(s)**
Patient Management System (PMS)	To facilitate the patient registration and stock up the information of patient for further use in other subsystems
Visit Scheduling System (VSS)	To support appointment making of the patient and remaindering the upcoming patient appointment automatically
Diagnostic and Treatment System (DTS)	To retrieve the patient's past medical history and facilitate the entire diagnostic and prescription process via the use of the template
Drug Interaction Detection System (DIDS)	To detect any drug-drug interaction and patient-drug interaction for the medicine(s) to be prescribed in DTS, and alert both the doctors and nurses during the drug prescription process if any errors are detected
Prescription Dispensing System (PDS)	To transfer the prescribed result (from DTS) to nurse's display screen in real-time manner and facilitate the picking appropriate medicine(s) process
Billing System (BS)	To convert the payment of each visit to insurance company and company's internal financial system automatically
Reporting System (RS)	To generate all kind of report base on the query of users (e.g. monthly total visit report and monthly visit per doctor)

4.3.3 Phase 3: System Implementation

System implementation phase is to deploy all the system components with following to the proposed project plans developed in the second phase. Apart from these, training is also an important activity that should be done in this phase before the new system processes start.

4.3.3.1 Implementation

Unexpected problems and issues will probably occur during the actual implementation. Therefore, to prevent this from happening, it is recommended that before the actual uses of the system, a part of the business that can be segregated and brought online should be identified. This segment of the business, which should be something that is expected to continue after the completion of implementation, should have no major production or technical issues, and will be a credible test of the overall system. In this way, it allows for the new policies and procedures to be tested in actual use with a minimum data set that people are familiar with before the entire company is committed.

4.3.3.2 Staff Training

Training should be provided to those related staff after the test of the installation have been complete. Different training is going to assign to different staff representatives. Because different person response for different parts in the system so different training is required. For operators, trainings which relate to the operating procedures and use of front-end programs or interfaces should be given because it directly determines whether the system implementation success or not. Training can be classified to two types, practical and documentations. In this case, operator should be trained in person to teach them how to use the TEMRS and MedicPDSS from the point of patient registration to the end of medicine prescribing as well as the payment. Documentations can be offered to both operators and system maintenance representatives to familiar their roles. Examples of documentation for operators are user and reference guides of the TEMRS and MedicPDSS.

4.3.4 Phase 4: Evaluation and Maintenance

This phase is to review the new system process performance and conduct baseline measurement for the new system, thereby allowing the project team to modify and fine tune the new system according to the evaluation results. Moreover, suggested solution for the improvement in the system performance will be proposed.

4.3.4.1 System maintenance and monitoring

System maintenance and monitoring is essential in any ICT areas. After the whole company has completed employed the system, this is an excellent time to review the performance measures established at the beginning of the project for the system, and evaluate the results.

In the case of HPMS, expected results should include the satisfaction rate of those using the system, increased level of productivity, less searching time, and an improvement in the overall patient safety level. These benefits enable continuous improvement without additional resources.

In order to better communicate the errors that occur in the system, a reporting system is highly recommended in HPMS. This will also provide a way of handling any skepticism. Medical staff is educated to report the system errors through either email or phone; so we can immediately be made aware of any problems and addresses them in the early stages.

4.4 An illustrated Example – from TEMRS to MedicPDSS

The decision support approach has been tested in HPMS to validate the feasibility of this solution in an actual operational environment. Totally, seven phases are involved in the adoption process of MedicPDSS (Figure 4.4).

4.4.1 Phase 1: Diagnosis by Medical Expert

The system interface for the GPs to make treatment is shown in Figure 4.5. After registering in TEMRS, the patient information, including patient name, sex, age, allergies, past medical history, are transferred to the GP's computer. In order to obtain a better result in the decision support approach, the symptoms and diagnosis are pre-defined in the system based on the template formation for each diagnosis, in which GPs just simply select and check the box under the symptoms/diagnosis column. On the other hand, for those symptoms and diagnosis that have not been encountered before, an input area is designed for GPs to type in specific information.

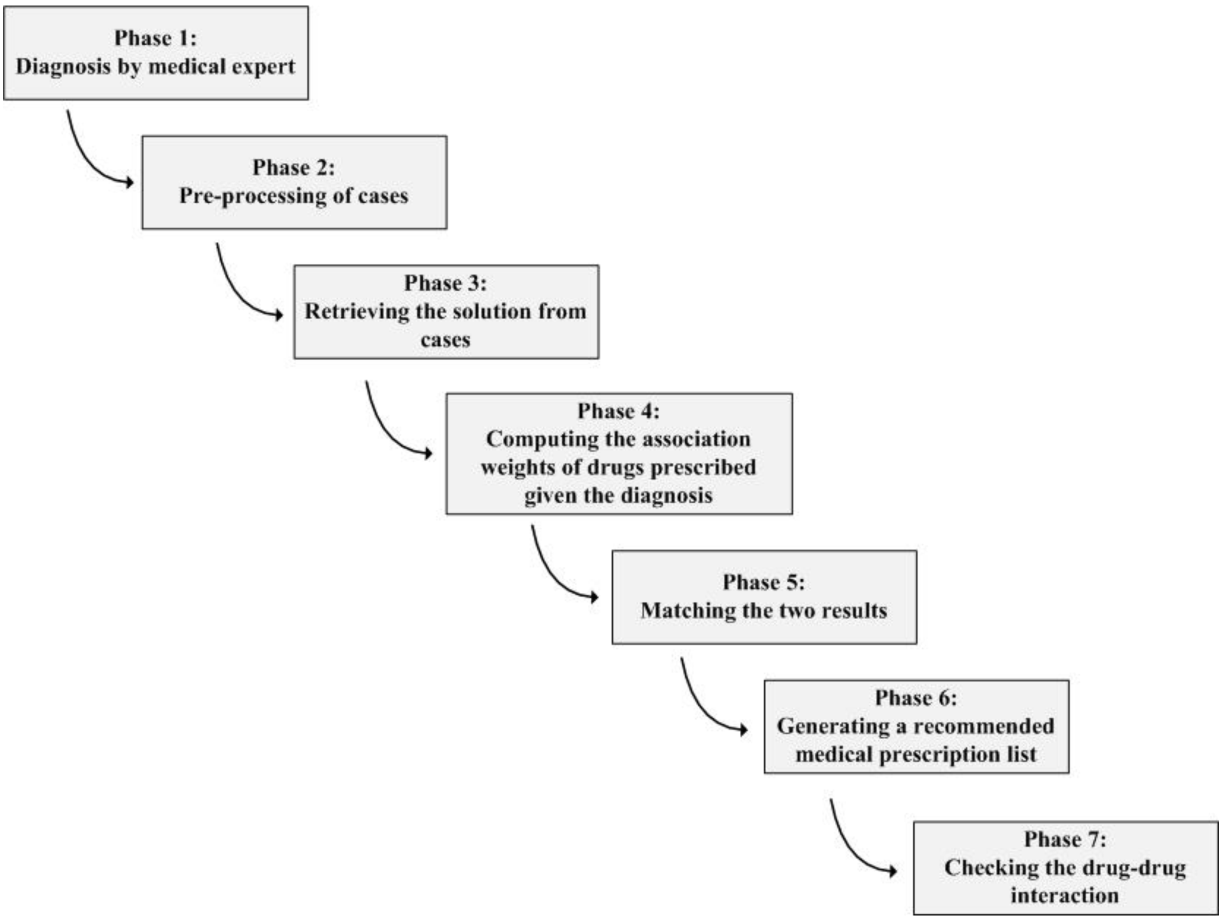


Figure 4.4 – Phases in an Illustrated Example

4.4.2 Phase 2: Pre-processing of Cases

This phase focuses on turning the data warehouse into a data mart for easy access to frequently needed data. Before retrieving the cases to find similar solutions, a pre-processing method is used to index and extract the specific information from the data warehouse. Some irrelevant data is removed in the knowledge base. For example, “referral” does not have any effect on the decisions made in drug prescription and is thus removed.

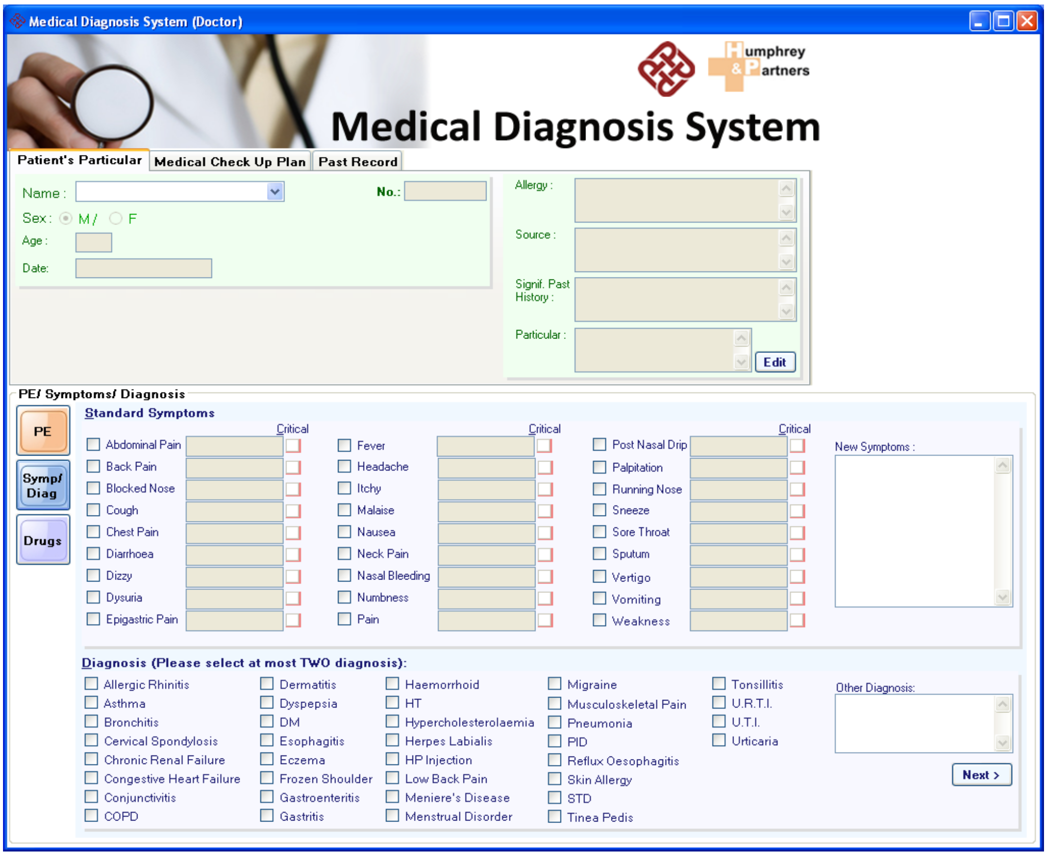


Figure 4.5 – Diagnosis of Medical Expert

4.4.3 Phase 3: Retrieving the Solution from Cases

After the GP decides the diagnosis and the pre-processing phase, all the relevant information is gathered to perform the CBR process. Table 4.3 summarizes the attributes for case featuring.

It involves the patient information and past treatment details (such as last record, number and duration of sick leaves, payment, diagnosis, symptoms, additional services). Before storing in the case base, all these cases will be validated by the BoDs in HPMS who are specialists in various medical disciplines. With their experience, all the stored cases are validated and the collection of these cases covers a wide range of illnesses treated by a large group of physicians. The main purpose of CBR is to retrieve similar cases of patients suffering from the same condition. If the diagnosis and patient information match perfectly with the existing case, the solution of the existing case will be used as the reference to the physician without any change. However, if no exact match is found, Eq. (3.6) is applied to retrieve and propose the most appropriate medical prescription list. All the weights of the features are given by the BoDs in HPMS. On the basis of the data captured from TEMRS, the BoDs discuss the weightings one by one and finally reach a solution. This helps in ranking all the cases in the knowledge base. A typical case in the knowledge base is shown in Figure 4.6. It contains the problems (description of the treatments with patient information) and the medical prescription choice with the association weighting for further matching (see Section 4.4.4).

4.4.4 Phase 4: Computing the Association Weights of Drugs being Prescribed given the Diagnosis

By using Eq. (3.11), Eq. (3.12) and Eq. (3.13) to generate the interesting association rules, we can rank all the drugs prescribed in descending order of probability based on the input from phase 1. The probability is computed based on the frequency of drug selection captured from the past instances of prescription.

Table 4.3 – Summary of the Case Attributes

Attribute	**Possible values**
Patient number	Unique ID (e.g. 34458, 32251, 1121)
Age	Positive Integer (1-100)
Sex	M, F
Body Weight(kg)	Positive Integer (1-100)
Height(cm)	Positive Integer (1-250)
Last Record	Positive Integer (today – last treatment date)
Number of days of sick leave	Positive Integer (0-30)
Payment	Positive Integer (20-1000)
Diagnosis	Multi-value ($\{URTI, Gastroenteritis, ..., Rhinitis\}$)
Symptoms	Multi-value ($\{Fever, Cough, ..., RunningNose\}$)
Days of medication	Positive Integer (0-5)

4.4.5 Phase 5: Matching the Two Results

It is realized that the experience of GPs is directly proportional to the number of cases they have dealt with. Therefore, this phase aims at combining the results from the two different models by weighting with their experience in order to reduce the bias of the drug choice.

Similar to phase 2, the weight is provided by the BoDs with reference to the number of visits to the GP, past history and patient revisit rate. The weight is adjustable from low to high (on a scale from 0 to 100%). This is useful when there is a change in performance of a particular GP.

Patient Information

Patient ID:	26734
Patient Name:	Jacky Ting Siu Long
Age:	23
Sex:	M
Body Weight:	75kg
Height:	171cm
Last visit:	1/2/2008
Diagnosis:	URTI
Symptoms:	Fever, Cough
Payment:	HKD 120
Day of Medication:	5
Number of Sick Leave(s):	0

Similar Case Retrieved

Attribute	Weight (0-1) (w)	New Case (a)	Old Case (b)	Similarity (c) = $(a-b)^2$	Weighted Similarity (d) = $(w^2 * c)$
Patient number	0.8	26734	26734	0	0
Age	0.7	23	23	0	0
Sex	0.6	M	M	0	0
Body Weight(kg)	0.5	75	72	9	2.25
Height(cm)	0.5	171	171	0	0
Last Record	0.7	20	60	1600	784
Number of sick leave	0.2	0	2	4	0.16
Payment	0.3	120	150	900	81
Diagnosis	0.9	URTI	URTI	0	0
Day of medication	0.3	5	3	4	0.36
Similarity	-	-	-	-	**29.46**

Output

Drug Prescribed for Last Case

Cocillana
Bisolvon
Lysozyme
Ponston
Actifed

Figure 4.6 – An Example Case and the Proposed Solution

4.4.6 Phase 6: Generating a Recommended Medical Prescription List

After combining the results from phase 4, the GP can have the recommended medical prescription list regarding to the patient's problems. Thus, the most commonly prescribed drugs from two different models will be placed on the top, whereas the remaining drugs will be ranked in descending order of the probability of their being prescribed. Figure 4.7 shows the final result of the recommended medical prescription list.

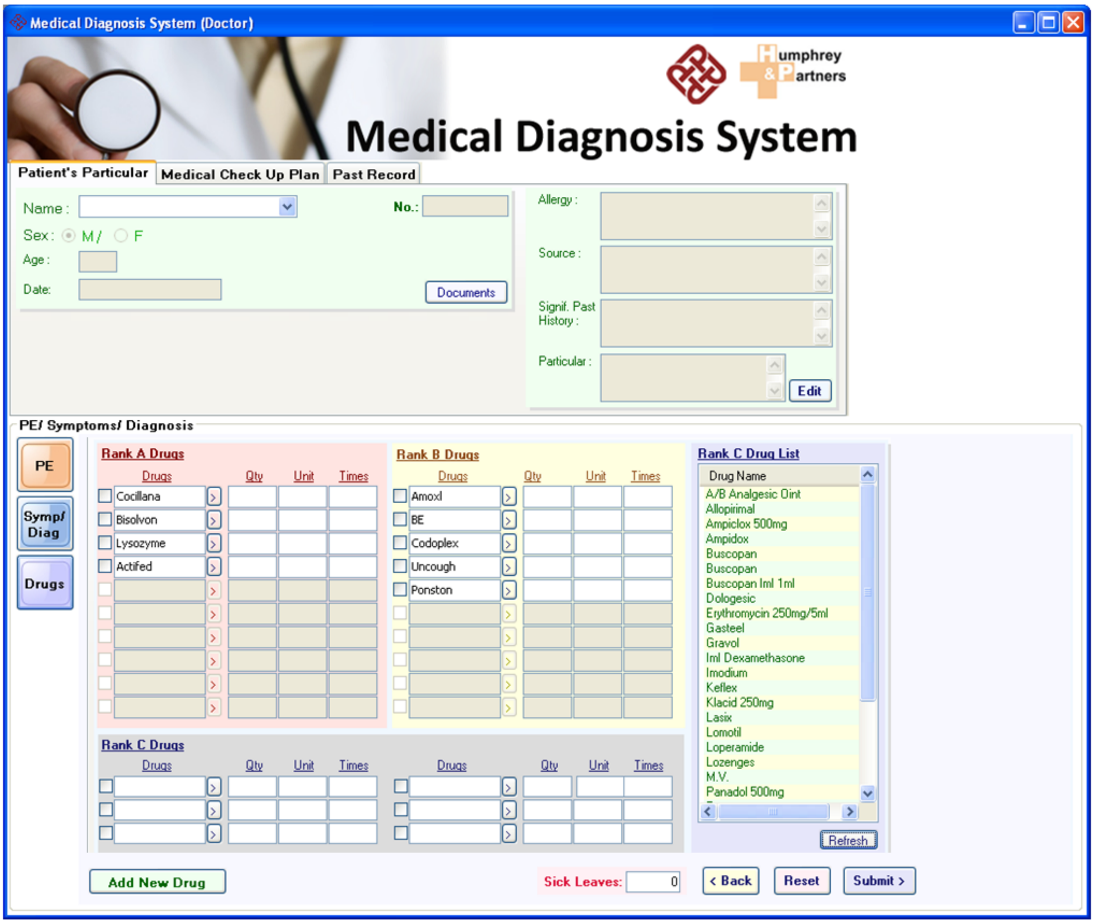


Figure 4.7 – The Recommended Medical Prescription List Produced in MedicPDSS

4.4.7 Phase 7: Checking the Drug-drug Interaction

Upon the medicines selection of physicians, MedicPDSS automatically check the drug-drug interaction of the selected drugs from the rules stored in the Interaction Rules Database. As shown in Figure 4.8, this function determines whether the medicines dispensed to the patient have interaction or not. If an interaction exists, a large pop up alert message box is generated to warn the medical workers.

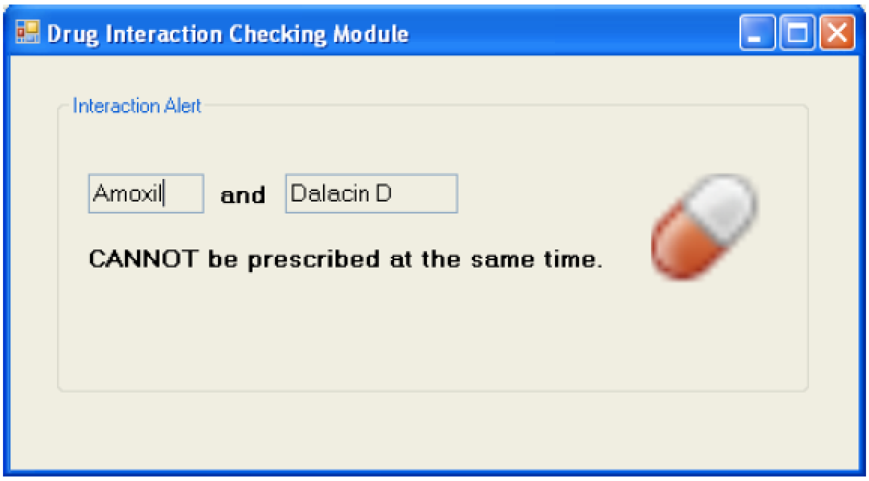


Figure 4.8 – Drug-Drug Interaction Checking Function in MedicPDSS

4.5 Summary

A case study was carried out at a medical organization in Hong Kong during the implementation of MedicPDSS. A structural framework, which consists of four phases (i.e. preparation, solution and system design, system implementation, as well as evaluation and maintenance) was introduced to support the design and development of MedicPDSS. The entire project lasted for nine months, in which the first four months were taken to complete the first three phases (from preparation to system implementation) whereas the remaining five months were used for the system evaluation and maintenance. In total, twenty medical experts (i.e. six nurses, six administrative staff and eight doctors) were participated in this study.

An illustrated example was also provided for validating the feasibility of MedicPDSS in the actual operational environment. In order to better illustrate the practicability of MedicPDSS, the adoption process of MedicPDSS was discussed step by step in the prescription process, in which it begins with the diagnosis determination of physicians and ends at the detection process of drug-drug interaction.

Chapter 5 Performance Evaluation and Discussion

In this study, MedicPDSS infrastructure has been designed to facilitate decision making in the drug selection process. It has been specifically designed for healthcare professions that encounter the complex and vast amount of drug information. MedicPDSS has been implemented in a local Hong Kong medical organization to validate its feasibility in providing reliable decision support in medical prescription process (see Chapter 4). In this chapter, the results and discussion of system implementation are presented. First, system evaluation of TEMRS as well as the supportive template building algorithm – automatic knowledge elicitation is discussed. Second, performances of the other two computational intelligence algorithms (i.e. rule-associated case-based reasoning, and the drug information extraction) are evaluated quantitatively. Third, overall results of implementing MedicPDSS in Humphrey and Partners Medical Services Limited (HPMS) are studied by qualitative and quantitative measures. Fourth, lessons learnt and ethical issues arisen by the system implementation are discussed. Fifth, limitations of the study are derived and the contributions of MedicPDSS to healthcare professionals are discussed finally.

5.1 Evaluation of TEMRS and Automatic Knowledge Elicitation Algorithm

In order to evaluate the efficiency and performance of the TEMRS and the corresponding supportive algorithm - Automatic Knowledge Elicitation Module, a survey was taken as a qualitative and quantitative evaluation tool to collect feedback from physicians (Cooper and Schindler, 2003). Furthermore, because there is no existing template building via the knowledge extraction from EMRs for doing a comparison with this study, therefore expert opinions and feedbacks are used for the evaluation.

5.1.1 Evaluation Settings

In this case study, evaluation of the system took place over a period of one month, starting on 1 October 2010 and concluding on 31 October 2010. 5-point Likert scale questionnaires were constructed and distributed to users at the end of the evaluation period to acquire the feedback on the system performance, to enable comparison between the results of tradition human-based and automatic knowledge acquisition methods (i.e. without template formation in the EMR). In the questionnaire, there were four main parts: the frequencies of usage of the system, user-friendliness, system performance and system maintenance. Different criteria were established under the four areas and are shown in Table 5.1. After the system was used for one month, questionnaires were distributed to selected respondents by surface mail in the late October 2010. A pre-paid, pre-addressed envelope was included to facilitate the return of completed questionnaires. With eight medical experts participating in this evaluation, eight questionnaires were distributed and eight valid responses were returned, making a 100% response rate.

Table 5.1 – Criteria Determination

<table>
<tr><th colspan="2">Evaluation Aspect</th><th>Criteria</th></tr>
<tr><td colspan="2">Frequency of Use</td><td>Patient Information Retrieval</td></tr>
<tr><td rowspan="3">System User-friendliness</td><td rowspan="3">System Design</td><td>Variety of attributes</td></tr>
<tr><td>Simplicity</td></tr>
<tr><td>Ease to learn and use</td></tr>
<tr><td rowspan="3">System Performance</td><td>Information Retrieval</td><td rowspan="3">• Quality and accuracy
• Sufficiency of content
• Ease to understanding
• Effectiveness</td></tr>
<tr><td>Knowledge Representation</td></tr>
<tr><td>Template Formation</td></tr>
<tr><td colspan="2" rowspan="2">System Maintenance</td><td>Sufficiency of technical support</td></tr>
<tr><td>Efficiency of maintenance services</td></tr>
</table>

5.1.2 Participants and Description of Data Collected

Table 5.2 shows the year of experience and medical specialty of each doctor. Among the evaluation period, 200 cases were collected and corresponding population data was summarized in Table 5.3. The average year of experience of doctors and patient age were 10.12 and 35.53 respectively.

Table 5.2 – Characteristics of the Physicians Participated in this Study

Doctor	Year of Experience	Specialty
A	2	Respiratory
B	3	Ear, Nose and Throat (ENT)
C	10	Pediatrics
D	15	Gynecology
E	30	General surgery
F	10	Gynecology
G	6	Respiratory
H	5	Dietetics

Table 5.3 – Statistical Summary of the Population Data

	Min.	Max.	Mean	St. Dev.
Patient age (full year)	0	80	35.53	16.34
Patient gender	0	1	0.52	-

Since TEMR was aimed to facilitate the clinical operations and enhance the data entry of each physician, we identified the typical diagnoses encountered in HPMS because this result can induce us to determine frequent occurrence of diagnoses in the company and hence we can design specific template for that. Table 5.4 summarizes the top 5 diagnoses encountered in HPMS. As shown in the analysis, Upper Respiratory Tract Infections (U.R.T.I.) was the top disease encountered (i.e. 50.59% out of total) whereas Gastroenteritis rated at second (i.e. 10.19% out of total). This result was accepted as there were always sudden changes in temperature during the study period in Hong Kong. Consequently, we designed the template by analyzing the kind of information to be shown in the TEMRS for better data entry.

Table 5.4 – Top 5 Diagnoses Encountered in the Case Study

Diagnosis	Percentages (%)
U.R.T.I.	50.59% (1534/3032)
Gastroenteritis	10.19% (309/3032)
Dermatitis	5.87% (178/3032)
Dyspepsia	5.28% (160/3032)
Rhinitis	4.82% (146/3032)

5.1.3 Survey Results

The feedback of the eight physicians who participated in this case study was collected through questionnaires based on the criteria on Table 5.1. Around 70% of physicians were satisfied or highly satisfied with the system, as shown in Figure 5.1. It means that a high proportion of physicians agreed that this system is valuable in these aspects compared to those using a conventional approach (i.e. without a diagnostic template) especially in template formation. Other feedback is shown in Table 5.5 and Table 5.6.

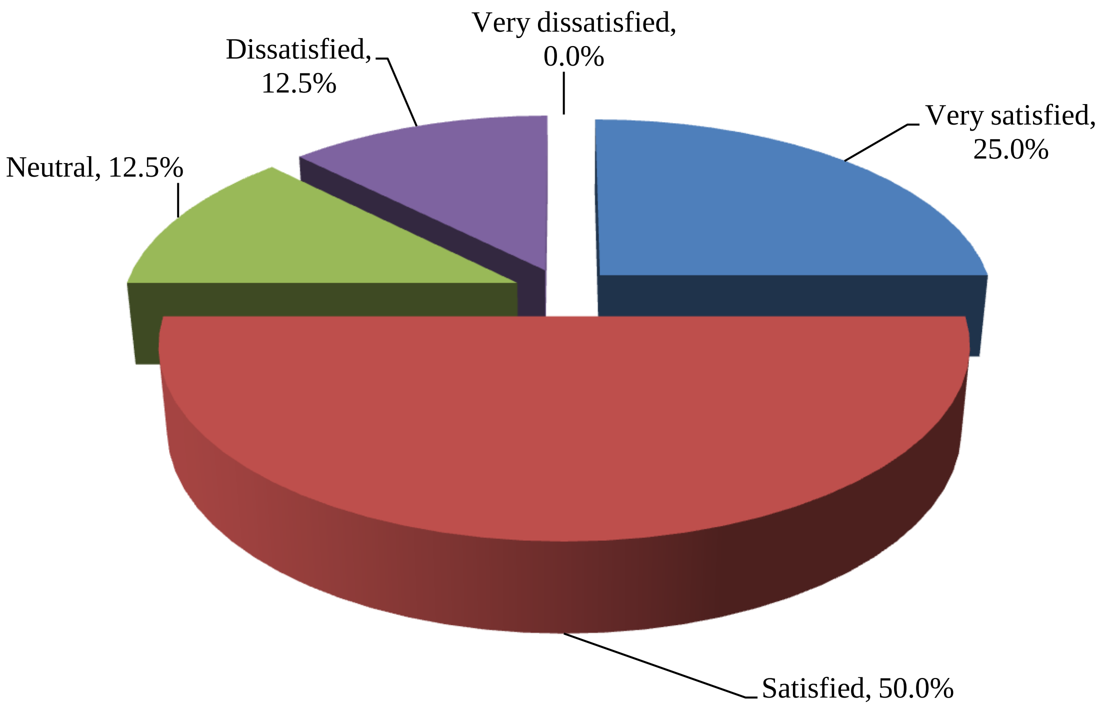


Figure 5.1 – Satisfactory Level of TEMRS

According to the results, the users were satisfied with the performance of TEMRS, especially in the knowledge representation scheme (i.e. visualization through a template) and information retrieval. Concerning the knowledge representation scheme, users appreciated the design of quality and accuracy aspects for the diagnostic concept being used in the corresponding template. In addition, they agreed that TEMRS was good in providing sufficient and ease-to-understand medical knowledge. As a result, the interactions between the physicians were significantly increased compared with the conventional knowledge elicitation mechanism. In addition to the most satisfied area (i.e. knowledge representation scheme and information retrieval), almost all the users agreed the worth of the variety of attributes, simplicity, and ease of learning in which over 95% of users felt that these features were important for the implementation of TEMRS. Furthermore, an interesting finding was related to the system maintenance. Around 90% of users were satisfied with the technical support services and they suggested a user guide be provided with frequently asked question for them to understand more about the manipulation of the system and handling errors.

5.2 Performance Evaluation of RACER and Drug Information Extraction Algorithm

5.2.1 Design of the Experiment Setup for Performance Evaluation of RACER

Figure 5.2 depicts the experiment setup for measuring the performance of RACER. Real case data is collected from HPMS. Since each patient’s medical history, including personal information, information on medical allergy, past visit’s diagnosis and therapeutic result, is recorded in a secure TEMRS, therefore GP-related patient records are retrieved and used in this experiment. In total, 800 cases which ranged from November 2010 to February 2011 are used in this experiment.

Table 5.5 – Result of Performance Evaluation in Frequency of Use, System User-friendliness and System Maintenance (i.e. The Figure Left to the Percentage Represents the Number of Respondents Selecting the Scale whereas the Percentage in Bracket Represents the Average Scores Among the Respondents)

Frequency of Use					
	Frequency				
	Never	Seldom	Sometimes	Often	Frequently
Patient Information Retrieval	0 (0%)	0 (0%)	1 (12.5%)	5 (62.5%)	2 (25.0%)

System User-friendliness					
	Level of Satisfaction				
System Operational Design	Very Dissatisfied	Dissatisfied	Neutral	Satisfied	Very Satisfied
Variety of Attributes	0 (0%)	0 (0%)	0 (0%)	6 (75.0%)	2 (25.0%)
Simplicity	0 (0%)	0 (0%)	2 (25.0%)	4 (50.0%)	2 (25.0%)
Ease to learn	0 (0%)	1 (12.5%)	2 (37.5%)	4 (50.0%)	1 (12.5%)
Overall	0%	4.2%	16.7%	58.3%	20.8%

System Maintenance					
Sufficiency of technical support	0 (0%)	2 (25.0%)	1 (12.5%)	3 (37.5%)	2 (25.0%)
Efficiency of maintenance services	0 (0%)	0 (0%)	2 (37.5%)	4 (50.0%)	2 (25.0%)
Overall	0%	11.8%	23.5%	41.2%	23.5%

Table 5.6 – Result of Performance Evaluation in Information Retrieval, Knowledge Representation (i.e. The Figure Left to the Percentage Represents the Number of Respondents Selecting the Scale whereas the Percentage in Bracket Represents the Average Scores Among the Respondents)

System Performance					
Quality and accuracy	Very Dissatisfied	Dissatisfied	Neutral	Satisfied	Very Satisfied
Information Retrieved	0 (0%)	0 (0%)	1 (12.5%)	5 (62.5%)	2 (25.0%)
Knowledge Representation Scheme	0 (0%)	0 (0%)	1 (12.5%)	6 (75.0%)	1 (12.5%)
Template being designed	0 (0%)	0 (0%)	2 (25.0%)	3 (37.5%)	3 (37.5%)
Overall	0%	0%	16.7%	58.3%	25.0%
Sufficiency of content					
Information Retrieved	0 (0%)	1 (12.5%)	2 (25.0%)	3 (37.5%)	2 (25.0%)
Knowledge Representation	0 (0%)	1 (12.5%)	3 (37.5%)	3 (37.5%)	1 (12.5%)
Template being designed	0 (0%)	0 (0%)	2 (25.0%)	5 (62.5%)	1 (12.5%)
Overall	0%	8.3%	29.2%	45.8%	16.7%

Ease to understanding					
Information Retrieved	0 (0%)	1 (12.5%)	2 (25.0%)	3 (37.5%)	2 (25.0%)
Knowledge Representation	0 (0%)	1 (12.5%)	3 (37.5%)	3 (37.5%)	1 (12.5%)
Template being designed	0 (0%)	0 (0%)	2 (25.0%)	5 (62.5%)	1 (12.5%)
Overall	0%	8.3%	29.2%	45.8%	16.7%
Effectiveness					
Information Retrieved	0 (0%)	0 (0%)	2 (25.0%)	4 (50.0%)	2 (25.0%)
Knowledge Representation	0 (0%)	1 (12.5%)	1 (12.5%)	4 (50.0%)	2 (25.0%)
Template being designed	0 (0%)	1 (12.5%)	1 (12.5%)	5 (62.5%)	1 (12.5%)
Overall	0%	8.3%	16.7%	54.2%	20.8%

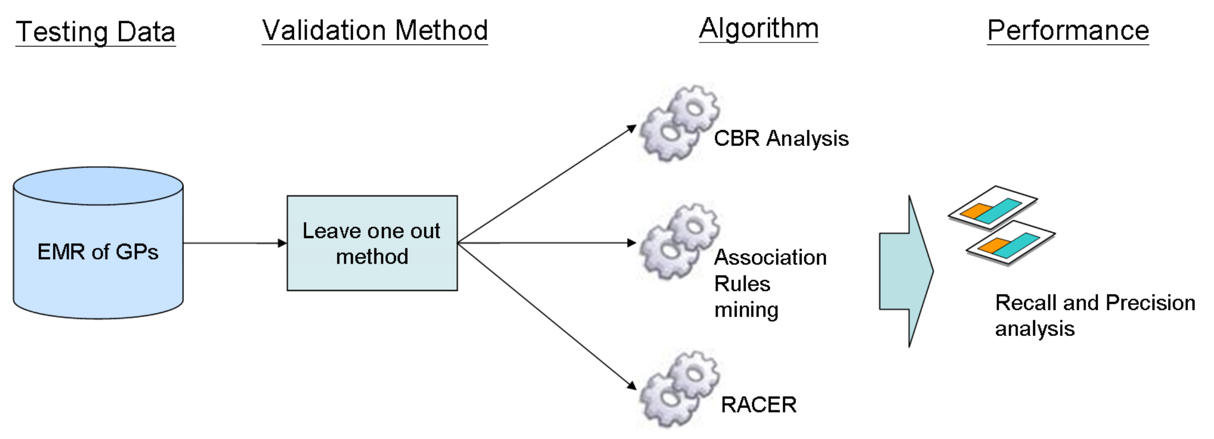


Figure 5.2 – The Experiment Setup for Measuring the Performance of RACER

As shown in Figure 5.2, leave-one-out method is used as the validation method for determining how accurately a learning algorithm will be able to predict data that it was not trained on. Leave-one-out cross validation is useful because it does not waste data. When using the leave-one-out method, the learning algorithm is trained multiple times, using all but one of the training cases. The form of the leave-one-out method is shown as Figure 5.3.

For *i* = 1 to *N* (where *N* is the number of training cases)
    Temporarily remove the ***i***-th case from the training set
    Train the learning algorithm on the remaining *N* - 1 points
    Test the removed case and note the accuracy
End For
Calculate the overall accuracy over all *N* cases

Figure 5.3 – Algorithm of Leave-one-out Method

5.2.1.1 Description of Data Collected

Numerous data are stored in the TEMRS in which not all the data is useful in supporting the prescription making. After discussing with the medical practitioners of HPMS, four categories of data are used in this experiment. The four categories are:

(i) Demographic category: Patient's sex and age

(ii) Allergic category: Patient's allergy on medication

(iii) Diagnostic category: Symptoms and diagnosis in each case

(iv) Therapeutic category: Medicines prescribed in each case

Table 5.7 elucidates the features used in each category and illustrates whether the feature is a problem feature or a solution feature.

Table 5.7 – Features of Diagnosis that are Used in the Experiment

Category	**Feature**	**Possible values**	**Problem feature or solution feature**
Demographic	• Sex	Boolean value (M, F)	Problem
	• Age	Single value (Baby, children, youth, adult, elderly)	Problem
Allergic	• Allergy on medication	Single or Multi-value (e.g. NKDA)	Problem
Diagnostic	• Symptoms	Single or Multi-value (e.g. itchy, nasal discharge, nasal congestion, and so on)	Problem
	• Diagnosis	Single or Multi-value (e.g. URTI)	Problem
Therapeutic	• Medicines prescribed	Single or Multi-value (e.g. Dexophen 30mg, Bisolvon Co, Actifed Co, and so on)	Solution

5.2.1.2 Measure and Procedure

A series of experiments have been carried out for measuring the performance of RACER. The experiment setting is shown is Table 5.8. To verify the scalability of RACER, the experiments are carried out with different number of training cases (i.e. 100 to 800 cases with a 100 cases increment). Three different sets of minimum support and confidence are used in the association rules mining (i.e. 0, 0; 0.1, 0.4; and 0.2, 0.6). Three different sets of threshold values for determining the maximum number of retrieved cases are used (i.e. no. of training cases/10, no. of training cases/5, and no. of training cases/2). Three different sets of threshold values for determining the maximum number of retrieved cases are used (i.e. 5, 6 and 7). Equal feature weightings are used in the CBR and RACER analysis. Only the first most similar case is retrieved in the CBR analysis. Recall and precision analysis are applied for the performance measurement by comparing the suggested solutions of the three analysis method against the actual solution. The recall and precision are defined as Eq. (5.1) and (5.2), respectively.

$$recall = \frac{|d_s \cap d_p|}{|d_p|} \tag{5.1}$$

$$precision = \frac{|d_s \cap d_p|}{|d_s|} \tag{5.2}$$

where

d_s and d_p are the medicine lists of the suggested solution and the actual solution, respectively

$|d_s|$ is the number of medicines in d_s

$|d_s \cap d_p|$ is the number of medicines jointly appearing in d_s and d_p.

5.2.1.3 Result of Performance Evaluation

The results are summarized in Figures 5.4 to 5.8. Figure 5.4 shows the precision and recall of the algorithms with the minimum support = 0, minimum confidence = 0, maximum no. of retrieved cases = no. of training cases/10, and maximum no. of suggested medicines = 5. The figure reveals that RACER outperforms the other two approaches in both recall and precision in this setting. Figures 5.5 and 5.6 show the precision and recall of the algorithms with a higher minimum support and confidence. The recall and precision of CBR and RACER remain steady, whereas association rules mining has a higher precision rate but a very low recall. Figures 5.7 and 5.8 show the effects on precision and recall of RACER by different sets of maximum number of retrieved cases and different sets of maximum number of suggested medicines, respectively. The results show that the precision and recall remain very stable with the maximum number of retrieved cases and it has a higher recall but lower precision when using a higher number of suggested medicines. For association rules mining, two parameters (i.e. the minimum support and confidence) are needed to be adjusted to control the recall and precision. RACER is only required to adjust one single parameter that is the maximum number of suggested medicines. In addition, the meaning of support and confidence is technical and difficult to be understood, while the meaning of maximum number of suggested medicines is much more simple and obvious.

Table 5.8 – Experiment Setting for Measuring the Performance of RACER

Test ID	**No. of training cases**	**Minimum support, Minimum confidence**			**Maximum no. of retrieved cases**			**Maximum no. of suggested medicines**		
1	100	0, 0	0.1, 0.4	0.2, 0.6	10	20	50	5	6	7
2	200	0, 0	0.1, 0.4	0.2, 0.6	20	40	100	5	6	7
3	300	0, 0	0.1, 0.4	0.2, 0.6	30	60	150	5	6	7
4	400	0, 0	0.1, 0.4	0.2, 0.6	40	80	200	5	6	7
5	500	0, 0	0.1, 0.4	0.2, 0.6	50	100	250	5	6	7
6	600	0, 0	0.1, 0.4	0.2, 0.6	60	120	300	5	6	7
7	700	0, 0	0.1, 0.4	0.2, 0.6	70	140	350	5	6	7
8	800	0, 0	0.1, 0.4	0.2, 0.6	80	160	400	5	6	7

All in all, the results exhibits that the performance of RACER remains very stable by using different sets of parameters. The results are almost the same (i.e. only a few percentage differences) when different settings of parameters are used. It is not necessary to know what the appropriate settings for the RACER are in advance, which makes RACER robust.

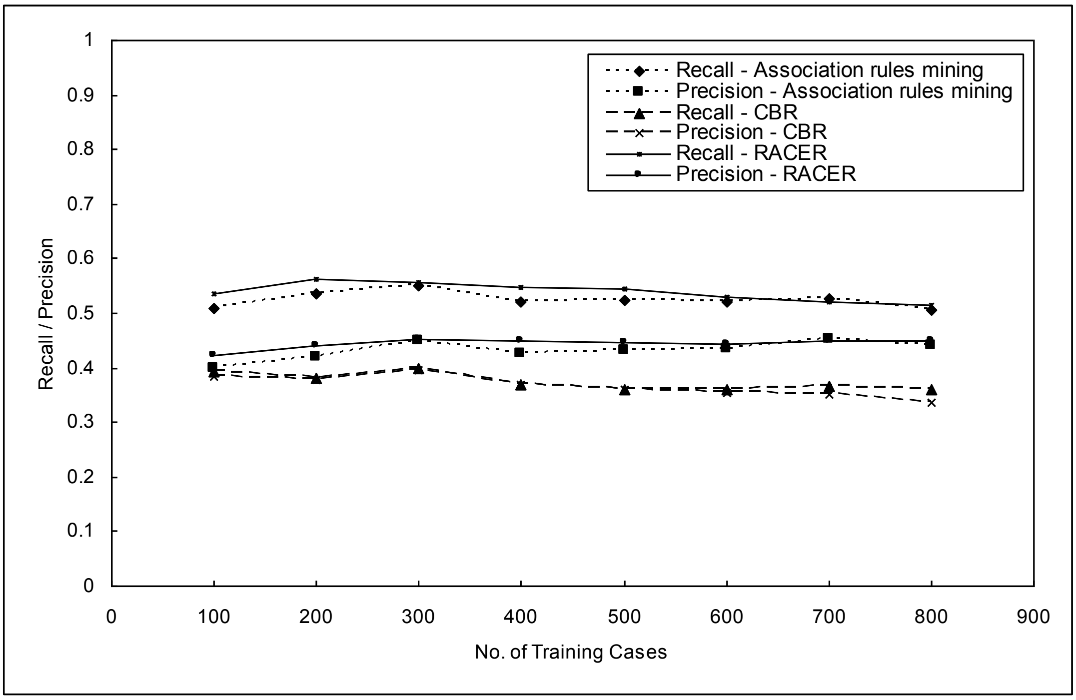


Figure 5.4 – The Precision and Recall of the Algorithms (Minimum Support = 0, Minimum Confidence = 0, Maximum No. of Retrieved Cases = No. of Training Cases/10, Maximum No. of Suggested Medicines = 5)

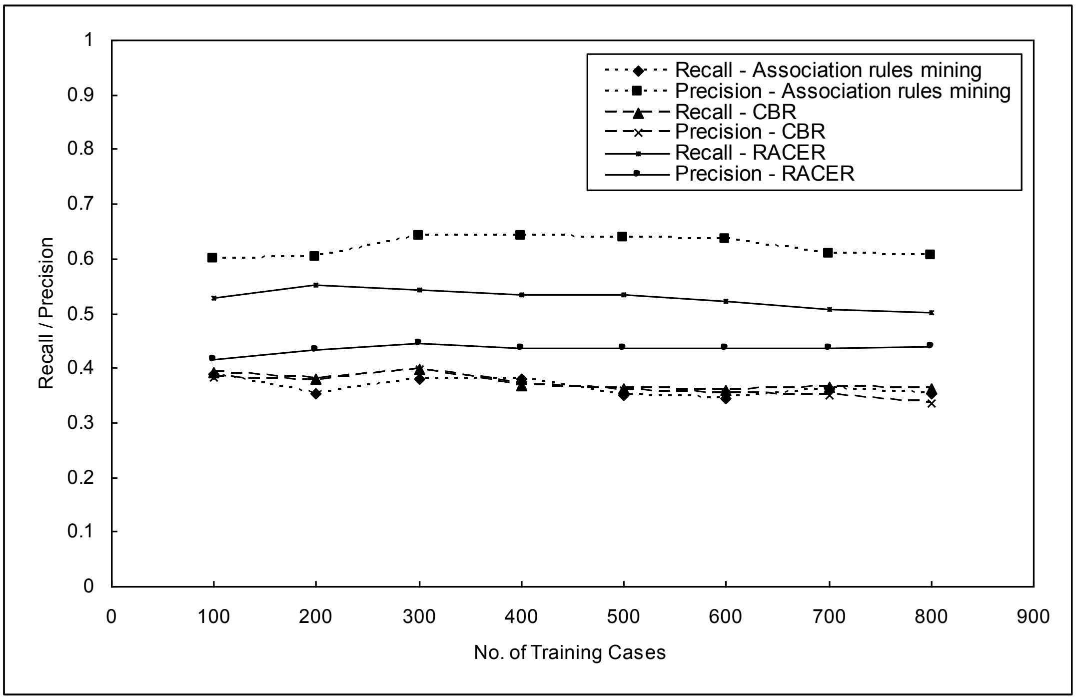


Figure 5.5 – The Precision and Recall of the Algorithms (Minimum Support = 0.1, Minimum Confidence = 0.4, Maximum No. of Retrieved Cases = No. of Training Cases/10, Maximum No. of Suggested Medicines = 5)

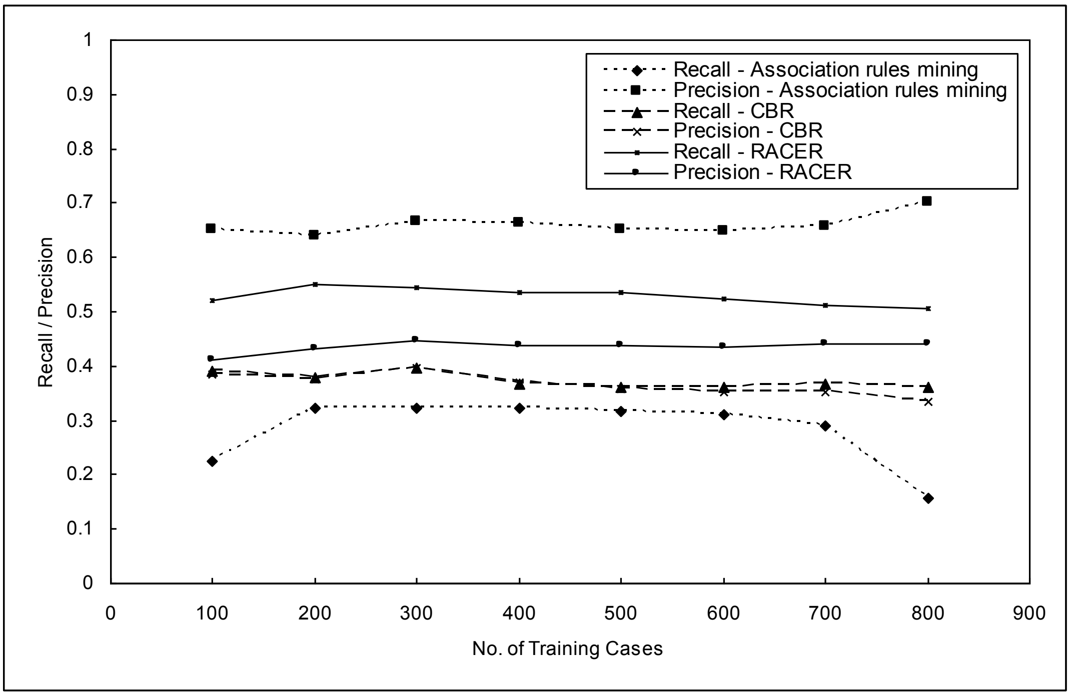


Figure 5.6 – The Precision and Recall of the Algorithms (Minimum Support = 0.2, Minimum Confidence = 0.6, Maximum No. of Retrieved Cases = No. of Training Cases/10, Maximum No. of Suggested Medicines = 5)

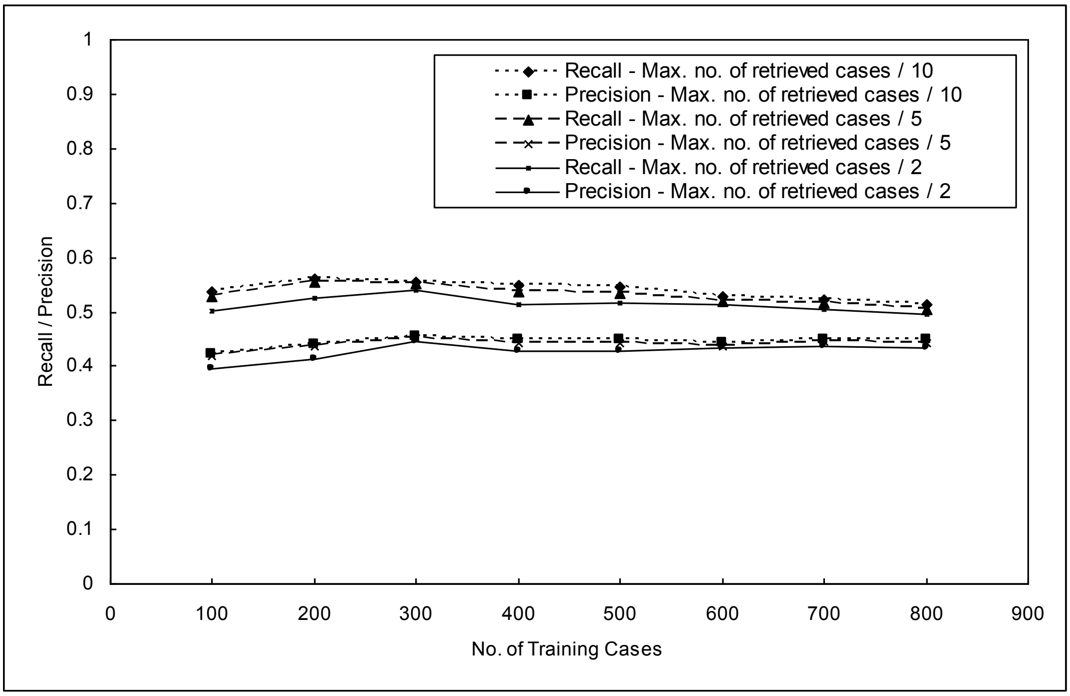


Figure 5.7 – The Precision and Recall of RACER with Different Sets of Maximum of Retrieved Cases (Minimum Support = 0, Minimum Confidence = 0, Maximum No. of Suggested Medicines = 5)

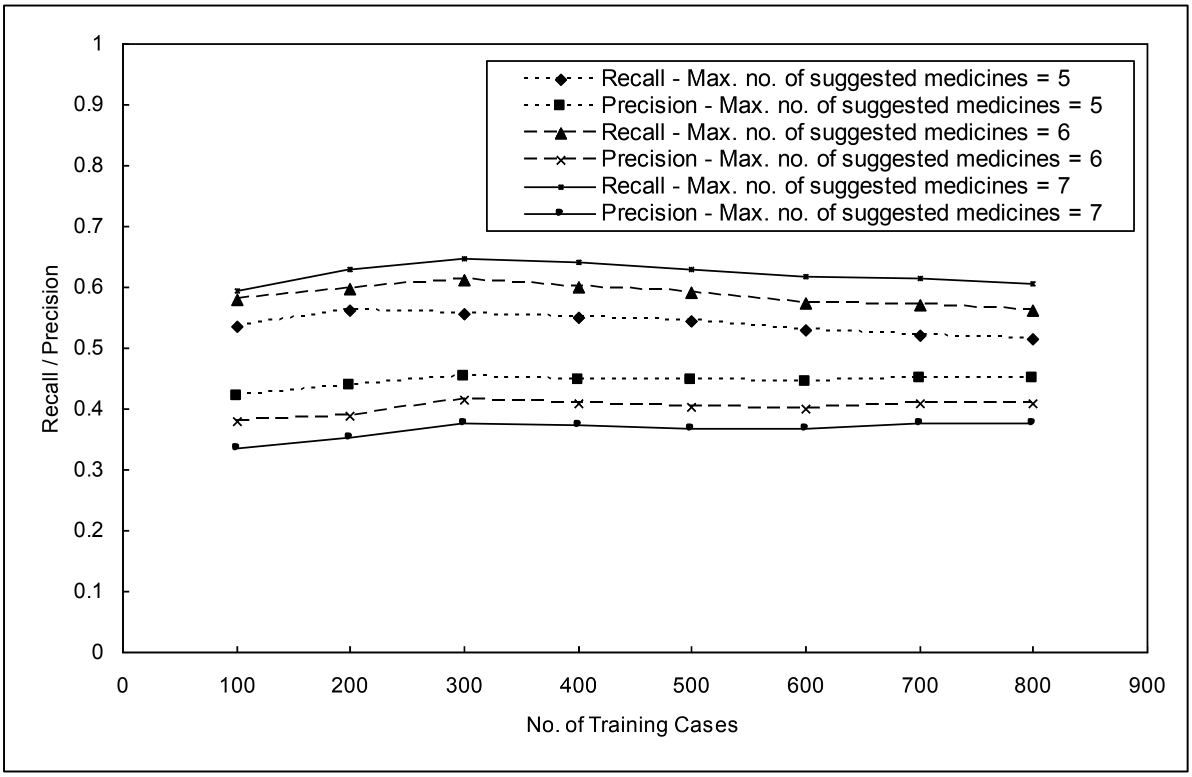


Figure 5.8 – The Precision and Recall of RACER with Different Sets of Maximum of Suggested Medicines (Minimum Support = 0, Minimum Confidence = 0, Maximum No. of Retrieved Cases = No. of Training Cases/10)

5.2.2 Design of the Experiment Setup for Performance Evaluation of Drug Information Extraction Algorithm

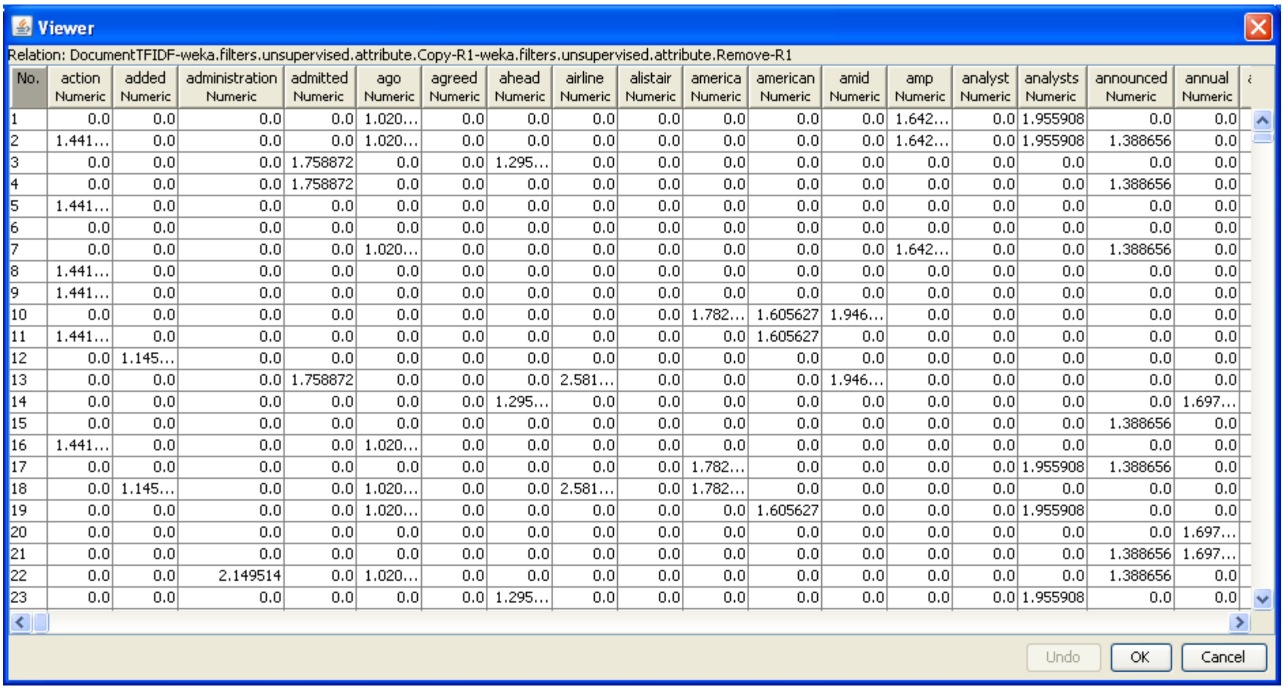
The goal of the Drug Information Extraction Algorithm is to identify the document with drug-drug interactions found, therefore an experiment is used to determine whether this algorithm can correctly classify the documents in specific categories.

5.2.2.1 Description of Date Collected

In this performance evaluation, 400 documents, where 100 documents for each predefined category (i.e. disease, therapy, drug and vaccine) are provided by the physicians of HPMS and used to evaluate the proposed document classification methodologies. To start with the evaluation of the classifiers' performance, the total 400 documents are split into two datasets, namely training set and testing set, in which 30% of the documents go into the training set, whereas the remaining 70% go into the testing set. The training set is the initial set of verified documents within each category for the formulation of solution to solve problems. The testing set is the controlled set with problems and reference answers for evaluating the performance of the different classifiers. In the representation of these documents, they have been vectorized into 1311 attributes (in term of numerical values) and 1 solution attribute (in term of nominal values). No missing data is among the attributes and all the numeric attributes are described in the Term Frequency-Inverse Document Frequency (TFIDF). An example of the data can be presented as Figure 5.9 and Table 5.9 summarizes the description data in both training and testing set.

5.2.2.2 Measure and Procedure

To evaluate the classification performance of the algorithm, a controlled simulation has been carried out which validates the system performance of the proposed Naïve Bayes (NB) classifier with other classifiers, which include Support Vector Machines (SVM), Neural Network (NN) and Decision Trees (DT). In particular, WEKA (Hall et al., 2009), an open-source data mining toolkit, is employed for the classifiers' performance evaluation. It compares the actual classification result of the testing set with the predicted classification result generated by the selected classifier.



Viewer

Relation: DocumentTFIDF-weka.filters.unsupervised.attribute.Copy-R1-weka.filters.unsupervised.attribute.Remove-R1

No.	action Numeric	added Numeric	administration Numeric	admitted Numeric	ago Numeric	agreed Numeric	ahead Numeric	airline Numeric	alistair Numeric	america Numeric	american Numeric	amid Numeric	amp Numeric	analyst Numeric	analysts Numeric	announced Numeric	annual Numeric
1	0.0	0.0	0.0	0.0	1.020...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.642...	0.0	1.955908	0.0	0.0
2	1.441...	0.0	0.0	0.0	1.020...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.642...	0.0	1.955908	1.388656	0.0
3	0.0	0.0	0.0	1.758872	0.0	0.0	1.295...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	1.758872	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.388656	0.0
5	1.441...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	1.020...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.642...	0.0	0.0	1.388656	0.0
8	1.441...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	1.441...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.782...	1.605627	1.946...	0.0	0.0	0.0	0.0	0.0
11	1.441...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.605627	0.0	0.0	0.0	0.0	0.0	0.0
12	0.0	1.145...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13	0.0	0.0	0.0	1.758872	0.0	0.0	0.0	2.581...	0.0	0.0	0.0	1.946...	0.0	0.0	0.0	0.0	0.0
14	0.0	0.0	0.0	0.0	0.0	0.0	1.295...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.697...
15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.388656	0.0
16	1.441...	0.0	0.0	0.0	1.020...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
17	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.782...	0.0	0.0	0.0	0.0	1.955908	1.388656	0.0
18	0.0	1.145...	0.0	0.0	1.020...	0.0	0.0	2.581...	0.0	1.782...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
19	0.0	0.0	0.0	0.0	1.020...	0.0	0.0	0.0	0.0	0.0	1.605627	0.0	0.0	0.0	1.955908	0.0	0.0
20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.697...
21	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.388656	1.697...
22	0.0	0.0	2.149514	0.0	1.020...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.388656	0.0
23	0.0	0.0	0.0	0.0	0.0	0.0	1.295...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.955908	0.0	0.0

Undo OK Cancel

Figure 5.9 – Experiment Data for the Drug Information Extraction Algorithm

Table 5.9 – Data Description for the Drug Information Extraction Algorithm

	Training Data	**Testing Data**
Number of instances	1200	2800
Number of attributes	1312 (Numeric – 1311; Nominal -1)	1312 (Numeric – 1311; Nominal -1)
Missing data	No	No

To test and evaluate the model, 70% of the dataset are used. Instances are extracted and then served as a benchmarking dataset for machine learning problems. By comparing the actual class of the instance with the predicted one (i.e. generated by the classification model), system performance can be measures in term of recall, precision, and F-measure. These can be mathematically defined as below.

$$\text{recall} = \frac{\text{Number of documents retrieved that are relevant}}{\text{Total number of documents that are relevant}} \tag{5.3}$$

$$\text{precision} = \frac{\text{Number of documents retreived that are relevant}}{\text{Total number of documents that are retrieved}} \tag{5.4}$$

$$\text{F-measure} = \frac{\text{2 x recall x precision}}{\text{recall + precision}} \tag{5.5}$$

5.2.2.3 Result of Classification Performance

The model is built based on the “Naïve Bayes” classifier developed in Weka. Table 5.10 summarizes the result of using Na ïve Bayes classifier to classify the documents. However, it surprisingly finds that the results of preprocessed dataset (95.5%) are worse than those which have not preprocessed (96.9%). Therefore, it is required to adjust the preprocessed model in order to achieve a better result. Considering the preprocessing phase is common to adopt in all case, therefore the adjustment is made in the feature selection phase. In the present study, Cfs Subset Evaluator and rank search (with Gain ration metric) are used for the feature selection. Therefore, another technique for rank search has been tried to adopt. This time, Chi-square feature selection has been adopted and 89 attributes are selected (Figure 5.10).

Table 5.10 – Classification Accuracy of Naïve Bayes Classifier (By Using the Dataset with Preprocessing and without Preprocessing)

	Correctly Classified Instances	**Incorrectly Classified Instances**	**Precision**	**Recall**	**F-Measure**
Without Preprocessing and Feature selection	2713 (96.9%)	87 (3.1%)	0.969	0.969	0.969
With Preprocessing and Feature selection	2675 (95.5%)	125 (4.5%)	0.956	0.955	0.955

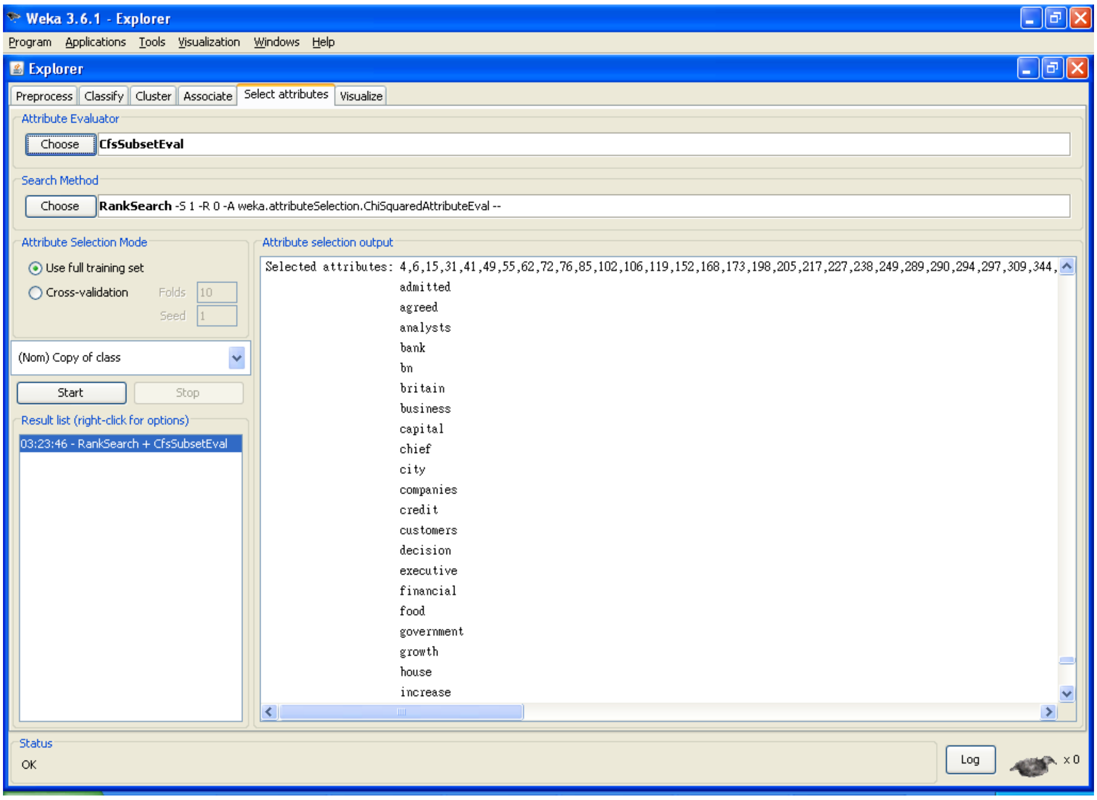


Figure 5.10 – Feature Selection Result using the Cfs Subset Evaluator and Rank Search (with Chi-square Feature Selection)

The result has been improved after using Chi-square feature selection, as depicted in Table 5.11. It is proven that preprocessing and feature selection are useful in achieving better classification result. Furthermore, another critical point can be found is that the time used to build the model is significantly improved after the number of features has been greatly reduced from 9.66 seconds to around 0.19 seconds (Table 5.12).

Table 5.11 – Classification Accuracy of Naïve Bayes Classifier (By Using the Dataset with Different Feature Selection Techniques)

	Correctly Classified Instances	**Incorrectly Classified Instances**	**Precision**	**Recall**	**F-Measure**
Without Preprocessing and Feature selection	2713 (96.9%)	87 (3.1%)	0.969	0.969	0.969
With Preprocessing and Feature selection – Gain Ratio	2675 (95.5%)	125 (4.5%)	0.956	0.955	0.955
With Preprocessing and Feature selection – Chi-square	2717 (97.0%)	83 (3.0%)	0.970	0.970	0.970

After discussing the importance of preprocessing and feature selection, experiment is to test whether naïve Bayes is the best classifier among other classifiers. To serve for this purpose, three different classifiers have been applied for testing. These classifiers are: SVM (the “SMO” function in WEKA), NN (the lazy “IBk”), and DT (the tree “J48”). In this experiment, the preprocessed dataset (with 90 attributes) are used for evaluation. Table 5.13 summarized all the accuracy results with the precision, recall, and F-Measure. As shown in the table, the accuracy result of naïve Bayes is the best among other classifiers. Although SVM gets similar results as Naïve Bayes, the times taken to build the model is dissatisfactory. Compared with the times used for building a Naïve Bayes classifier (0.19 seconds), SVM requires 2.69 seconds, which is 14 times of Naïve Bayes classifier, as depicted in Table 5.14. As a result, Naïve Bayes is reported to be the best text classifier.

Table 5.12 – Times Taken to Build the Naïve Bayes Classifier (By Using the Dataset with Preprocessing and without Preprocessing)

	Times taken to build model (seconds)
Without Preprocessing and Feature selection	9.66
With Preprocessing and Feature selection – Gain Ratio	0.14
With Preprocessing and Feature selection – Chi-square	0.19

5.3 Results of MedicPDSS Implementation in HPMS

5.3.1 Users’ Feedbacks

After implementing the medical prescription support approach to facilitate decision support in the drug selection process, the performance result is compared with those derived from the existing experience-based approach (i.e. based on the human experience and knowledge to make the prescription). Eight GPs work on rotation in two different clinics and they use the system in the course of their normal work during the evaluation period (i.e. 1 October 2010 to 28 February 2011). They were invited to provide user feedback about the usage of the system through interviews. The purpose of the interview was concerned with the following dimensions:

- User satisfaction: Is the system useful for them?
- Ease of use: Is it easy to learn and use?
- Flexibility: Is it easy to cope with developments in the future?
- Effectiveness: Can the system provide the appropriate prescription references to GPs? Can the system reduce errors in prescription?

Table 5.13 – Classification Results of Different Classifier

	Correctly Classified Instances	**Incorrectly Classified Instances**	**Precision**	**Recall**
Naïve Bayes	260 (93.0%)	20 (7.0%)	0.930	0.930
SVM	259 (92.9%)	21 (7.1%)	0.921	0.921
NN	246 (88.0%)	34 (12.0%)	0.880	0.880
DT	242(86.4%)	38 (13.6%)	0.864	0.864

Table 5.14 – Times Taken for Each Classifier to Build Model

	Times taken to build model (seconds)
Naïve Bayes	0.19
SVM	2.69
NN	9.66
DT	1.8

The result of the interviews is summarized and presented in Table 5.15. From the result, it is found that the physicians agree that the system can improve their work in the different dimensions mentioned above; and GPs are willing to use it in future. Furthermore, most young physicians report that they welcome MedicPDSS since it allows them to acquire more prescription knowledge from their seniors. In particular for the new medicine selection, they commented that more attention has been paid to the peer-based prescription decisions. Although some physicians are refused to use the computerized system as they are not so familiar with general computer skill, they claimed that they can share their prescription decision and experiences to peers interactively. They also commented that they will treat the knowledge retrieved by MedicPDSS is a kind of advisory information for them to learn more from a large of peers, especially in the case of encountering unacquainted situations. Although MedicPDSS cannot provide the golden standard of prescription and concept of evidence-based medicine (due to the retrieved knowledge does not take any critical examination), one point the physicians all agreed is that the information of MedicPDSS is more objective than that in the past knowledge extraction method (e.g. attending seminars).

Table 5.15 – User Feedback for the MedicPDSS Performance

	Very Dis -satisfied	**Dissatisfied**	**Normal**	**Satisfied**	**Very Satisfied**
Overall system performance					
Data input	0%	10%	20%	45%	25%
Information retrieval	0%	0%	25%	50%	25%
Decision support function	0%	15%	25%	40%	20%
Data input					
Efficiency (compared with the old process)	0%	5%	20%	50%	25%
Simplicity	0%	10%	45%	30%	15%
Design of user interface	0%	10%	30%	30%	30%
Information retrieval					
Correctness of content	0%	25%	30%	20%	25%
Sufficiency of content	0%	15%	25%	40%	20%
Ease to understanding	0%	0%	40%	50%	10%
Decision support function					
Efficiency (compared with the old process)	0%	0%	30%	45%	25%
Usefulness of prescription advice	0%	15%	40%	25%	20%

5.3.2 Evaluation of the Hit Rate in the Three Ranks of MedicPDSS

As proposed, three different ranks (i.e. Rank A, B, and C) are introduced in the proposed methodology. To verify the performance in each rank (i.e. the ratio of the number of correct medicine(s) produced in each rank among the total number of existing relevant medicine(s) in each rank), we measured the hit rate of GPs in each visit. After the clinical investigation performed by the physician, a range of drugs will be recommended and listed under Rank A, B or C. The hit rate refers to the number of matches between the MedicPDSS's recommendations and the drugs actually prescribed by the physicians. The experiment setup is depicted in Figure 5.11.

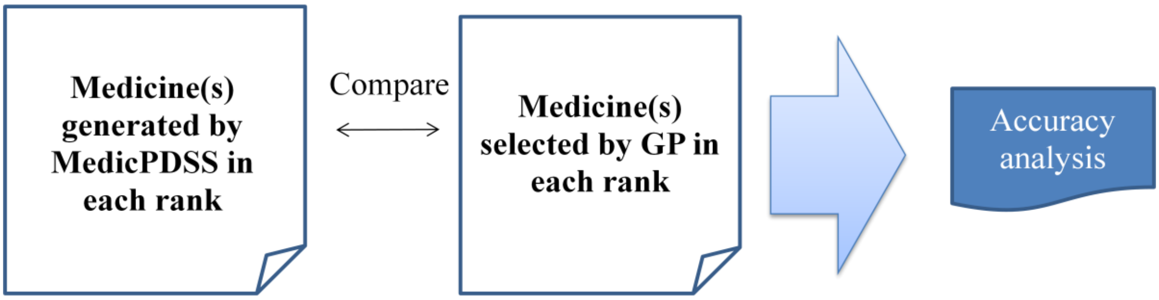


Figure 5.11 – The Experiment Setup for Measuring the Hit Rate of MedicPDSS

The results of performance evaluation in different ranks are shown in Table 5.16. It is noted that the hit rate of solution retrieval of Rank B is higher than that of Rank A and Rank C because most of the medicines are obtained using either CBR or ARM. From the result, the suggested medicines allow the physician to decide on a prescription because on average at least one medicine has been prescribed in each rank. Furthermore, most of the medicines that will be prescribed can be found in either Rank A or Rank B, in which physicians can select around 2 to 3 medicines (out of the actual solution of 5 to 7 medicines being selected) in the recommended medication list in MedicPDSS. These results show that the proposed system allows physicians to identify the required drugs easily.

Table 5.16 – Evaluation of Medicines Selected in Different Ranks of MedicPDSS

	Rank A	**Rank B**	**Rank C**
Average hit rate	1.40	1.87	1.47
Standard Derivation	0.81	1.01	0.90
Minimum number of medicine retrieved	0	0	0
Maximum number of medicine retrieved	2	4	3

5.4 Lessons Learnt by Exploring the Issues Raised in the Implementation of MedicPDSS

MedicPDSS is an enabling technology, as illustrated in the case study, that a healthcare organization can adopt to reduce the user resistance in new technology introduction and improve patient safety (like reducing errors in prescribed decision and enhancing the detection of drug-drug interactions) as well as business operations (like enhancing the management and retrieval of medical records, improving the registration and treatment process, and so on). However, it is learnt that several issues have been occurred during the implementation in which the practitioners should be aware of. The major issues of implementation are described below.

5.4.1 Cost Issues

Information technology involves people to invest on huge amount of money for the acquisition of hardware and software. Compared with traditional paper-based approach, the company requires to purchase more computers (with Internet-enabled connection) and printers for the MedicPDSS. Furthermore, there are limited success stories in adopting such technologies in which it is difficult to realize the return-on-investment (ROI) of such implementation (Miller and Sim, 2004). Subsequently, the new technology adoption will be adversely affected by both factors. A careful cost-benefit estimation and evaluation is required to make for overcoming this limitation.

5.4.2 Security Issues

Since the MedicPDSS are capable to store all the patient information, such as their demographic information, past medical history, and current health condition, security audit and user authentication are important to ensure the confidentiality and security of the data. Furthermore, the inter-systems featured in the MedicPDSS are connected and communicated via the Internet, in which data protection under transmission should be placed as the primary issues surrounding the adoption of MedicPDSS to protect the privacy and integrity of the patient's information.

5.4.3 Human Issues

Although there are training and educational program for operating the system, it is realized that people may not have confidence in the new system because it is relatively new to them. On the other hand, user resistance to change is another critical factor that impacts the implementation of MedicPDSS. Consequently, significant amounts of support, in both the phase 3 (i.e. implementation) and 4 (i.e. post-implementation), are required to enhance their trust in adopting the technology. In addition to this behavioral matter, there is scarce resource of healthcare IT professionals in Asia, especially in Hong Kong. Thus, improving the computer skill of medical staff remains an open question in future works.

5.4.4 Technical Issues

MedicPDSS heavily depends on both the hardware and software; therefore when either one of each is malfunction, the users will no longer proceed to use the system. Consequently, level of technology satisfaction may decrease which affect the system adoption of users. Thus, a technical support is necessary to remain both the hardware and software remain available and accessible in a timely manner (Wager et al., 2001). Furthermore, various risk management tools (such as business continuity planning and disaster recovery planning) should be employed to respond immediately to prevent service breakdown and get back to work after disruption (Barlow et al., 2007).

5.4.5 Data Migration Issues

In current practice of Hong Kong medical organization, physical paper-based records are used to store the patient information and their past medical history, however all this information is difficult to convert or migrate to the database featured in MedicPDSS. Furthermore, lack of integration with other applications is also explored. Efficient data migration methods may be required for implementation.

5.4.6 Standardization Issues

There is lack of standardization on codes. The terminological data standards would still remain an open question for both healthcare delivery and clinical care (Richesson and Krischer, 2007). It is important to reach a consensus on using the consistent standards and medical terminologies in information sharing across all the parties. For example, in this case study, the company and its partners agreed to use the Heath Level Seven (HL7) messaging standard for the exchange, management and integration of data that support clinical management, delivery and evaluation of healthcare services.

5.4.7 Ethical Issues

The decision making process of GPs is first considered as an ethical issue. It is noted that few physicians might have the perception that MedicPDSS was designed as an autonomous system that replaces their human judgment. In this way, the right of a patient to obtain the best form of medical treatment or service is assured. Given that MedicPDSS aims to enhance knowledge in the medical prescription process, a list of appropriate medicines (instead of several medicines) will be generated in the system. In this regard, physicians can make use of this information (or they may even ignore the information) and their own clinical judgment to provide the most suitable medication to a patient. In other words, it is important to let the physician understand that the proposed system is a kind of decision support tool on which they should not completely rely in making decisions.

Another ethical issue is related to privacy and confidentiality of the information provided by the patients and physicians. It is recognized that MedicPDSS makes use of the electronically stored health information to infer the medical prescription decision support. In this way, the privacy and confidentiality of the information provided by a patient is not entirely recorded in the TEMRS, but rather it is retrieved for use in the MedicPDSS. It is claimed that confidentiality and privacy might be threatened with the use of such a system. Thus, one of the solutions to counteract this issue is to get the consent of patients, making them understand that the information is used for enhancing the case base of the system and will not be used for other purposes such as education and commercial purposes. Another solution is to introduce carefully thought-out policies that outline the system use of permissions and restrictions to reduce any ethical lapses.

5.5 Limitations and Assumptions of the Study

The limitations of both CBR and ARM have been mentioned earlier (i.e. it is impossible to rely solely on past experience to treat the current situation), and it was observed that our approach works excellently when the patient condition in each visit was similar. In this case study, we found that more than 75% of the visits were similar to each other (e.g. the patients got similar or the same diagnoses) in which the drug selection is nearly the same as the previous visit. This may be due to the reasons that GPs employ the same rules or standards to treat the patients for the same diseases each time. On the limitation of the case study and the experimental set-up, the size of the knowledge base (i.e. the company and GPs involved) and the number of drugs available in the database are too small. It is aware that the relatively small data set does limit the findings of the study, however it is believed that the results obtained show that MedicPDSS could be applied to a larger store of records in making suggestions on a range of medicines that could be used in medical prescription. Normally, more information can provide better decision support in CBR and ARM. As the approach can be launched in other medical centers, the knowledge base and drug information have the potential to grow rapidly, and become more knowledgeable to support the current complex medical prescription problems. It is also interesting to note that even though making the prescription is a complex process involving numerous variables (up to a hundred) in making decision, the proposed decision support approach can greatly assist the domain expert by reducing the prescription choices and by identifying appropriate medicine for the physician's consideration. There is considerable saving in time compared with the conventional statistical approach for retrieving the previous prescription of each patient.

There are also few limitations of this study. First, the physician sample using the system was small and a much larger number of physicians should be participated in further study to provide a more complete assessment. Second, there are twenty diagnosis encountered in the medical organization selected in this case study. For a more comprehensive study, one would expect to assess more diagnoses that demonstrate the feasibility of MedicPDSS under different situation. Third, each organization has its own unique environment and motivation towards the new technology adoption; and this may induce biases in the assessment. However, despite these limitations, the current study attempted to identify essential factors to consider when implementing MedicPDSS in other healthcare organizations.

In this study, we assume that a prescription is useful when most of the physicians, are seen to make use of it. So we summarize all related prescription information to form peer-based evidence, instead of using individual knowledge. This is a powerful decision support method that allows for acquiring the knowledge of a large group of physicians and hence model such knowledge through a knowledge-based system to support further decision making. These issues constitute interesting and promising directions for future research in how to enhance the quality of knowledge sharing in the decision making context.

5.6 Discussion of MedicPDSS Contribution

One of major aims in MedicPDSS is to automatically capture the knowledge stored in the EMR for building the diagnostic template. Compared with the manual template formation, Table 5.17 highlights that, as far as practical aspects of knowledge elicitation are concerned, in comparison with the conventional manual approach, the automatic approach has the advantage of less subjectivity and in being more reliable in knowledge sharing with less human effort, and the overall quality of medical services can be improved by means of this collective knowledge.

Furthermore, the proposed integrated approach that utilizes collective perspective (i.e. macro-view) results in supporting the individual perspective (i.e. micro-view) to enhance decision support in the medical prescription process. The rationale of integrating ARM into CBR is to provide double loop learning that uses peer-based evidence to provide more information about a past solution retrieved in isolation (i.e. single loop learning). Table 5.18 highlights that, as far as practical aspects of knowledge sharing are concerned, in comparison with the CBR approach alone, MedicPDSS presents the advantages of combining the strength and complementing the weakness of conventional CBR-based Knowledge-based System (KBS).

Table 5.17 – Comparison between Conventional and Automatic Knowledge Elicitation Methods

Criteria	**Conventional Knowledge Elicitation Method**	**Automatic Knowledge Elicitation Method**
Level of human-effort required	High and human-based	Low and automatic
Quality of knowledge elicited	More subjective and not guaranteed as it is based on individual physician's knowledge and experience	Less subjective and more reliable as it is based on a group of physicians
Effectiveness	Less effective as it heavily depends on human-effort and time consuming	More effective as it is a systematic and automatic approach

In summary, MedicPDSS can contribute to the industry in both the decision support and knowledge sharing perspectives:

(i) Enhancement on prescription decision support

With the growth in the amount of information about drugs, it is difficult for physicians to make a good prescription without a flexible drugs list. Mistakes in prescription are not only harmful but in serious cases they can also be fatal. The MedicPDSS approach presented makes use of CBR to retrieve the micro-view of the physician's practices and ARM to model the macro-view. Subsequently, drug-drug interaction rules are captured from the literature and used to support the prescription. As a result, better decision support can be achieved by leveraging the internal (i.e. the physicians'' prescription decisions) and external (i.e. the reliable online information from the journal articles) source of prescription knowledge.

(ii) Knowledge sharing among physicians

In the past, knowledge sharing has been hindered due to various reasons, such as resource limitation and ineffective knowledge elicitation techniques. By means of the automatic knowledge elicitation technique and the knowledge representation/modeling technique for the decision support system, less human effort and time will be required. These improvements on efficiency and effectiveness are believed to facilitate knowledge sharing among physicians and ultimately enhance the quality of medical services provided.

Table 5.18 – Comparison of Conventional KBS and MedicPDSS

Criteria	**CBR-based KBS**	**MedicPDSS**
Quality of shared knowledge	More subjective as it is based on individual physician's knowledge and experience	More objective as it is based on large group of physicians
Interactivity	Information is retrieved through physician-patient and physician-diagnosis interaction	Information is retrieved through summarizing the peer evidence
Learning Cycle	Mostly single loop but sometimes can be double loop	Double loop
New Drug Selection	Depend on the physician's knowledge	Take into consideration the peer-based prescription decision to facilitate the own choice

5.7 Summary

In this chapter, the sub-system (i.e. TEMRS) and the three core modules (i.e. Automatic Knowledge Elicitation Module, Prescription Modeling Module, and Drug Information Extraction Module) of the MedicPDSS are evaluated. A series of experiments have been conducted for evaluating the performance of the modules. A survey was distributed for measuring the performance of the TEMRS of the Automatic Knowledge Elicitation Module. The results show that the users are satisfied with the system and they agree that the provision of diagnostic template is valuable for them in better acquiring the medical knowledge.

The performance of the RACER algorithm (in the Prescription Modeling Module) was evaluated by measuring the accuracy of the predictive function in a medical organization with real prescription cases. The accuracies of solution among different methodologies (i.e. CBR, ARM, and RACER) are measured. The results show that the accuracy of the RACER approach is higher than that of applying CBR or ARM separately.

The performance of the Drug Information Extraction Module was evaluated by measuring the precision and recall rate of the number of valid classified instances using 400 medical document provided by the users of the case company. The results show that the NB classifier outperforms the other benchmarking classifiers (such as support vector machine, neural network, and decision tree) in terms of precision and recall rate as well as the model building time.

Other than the performance evaluation of each module, two experiments have been carried out to evaluate the MedicPDSS as a whole. The first experiment is to administrate a questionnaire to the system users. The results indicate that the users agree that the system can improve their work in the prescription context and they are willing to use it in future.

Furthermore, young physicians generally welcome to use the MedicPDSS because they can acquire more prescription knowledge from their seniors. The second experiment is to evaluate the hit rate in the three ranks of the MedicPDSS. The results show that out of the actual solution of five to seven medicines being prescribed, around two to three medicines are selected in the first two ranks (i.e. Rank A and B). Therefore, it is concluded that the concept of considering both specific and general knowledge is useful and similar to the physicians' prescription behaviors.

Chapter 6 Conclusions

6.1 Research Summary

Medical prescription is a critical phase in the medical treatment of a disease. Up to now, healthcare organizations have depended heavily on knowledge in terms of patient medical history, drug prescription procedure, hazard reports and medical expertise. Better and effective utilization of these resources can enhance patient safety and alleviate medical errors. On the other hand, knowledge within the healthcare organization is internal information (i.e. know-how) which is based on past solutions, experience and rules to determine the services delivered. As a consequence, it is difficult for every clinical officer to transfer their knowledge to others. Thus, it is necessary for clinical officers working in the healthcare industry to be able to acquire others' experience and hence be intelligently alert to the possibility of any abnormal processes occurring.

There is a need for the establishment of a methodology, which enables knowledge extraction, retrieval, transfer, and reuse in physicians' prescription, as well as fulfilling the decision support in the medical prescription process. A bottom-up approach for the collection of medical prescription logics and decisions is much needed for saving the time and improving the patient safety with the updated knowledge. Electronic Medical Record (EMR) is an emerging tool used in medical informatics to computerize medical records and to establish a knowledge sharing platform among physicians (Hersh, 2009; Herschel et al., 2001). Since EMR stores various items of important medical data, it is argued that these data items can be turned into knowledge that is valuable in making clinical decisions. Furthermore, EMR is an explicit medical record that stores the physician's tacit knowledge being deployed in each diagnostic process (Herschel et al., 2001). However, discussions on approaches for eliciting knowledge from information stored in EMR are rare in the literature. Furthermore, integrating the template concept into traditional EMR is an essential component for an EMR to be successfully implemented. Automatic knowledge elicitation for building diagnostic template and probabilistic technique are incorporated to generate a flexible template formation, so that physicians can determine the therapy with suitable diagnostic concept in a time effective manner. With the advent of Template-based EMRS (TEMRS), related diagnostic concepts are captured and codified for further decision support in the medical prescription process.

Decision Support System (DSS) has been proposed as one of the most effective ways of medication errors reduction, since it integrates both knowledge-based and expert-based concepts to support GPs in selecting and deciding appropriate medicines to cure the patient (Garg et al., 2005). In the fact that physicians wish to rely directly on the past experience that stored in the historic patient data, select similar cases that had reliable outcomes and reuse the solution accordingly, which works similar to the inference process of DSS. Therefore, the quality of a DSS is highly depended on its inference mechanism. Based on the literature review, Case-based Reasoning (CBR) utilizes the specific knowledge of previously experienced and concrete problem situations (cases), while association rules mining relies on general knowledge of a problem domain and making associations along generalized relationships between problem descriptors and conclusions (Zhuang, et al., 2009). They are two distinct techniques that consist of their own strengths and limitations whereas important contributions have been made by integrating CBR and rules in numerous applications. However, lack of researches and empirical investigations have been done for the prescription related topics. Therefore, this study is focused on improving the solution extracted in CBR (especially the missing medicines) and providing relevant and objective evidences in the prescription support. It is a novel measure to rank the multiple values solution by combining the results from CBR and association rules mining, so as to assist the physicians in identifying which medicines are more appropriate for the patients.

A Medical Prescription Decision Support System (MedicPDSS) is designed to address the above mentioned limitations and challenges. MedicPDSS is capable of extracting and modeling comprehensive individual and collective prescription behaviors, with good accuracy, has been proposed in this project. With the growing amount and increasing complexity of information about drugs, it is difficult for physicians to make a good perception without a flexible drug list. Mistakes in prescription are not only harmful; in serious cases they can be fatal. The prescription support system presented makes use of CBR to retrieve the micro-view of the physician's practices and ARM to model the macro-view; subsequently, a rule-induction matching algorithm is introduced to match the results and hence categorize them intelligently into a drug list. The physician can then select the drugs and make informed decisions on prescription by taking into consideration the decisions made by other physicians in similar cases. Using the information and knowledge in sources available to the public, each physician's prescription choices can be monitored and checked for any drug contraindications so as to increase the safety of the drug prescribed.

6.2 Contribution of the Research

MedicPDSS's key contributions to healthcare organizations, as well as the physicians, are highlighted as follows:

(i) The unique feature of MedicPDSS is that it helps healthcare professionals achieve the prescription decision support initiative in recommending a range of medicines and taking into consideration of the updated drug-drug interactions for improving the prescription safety. It fills the gap in current prescription decision support research, which solely focuses on prescription rules formation manually, and thus provides a new research direction on system development using the state-of-the-art methodology.

(ii) MedicPDSS can elicit the knowledge stored in the traditional EMRS. Such knowledge elicitation and representation approach represents a breakthrough in knowledge-based system applications when compared with other existing knowledge elicitation solutions. Although the transformation of medical knowledge from EMRS into readable schema is still in its infancy, the present study highlights the practicability and usefulness of such a concept. This provides support and consideration for healthcare organizations in particular to further utilize their stored information (in EMRS) into meaningful and useful knowledge to support the medical judgments of physicians.

(iii) Knowledge is the key element to enhance the medical judgments. However, the current knowledge sharing practices is not well-developed because physicians do not share what they know with other parties. Even though each physician has the knowledge to make the prescription, it is important for them to learn from others' experiences as well. MedicPDSS provides a novel approach to automatically capture the prescription decisions to enhance the sharing amongst physicians through the determination of peer-based collective intelligence.

(iv) Unlike other medical domains (such as cancer diagnosing), the conclusion of decision support of prescription is more complex that consists of a number of medicines. Each medicine out of hundreds of medicines can be a part of the solution in prescription making. However, handling of multi-prescription values solution received lack of concern in the domain. In order to fill the research gap, MedicPDSS is the first model attempted to handle the multiple values solution by assigning weights to the medicines that retrieved from the prescription solutions.

(v) Traditionally, inspections are carried out only when patients are found to be sick after taking wrong medicine (such as the drug-drug interaction). However, with the help of MedicPDSS, the current situation is improved by bringing all the problematic cases to the attention in real time. Having the capability of capturing updated drug-drug interactions rules from the reliable literature, drug safety is assured consistently. On the other hand, patient safety is enhanced by reducing medication errors.

6.3 Research Limitations

The limitations of the research are addressed as follows:

(i) The scope of the proposed system is aimed to tackle the medical prescription for general diseases (such as influenza, U.R.T.I., etc.), while other special disease (such as cancer, asthma, etc.) are not considered.

(ii) The proposed system is incapable to deal with the therapeutic decision other than medicine prescription, such as laboratory testing, and vaccination.

(iii) The prescription modeling is only focused on the demographic information and medical history of the patient, the factors related to drug information (such as the drug cost) are not considered.

(iv) The proposed system requires a large amount of medical cases to perform the prescription modeling. To better model the prescription decisions to support the decision marking in new situations, it is necessary to store enough qualified cases in the repository. This requires healthcare professionals to validate the input cases, including the diagnostic decision as well as the medicines to be prescribed in each case, and it is a very time-consuming task.

(v) The computational processing time of CBR and ARM depends heavily on the number of cases stored in the repository. The retrieval time will become longer if the repository is large. Thus, a computer server with better quality should be used to fasten the processing time. Thus, the implementation cost of the proposed system becomes high when system users purchase a server with better processing power.

6.4 Suggestions for Future Works

The following suggestions for future works of MedicPDSS are provided to improve the system's capability and adaptability:

(i) The proposed system is implemented in a single case study (i.e. a medical clinic), as described in Chapter 4. It is suggested that the system can be implemented in other types of medical organization, such as hospital and testing laboratory, to further validate the system.

(ii) The proposed system is focused on the general diseases, therefore, it is suggested that the system can be employed in other specific diseases, such as cancer, dental services, to further validate the modeling methodology.

(iii) With the consideration of using demographic information and medical history of the patient in prescription behavior modeling, the next research stage is to select the appropriate features and parameters in order to enhance the decision support quality in the prescription solution and also to design a user-friendly interface for the GPs to apply MedicPDSS in their daily operations.

(iv) The present study is applied to text-based medical information. The integration of multimedia such as medical images and sound effects is needed. It will be an interesting area to investigate the integration of visual and aural media along with written medical records. For example, adoption of speech recognition technique is one of the possible directions to capture the conservation of patients and physicians for better informing the clinical decisions.

To summarize the benefits of this study, MedicPDSS is developed in a generic healthcare environment that has the potential and flexibility in decision support for further assisting in the process of the prescription of medicine. Moreover, the proposed system that has been developed can also be employed as a simulation or training program for medical students and even some young physicians so they can learn the appropriate prescription for a given situation. As a result, all the healthcare organizations (i.e. hospitals and clinics), the Government health department, physicians, and universities can use the system to enhance the quality of health services and improve the pool of medical knowledge. This is a powerful knowledge sharing method that allows for acquiring the knowledge of a large group of physicians and hence model such knowledge through a KBS to support further decision making. These issues constitute interesting and promising directions for future research in how to enhance the quality of knowledge to be modeled in the decision support context.

References

Aamodt, A. and Plaza, E. (1994), “Case-based reasoning: foundational issues, methodological variations, and system approaches”, *Artificial Intelligence Communications*, Vol. 7, No. 1, pp. 39-59.

Abidi, S. (2001), “Knowledge management in healthcare: Towards ‘knowledge-driven’ decision-support systems”, *International Journal of Medical Informatics,* Vol. 63, No. 1-2, pp. 5-18.

Agrawal, R. and Strikant, R. (1994), “Fast algorithm for mining association rules”, in *Proceeding of the 20th International Conference on Very Large Data Bases*, Morgan Kaufmann, Santiago, pp. 487-499.

Aizawa, A. (2003), “An information-theoretic perspective of TF-IDF measures”, Information *Processing and Management,* Vol. 39, No. 1, pp. 45-65.

Anderson, J.G. (2007), “Social, ethical and legal barriers to E-health”, *International Journal of Medical Informatics,* Vol. 76, No. 5-6, pp. 480-483.

Anis A.H., Carruthers S.G., Carter A.O., and Kierulf J., “Variability in prescription drug utilization: issues for research”, *Canadian Medical Association Journal*, Vol. 154, No. 5, pp. 635-640.

Antonie, M.L., Zaiane, O.R., and Coman, A. (2001), “Application of data mining techniques for medical image classification”, in *Proceedings Second International Workshop on*

Multimedia Data Mining, pp. 94-101.

Aronson, J.K. (2006), "A prescription for better prescribing", *British Journal of Clinical Pharmacology*, Vol. 61, No. 5, pp. 487-491.

Ash, J.S., Berg, M., and Coiera, E. (2004), "Some unintended consequences of information technology in health care: the nature of patient care information system-related errors", *The Journal of the American Medical Informatics Association*, Vol. 11, No. 2, pp. 104-112.

Atzmueller, M., Baumeister, J., and Puppe, F. (2003), "Evaluation of two strategies for case-based diagnosis handling multiple faults", in *Proceedings of the 2nd German Workshop on Experience Management (GWEM 2003)*, Luzern, Switzerland.

Baesens, B., Setiono, R., Mues, C., and Vanthienen, J. (2003), "Using neural network rule extraction and decision tables for credit-risk evaluation", *Management Science*, Vol. 49, No. 3, pp. 312-329.

Barbosa, J.G.P., Goncalves, A.A., Simonetti, V., and Leitao, A.R. (2009), "A proposed architecture for implementing a knowledge management system in the Brazilian national cancer institute", *Brazilian Administration Review*, Vol. 6, No. 3, pp. 247-262.

Barlow, S., Johnson, J., and Steck, J. (2004), "The economic effect of implementing an EMR in an outpatient clinical setting", *Journal of Healthcare Information Management*, Vol. 18, No.1, pp. 46-51.

Bates, D.W., Lucian, L.L., and Cullen, D.J. (1998), "Effect of computerized physician order entry on medication error prevention", *Journal of the American Medical Informatics Association*, Vol. 280, No. 15, pp. 1311-1316.

Bell, G.B. and Sethi, A. (2001), "Matching records in a national medical patient index", *Communications of the ACM*, Vol. 44, No. 9, pp. 83-88.

Berrar, D., Sturgeon, B., Bradbury, I., Downes, C.S., and Dubitzky, W. (2003), "Microarray data integration and machine learning methods for lung cancer survival prediction", in *Proceedings of the 4th International Conference on Critical Assessment of Microarray Data Analysis 2003*, pp. 43-54.

Bichindaritz, I. and Marling, C. (2006), "Case-based reasoning in the health sciences: what's next?", *Artificial Intelligence in Medicine*, Vol. 36, No. 2, pp. 127-135.

Bichindaritz, I. and Potter, S. (2004), "Knowledge based phylogenetic classification mining", in *Proceedings of the 4th Industrial Conference on Data Mining 2004*, pp. 163-172.

Boreham, N.C., Mawer, G.E., and Foster, R.W. (2000), "Medical students' errors in pharmacotherapeutics", *Medical Education*, Vol. 34, No. 3, pp. 188-193.

Bouaud, J., Seroussi, B., Antoine, C.C., Gozy, M., Khayat, D., and Boisvieux, J.F. (1998), "Hypertextual navigation operationalizaing generic guidelines for patient-specific therapeutic decisions", in *Proceedings of AMIA Symp.*, pp. 488-492.

Bradley, C.P. (1991), "Decision making and prescribing patterns - a literature review", *Family Practice*, Vol. 8, pp. 276-287.

Bradley, C.P. (1992), "Factors which influence the decision whether or not to prescribe: the dilemma facing general practitioners", *British Journal of General Practice*, Vol. 42, pp. 454-458.

Breault, J.L., Goodall, C.R., and Fos, P.J. (2002), "Data mining a diabetic data warehouse", *Artificial Intelligence in Medicine*, Vol. 26, No. 1-2, pp. 37-54.

Brender, J., Nøhr, C., and McNair, P. (2000), "Research needs and priorities in health informatics", *International Journal of Medical informatics*, Vol. 58-59, pp. 257-289.

Brossette, S.E., Sprague, A.P., Jones, W.T., and Moser, S.A. (2000), "A data mining system for infection control surveillance", *Methods of Information in Medicine*, Vol. 39, No. 4-5, pp. 303-310.

Carter, J.H. (2004), "How EMR software can help prevent medical mistakes", Available at http://www.acpinternist.org/archives/2004/09/emr.htm (Retrieved from 20 March 2011).

Chakrabarti, S., Roy, S., and Soundalgekar, M.V. (2003), "Fast and accurate text classification via multiple linear discriminant projection", *The International Journal on Very Large Data Bases*, Vol. 12, No. 2, pp. 170-185.

Chan, T.Y.K. (2006), "Prevention of medication error and unintentional drug poisoning in the elderly", *Hong Kong Medical Journal*, Vol. 12, No. 5, pp. 332-333.

Chen, H. (2001), *Knowledge management systems: a text mining perspective*, Tucson, AZ, The University of Arizona.

Chen, M.M., and Landefield, C.S. (1994), "Physicians' behavior and their interactions with drug companies", *Journal of the American Medical Association*, Vol. 271, pp. 648-649.

Chien, C.F. and Chen, L.F. (2008), "Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry", *Expert Systems with Applications*, Vol. 34, No. 1, pp. 280-290.

Cho, Y.H., Kim, J.K., and Kim, S.H. (2002), "A personalized recommender system based on web usage mining and decision tree induction", *Expert Systems with Applications*, Vol. 23, No. 3, pp. 329-342.

Christine, A.P. (2001), "Adverse drug-drug and food-drug medication interactions", Available at http://ohioline.osu.edu/ss-fact/0129.html (Retrieved from 1 March 2011).

Cios, K.J., Pedrycz, W., Swiniarski, R.W., and Kurgan, L.A. (2007), *Data mining: a knowledge discovery approach*, Springer, New York.

Classen S., Mkanta, W., Walsh, K., and Mann, W. (2005), "The relationship of classes of commonly prescribed medications to functional status and quality of life", *Physical & Occupational Therapy in Geriatrics*, Vol. 24, No. 1, pp. 25- 43.

Cohen, K.B. and Hunter, L. (2008), "Getting started in text mining", *PLoS Computational Biology*, Vol. 4, No. 1, Doi:10.1371/journal.pcbi.0040020.

Cooke, N.J. (1994), "Varieties of knowledge elicitation techniques", *International Journal of Human-Computer Studies*, Vol. 41, No. 6, pp. 801-849.

Cooke, N.J. (1999), "Knowledge elicitation", *Handbook of Applied Cognition*, pp. 479-509.

Cooper, D.R. and Schindler, P.S. (2003), *Business Research Methods*, McGraw Hill, New York.

Coscelli, A. (2000), "The importance of doctors' and patients' preferences in the prescription decision", *Journal of Industrial Economics,* Vol. 48, pp. 349-369.

Cothey, V. (2004), "Web-crawling reliability", *Journal of the American Society for Information Science and Technology,* Vol. 55, No. 14, pp. 1228-1238.

Coulter, D.M., Bate, A., Meyboom, R.H.B., Lindquist, M., and Edwards, R. (2001), "Antipsychotic drugs and heart muscle disorder in international pharmacovigilance: data mining study", *British Medical Journal*, Vol. 322, pp. 1207-1209.

Daengdej, J., Lukose, D., and Murison, R. (1999), "Using statistical models and case-based reasoning in claims prediction: experience from a real-world problem", *Knowledge-based Systems*, Vol. 12, No. 5-6, pp. 239-245.

D'Alessandro, D.M., Kreiter, C.D., and Peterson, M.W. (2004), "An evaluation of information-seeking behaviors of general pediatricians", *Pediatrics*, Vol. 113, No. 1, pp. 64-69.

Dawes, M. and Sampson, U. (2003), "Knowledge management in clinical practice: systematic review of information seeking behavior in physicians", *International Journal of Medical Informatics*, Vol. 71, No. 1, pp. 9-15.

Delaney, B.C., Fitzmaurice, D.A., Riaz, A. and Hobbs, F.D.R. (1999), "Can computerised decision support systems deliver improved quality in primary care?", *British Medical Journal*, Vol. 319, No. 7220, pp. 1281.

Delen, D., Walker, G., and Kadam, A. (2005), "Predicting breast cancer survivability: a comparison of three data mining methods”, *Artificial Intelligence in Medicine*, Vol. 34, No. 2, pp. 113-127.

Dempster, A.P. (1968), “Generalization of Bayesian inference”, *Journal of the Royal Statistical Society*, Vol. 30, pp. 205-247.

Denig, P., Haaijer-Ruskamp, F.M., and Zijsling, D.H. (1988), “How physicians choose drugs”, *Social Science & Medicine,* Vol. 27, No. 12, pp. 1381-1386.

DesRoches, C.M., Campbell, E.G., Rao, S.R., Donelan, K., Ferris, T.G., Jha, A., Kaushal, R., Levy, D.E., Rosenbaum, S., Shields, A.E., and Blumenthal, D. (2008), “Electronic health records in ambulatory care — a national survey of physicians”, *The New England Journal of Medicine*, Vol. 359, No. 1, pp. 50-60.

Downing, G.J., Boyle, S.N., Brinner, K.M., and Osheroff, J.A. (2009), “Information management to enable personalized medicine: stakeholder roles in building clinical decision support”, *BMC Medical Informatics and Decision Making*, Vol. 9, pp. 4.

Druzdzel, M. and Henrion, M. (1990), “Using scenarios to explain probabilistic inference”, in *Working Notes of the AAAI-90 Workshop on Explanation*, pp. 133-141.

Duda, R.O., Hart, P.E., and Stork, D.G. (2001), *Pattern classification*, Wiley and Sons, New York.

Dussart, C., Pommier, P., Siranyan, V., Grelaud, G., and Dussart, S. (2008), “Optimizing clinical practice with case-based reasoning approach”, *Journal of Evaluation in Clinical Practice*, Vol. 14, No. 5, pp. 718-720.

Ellis, D., (1996), “The dilemma of measurement in information retrieval research”, *Journal of the American Society for Information Science*, Vol. 47, No, 1, pp. 23-36.

Fayyad, U. and Uthurusamy, R. (1996), “Data mining and knowledge discovery in databases”, *Communications of the ACM 39*, pp. 24-37.

Fayyad, U.M., Piatetsky-Shapiro, G., and Smyth, P. (1996), “From data mining to knowledge discovery: An overview”, in *Advances in Knowledge Discovery and Data Mining. Calif.: AAAI Press*, pp. 1-34.

Fellows, R. and Liu, A. (2003), *Research methods for construction*, Blackwell Publishing.

Ford, D.N. and Sterman, J.D. (1998), “Expert knowledge elicitation to improve mental and formal models”, *System Dynamics Review*, Vol. 14, No. 4, pp. 309-330.

Frawley, W., Piatetsky-Shapiro, G., and Matheus, C. (1992), “Knowledge discovery in databases: an overview”, *AI Magazine*, pp. 213-228.

Friedman, C.P., Elstein, A.S., and Wolf, F.M. (1999), “Enhancement of clinicians’ diagnostic reasoning by computer-based consultation: a multisite study of 2 systems”, *Journal of the American Medical Informatics Association*, Vol. 282, No. 19, pp. 1851-1855.

Galland, L. (1997), *The four pillars of healing*, New York: Random House.

García, M.N.M., Román, I.R., Peñalvo, F.J.G., and Bonilla, M.T. (2008), “An association rule mining method for estimating the impact of project management policies on software quality, development time and effort”, *Expert Systems with Applications*, Vol. 34, No. 1, pp. 522-529.

Garg, A.X., Adhikari, N.K.J., McDonald, H., Rosas-Arellano, M.P., Devereaux, P.J., Beyene, J., Sam, J., and Haynes, R.B. (2005), "Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: A systematic review", *Journal of the American Medical Association*, Vol. 293, No. 10, pp. 1223-1238.

Gebus, S. and Leiviska, K. (2009), "Knowledge acquisition for decision support systems on an electronic assembly line", *Expert Systems with Applications*, Vol. 36, No. 1, pp. 93-101.

Gill, P., Scrivener, G., Lloyd, D., and Dowell, T. (1995), "The effect of patient ethnicity on prescribing rates", *Health Trends*, Vol. 4, pp. 111-114.

Golder, S.A. and Huberman, B.A. (2006), "Usage patterns of collaborative tagging systems", *Journal of Information Science*, Vol. 32, No. 2, pp. 198-208.

Gonzalez-Gonzalez, A.I., Dawes, M., Sanchez-Mateos, J., Riesgo-Fuertes, R., Escortell-Mayor, E., Sanz-Cuesta, T.M. and Hernandez-Fernandez, T. (2007), "Information needs and information-seeking behavior of primary care physicians", *The Annals of Family Medicine*, Vol. 5, No. 4, pp. 345-352.

Greenhalgh, T. and Gill, P. (1997), "Pressure to prescribe", *British Medical Journal*, Vol. 315, pp. 1482-1483.

Grefshelm, S.F. (2007), "Information needs and information seeking in a biomedical research setting: a study of scientists and science administrators", *Journal of the Medical Library Association*, Vol. 95, No. 4, pp. 426-434.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I.H. (2009), "The WEKA data mining software: an update", *SIGKDD Explorations*, Vol. 11, No. 1, pp. 10-18.

Hammond, K.J. (1989), *Case-based planning*, Academic Press Inc.

Hammond, K.W., Prather, R.J., Date, V.V., and King, C.A. (1990), "A provider-interactive medical record system can favorably influence costs and quality of medical care", *Computers in Biology and Medicine*, Vol. 20, No. 4, pp. 267-279.

Han, I., Chandler, J.S. and Liang, T.P. (1996), "The impact of measurement scale and correlation structure on classification performance of inductive learning and statistical methods", *Expert System with Applications*, Vol. 10, No. 2, pp. 209–221.

Hartge, F., Wetter, T., and Haefeli, W. E. (2006), "A similarity measure for case based reasoning modeling with temporal abstraction based on cross-correlation", *Computer Methods and Programs in Biomedicine*, Vol. 81, No. 1, pp. 41-48.

Haux, R. (2006), "Health information systems - past, present, future", *International Journal of Medical Informatics*, Vol. 75, No. 3-4, pp. 268-281.

Hearst, M. (1999), "Untangling text data mining", in the *Proceedings ACL'99: the 37th annual meeting of the association for computational linguistics*, University of Maryland, June 20–26, 1999.

Heckerman, D. (1995), *A tutorial on learning Bayesian networks*, Technical Report MSR-TR-95-06, Microsoft Research.

Herschel R.T., Nemati, H., and Steiger, D. (2001), "Tacit to explicit knowledge conversion: knowledge exchange protocols", *Journal of Knowledge Management,* Vol. 5, No. 1, pp. 107-116.

Hersh, W. (2009), "A stimulus to define informatics and health information technology", *BMC Medical Informatics and Decision Making*, Vol. 9, pp. 24.

Herzig, S., Jakobs, K.H., and Michel, M.C. (2002), "Novel teaching techniques in pharmacology", *Naunyn-Schmiedeberg's Archives of Pharmacology*, Vol. 366, No. 1.

Holbrook, A., Keshavjee, K., Troyan, S., Pray, M., Ford, P.T., and Compete, I. (2003), "Applying methodology to electronic medical record selection", *International Journal of Medical Informatics*, Vol. 71, No. 1, pp. 43-50.

Huang, M.J., Chen, M.Y., and Lee, S.C. (2007), "Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis", *Expert Systems with Applications*, Vol. 32, No. 3, pp. 856-867.

Hudlicka, E. (1996), "Requirements elicitation with indirect knowledge elicitation techniques: comparison of three methods", in *Proceedings of the 2nd International Conference on Requirements Engineering 1996,* pp. 4-11.

Hung, S.Y. and Liang, T.P. (2001), "Effect of computer self-efficacy on the use of executive support systems", *Industrial Management & Data Systems*, Vol. No. 5, pp.227-237.

Hung, S.Y., Yen, D., and Wang, H.Y. (2006), "Applying data mining to telecom churn management", *Expert Systems with Applications*, Vol. 31, No. 3, pp. 515-524.

Iakovidis, I. (1998), “Towards personal health record: current situation, obstacles and trends in implementation of electronic healthcare records in Europe”, *International Journal of Medical Informatics*, Vol. 52, No. 1-3, pp. 105-115.

Institute of Medicine. (2006), *Preventing Medication Errors: Quality Chasm Series*.

Isa, D., Kallimani, V.P., and Lee, L.H. (2009), “Using the self organizing map for clustering of text documents”, *Expert Systems with Applications*, Vol. 36, No. 5, pp. 9584-9591.

Jabr, N.H. (2007), “Physicians' attitudes towards knowledge transfer and sharing”, *Competitiveness Review: An International Business Journal incorporating Journal of Global Competitiveness*, Vol. 17, No. 4, pp. 248-260.

Jensen, F.V. (1996), *An introduction to Bayesian networks*, UCL Press, London, England.

Jensen, F.V. (2001), *Bayesian networks and decision graphs*, Springer-Verlag, New York.

Jiang, X.R. and Gruenwald, L. (2004), “Microarray gene expression data association rules mining based on BSC-tree and FIS-tree”, *Data & Knowledge Engineering*, Vol. 53, No. 1, pp. 3-29.

Kaplan, B. (2001), “Evaluating informatics applications—clinical decision support systems literature review”, *International Journal of Medical Informatics*, Vol. 64, No. 1, pp. 15-37.

Kaplan, B., Morelli, R., and Goethe, J. (1997), “Preliminary findings from an evaluation of the acceptability of an expert system in psychiatry”, *Extended Proceedings (CD-ROM) AMIA Symp*.

Kaur, H. and Wasan, S.K. (2010), “An integrated approach in medical decision-making for eliciting knowledge”, *Web-Based Applications in Healthcare and Biomedicine*, Vol. 7, pp. 215-227.

Kawamoto, K., Lobach, D.F., Willard, H.F., and Ginsburg, G.S. (2009), “A national clinical decision support infrastructure to enable the widespread and consistent practice of genomic and personalized medicine”, *BMC Medical Informatics and Decision Making*, Vol. 9, pp. 17.

Keshavjee, K., Bosomworth, J., Copen, J., Lai, J., Kucukyazici, B., Lilani, R., and Holbrook, A.M. (2006), “Best practices in EMR implementation: a systematic review”, in *Proceedings of the 11th International Symposium on Health Information Management Research (iSHIMR 2006)*, pp. 233-246.

Khan, A.S. and Hoffmann, A. (2003), “Building a case based diet recommendation system without a knowledge engineer”, *Artificial Intelligence in Medicine,* Vol. 27, No. 2, pp. 155–179.

Kirkwood, C.A., Andrews, B. and Mowforth, P. (1998), “Automatic detection of gait events: a case study using inductive learning techniques”, *Journal of Biomedical Engineering*, Vol. 11, pp. 511–516.

Klein, G.A., Calderwood, R., and MacGregor, D. (1989), “Critical decision method for eliciting knowledge”, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 19, No. 3, pp. 462-472.

Kohane, I.S., Greenspun, P., Fackler, J., Cimino, C. and Szolovits, P. (1996), “Building national electronic medical record systems via the World Wide Web”, *Journal of the American Medical Informatics Association*, Vol. 3, No. 3, pp. 191-207.

Kohn, L.T., Corrigan, J.M., and Donaldson, M.S. (2000), T*o err is human: Building a safer health system*, Washington, DC: National Academy Press.

Koppel, R., Metlay, J.P., Cohen, A., Abaluck, B., Localio, A.R., Kimmel, S.E., and Strom, B.L. (2005), “Role of computerized physician order entry systems in facilitating medication errors”, *The Journal of the American Medical Association*, Vol. 293, No. 10, pp. 1197-1203.

Koton, P. (1988), “Reasoning about evidence in causal explanations”, *Case-based Reasoning*, pp. 260-270.

Kotsiantis, S. and Kanellopoulos, D. (2006), “Association rules mining: A recent overview”, *GESTS International Transactions on Computer Science and Engineering*, Vol. 32, No. 1, pp. 71-82.

Kozier, B.J., Erb, G., Berman, A.J., and Snyder, S. (2004), *Fundamentals of nursing: concepts*, Process, and Practice, Upper Saddle River, NJ: Prentice Hall Health.

Krivda, C.D. (1995), *Data-Mining Dynamine*, Byte.

Ku, Y., Chiu, C., Liou, B.H., Liou, J.H., and Wu, J.Y. (2008), “Applying text mining to assist people who inquire HIV/AIDS information from Internet”, in *Proceedings ISI 2008 Workshops*, pp. 440-448.

Kumar, K.A., Singh, Y., and Sanyal, S. (2009), "Hybrid approach using case-based reasoning and rule-based reasoning for domain independent clinical decision support in ICU", *Expert Systems with Applications,* Vol. 36, No. 1, pp. 65-71.

Kuo, R.J., Lin, S.Y., and Shih, C.W. (2006), "Mining association rules through integration of clustering analysis and ant colony system for health insurance database in Taiwan", *Expert Systems with Applications,* Vol. 33, No. 3, pp. 794-808.

Kuo, R.J., Lin, S.Y., and Shih, C.W. (2007), "Mining association rules through integration of clustering analysis and ant colony system for health insurance database in Taiwan", *Expert Systems with Applications,* Vol. 33, pp. 794-808.

Kuperman, G.J., Teich, J.M., and Tanasijevic, M.J. (1995), "Improving response to critical laboratory results with automation: results of a randomized controlled trials", *Journal of the American Medical Informatics Association*, Vol. 6, No. 6, pp. 512-522.

Lai, J.K., Lau, F., and Shaw, N. (2009), "A study of information technology use and implementation of electronic medical record systems in BC medical practices", *BC Medical Journal*, Vol. 51, No. 3, pp. 114-121.

Lan, M., Sung, S.Y., Low, H.B., and Tan, C.L. (2005), "A comparative study on term weighting schemes for text categorization", in *Proceedings of International Joint Conference on Neural Networks (IJCNN-05),* Vol. 1, pp. 546–551.

Lau R. (2003), "Context-sensitive text mining and belief revision for intelligent information retrieval on the web", *Web Intelligence and Agent Systems An International Journal*, Vol.

1, pp. 1–22.

Laurence D.R., Bennett P.N., Laurence D.R., and Bennett P.N. (1996), *Clinical pharmacology*, Churchill Livingstone, Edinburgh.

Lee, C.H., Kim, Y.H., and Rhee, P.K. (2001), “Web personalization expert with combining collaborative filtering and association rule mining technique”, *Expert Systems with Applications*, Vol. 21, No. 3, pp. 131-137.

Lee, F., Teuch, J.M., Spurr, C.D., and Bates, D.W. (1996), “Implementation of physician order entry: User satisfaction and self-reported usage patterns”, *Journal of the American Medical Informatics Association*, Vol. 3, pp. 42-55.

Lee, S.M. and Abbott, P.A. (2003), “Bayesian networks for knowledge discovery in large datasets: basics for nurse researchers”, *Journal of Biomedical Informatics*, Vol. 36, No. 4, pp. 389-399.

Leung, G.M., Yu, L.H., Wong, O.L., Johnston, J.M., and Tin, Y.K. (2003), “Incentives and barriers that influence clinical computerization in Hong Kong: a population-based physician survey”, *The Journal of American Medical Informatics Association*, Vol. 10, No. 2, pp. 201-212.

Lewis, D.D. (1998), “Naive Bayes at forty: the independence assumption in information retrieval”, in *Proceedings of the 10th European conference on machine learning*, pp. 4–15.

Li, L., Tang, H., Wu, Z., Gong, J., Gruidl, M., Zou, J., Tockman, M., and Clark, R. (2004), "Data mining techniques for cancer detection using serum proteomic profiling", *Artificial Intelligence in Medicine*, Vol. 32, No. 2, pp. 71-83.

Lian, J., Cotrutz, C., and Xing, L. (2003), "Therapeutic treatment plan optimization with probability density-based dose prescription", *Medical Physics*, Vol. 30, No. 4, pp. 655-666.

Linnarsson, R. (1993), "Decision support for drug prescription integrated with computer-based patient records in primary care", *Medical Information,* Vol. 18, No. 2, pp. 131-142.

Ludwick, D.A. and Doucette, J. (2009), "Primary care physicians' experience with electronic medical records: barriers to implementation in a fee-for-service environment", *International Journal of Telemedicine and Applications*, Vol. 2009, pp. 1-9.

Lundin, D. (2000), "Moral hazard in physician prescription behavior", *Journal of Health Economics*, Vol. 19, No. 5, pp. 639-662.

Majkowskl, K. and Zirker, S.K. (2002), *The Challenges of Today's Prescribing Process*, RxHub LLC.

Manning, C.D., Raghavan, P., and Schütze, H. (2008), *Introduction to information retrieval*, Cambridge University Press, U.K.

Marling, C.R. and Whitehouse, P.J. (2001), "Case-based reasoning in the care of Alzheimer's disease patients", In D. W. Aha, & I. Watson (Eds), *Proceedings of the 4th international conference on case-based reasoning*, ICCBR-01, Springer, Berlin, pp. 702–715.

Matsumura, Y., Kuwata, S., Yamamoto, Y., Izumi, K., Okada, Y., Hazumi, M., Yoshimoto, S., Mineno, T., Nagahama, M., Fujii, A., and Takeda, H. (2007), "Template-based data entry for general description in medical records and data transfer to data warehouse for analysis", *Studies in Health Technology and Informatics*, Vol. 129, pp. 412-416.

Maxwell, S. and Walley, T. (2003), "Teaching safe and effective prescribing in UK medical schools: a core curriculum for tomorrow's doctors", *British Journal of Clinical Pharmacology*, Vol. 55, No. 6, pp. 496-503.

McCallum, A. and Nigam, K. (1998), "A comparison of event models for naive bayes text classification", in *Proceedings of AAAI-98 Workshop Learning for Text Categorization*.

McInnes, D.K., Saltman, D.C., and Kidd, M.R. (2006), "General practitioners' use of computers for prescribing and electronic health records: results from a national survey", *The Medical Journal of Australia*, Vol. 185, No. 2, pp. 88-91.

McKee, T.E. (1995), "Predicting bankruptcy via induction", *Journal of Information Technology*, Vol. 10, pp. 26-36.

Megalooikonomou, V., Ford, J., Shen, L., Makedon, F., and Saykin, A. (2000), "Data mining in brain imaging", *Statistical Methods in Medical Research*, Vol. 9, No. 4, pp. 359-394.

Meltzer, D., Manning, W.G., Morrison, J., Shah, M.N., Jin, L., Guth, T. and Levinson, W. (2002), "Effects of physician experience on costs and outcomes on an academic general medicine service: results of a trial of hospitalists", *Annals of Internal Medicine*, Vol. 137, No. 11, pp. 866–874.

Miller, R.H. and Sim, I. (2004), "Physicians' use of electronic medical records: barriers and solutions", *Health Affairs*, Vol. 23, No. 2, pp. 116-126.

Mitchell, T.M. (1997), *Machine learning*. Massachusetts: McGraw-Hill.

Mitra, S. and Acharya, T. (2003), *Data mining: multimedia, soft computing and bioinformatics*, John Wiley & Sons, Inc., New Jersey.

Mohd, H. and Mohamad, S.M.S. (2005), "Acceptance model of electronic medical record", *Journal of Advancing Information and Management Studies,* Vol. 2, No. 1, pp. 75-92.

Monane, M., Matthias, D.M., Nagle, B.A., and Kelly, M.A. (1998), "Improving prescribing patterns for the elderly through an online drug utilization review intervention: a system linking the physician, pharmacist, and computer", Jo*urnal of the American Medical Informatics Association*, Vol. 280, pp. 1249-1252.

Montani, S., Magni, P., Bellazzi, R., Larizza, C., Roudsari, A.V., and Carson, E.R. (2003), "Integrating model-based decision support in a multi-modal reasoning system for managing type 1 diabetic patients", *Artificial Intelligence in Medicine*, Vol. 29, No. 1-2, pp. 131-151.

Naoum, S.G. (1998), *Dissertation research and writing for construction students*, Burlington, Elsevier Butterworth-Heinemann.

Naylor, C.D. (2002), "Putting evidence into practice", *The American Journal of Medicine*, Vol. 113, No. 2, pp. 161-163.

Nelson, K.M. and Talbert, R.L. (1996), "Drug-related hospital admissions", *Pharmacotherapy,* Vol. 16, pp. 701–707.

Oliven, A., Michalake, I., Zalman, D., Dorman, E., Yeshurun, D., and Odeh, M. (2005), “Prevention of prescription errors by computerized, on-line surveillance of drug order entry”, *International Journal of Medical Informatics*, Vol. 74, pp. 377-386.

Olston, C. and Najork, M. (2010), *Web crawling*, Now Publishers Inc., Hanover, M.A., U.S.A.

Oren, E., Shaffer, F., and Guglielmo, B. (2003), “Impact of emerging technologies on medication errors and adverse drug events”, *American Journal of Health-System Pharmacy*, Vol. 60, pp. 1447-1458.

O'Sullivan, D., McLoughlin, E., Bertolotto, M., and Wilson, D. C. (2007), “Mobile case-based decision support for intelligent patient knowledge management”, *Health Informatics Journal*, Vol. 13, No. 3, pp. 179-193.

Pantazi, S.V., Arocha, J.F., and Moehr, J.R. (2004), “Case-based medical informatics”, *BMC Medical Informatics and Decision Making*, Vol. 4, pp. 4.

Park, Y.J., Choi, E., and Park, S.H. (2009), “Two-step filtering datamining method integrating case-based reasoning and rule induction”, *Expert Systems with Applications*, Vol. 36, No. 1, pp. 861-871.

Patwardhan, S. and Pedersen, T. (2006), *Using WordNet-based context vectors to estimate the semantic relatedness of concepts*, National Science Foundation Faculty, U.S.A.

Payne, P.R.O., Mendonca, E.A., Johnson, S.B., and Starren, J.B. (2007), “Conceptual knowledge acquisition in biomedicine: a methodological review”, *Journal of Biomedical*

Informatics, Vol. 40, No. 1, pp. 582-602.

Philips, W.R. and Haynes, D.G. (2001), "The domain of family practice: scope, role, and function", *Family Medicine Journal*, Vol. 33, No. 4, pp. 273-277.

Philips-Wren, G., Sharkey, P., and Morss S. (2008), "Mining lung cancer patient data to assess healthcare resource utilization", *Expert Systems with Applications: An International Journal*, Vol. 35, No. 4, pp. 1611-1619.

Pizzi, N.J. (2009), "Information processing in biomedical applications", *Human-Centric Information Processing Through Granular Modelling*, Vol. 182, pp. 289-311.

Poole, D., Mackworth, A., and Goebel, R. (1997), *Computational intelligence: a logical approach*, New York: Oxford University Press.

Prather, J.C., Lobach, D.F., Goodwin, L.K., Hales, J.W., Hage, M.L., and Hammond, W.E. (1997), "Medical data mining: knowledge discovery in a clinical data warehouse", in *Proceedings AMIA Annual Fall Symposium*, pp. 101-105.

Quaglini, S., Bellazzi, R., Berzuini, C., Stefanelli, M., and Barosi, G. (1992), "Hybrid knowledge-based systems for therapy planning", *Artificial Intelligence in Medicine*, Vol. 4, No. 3, pp. 207-226.

Radev, D.R., Jing, H.Y., and Tam, D. (2004), "Centroid-based summarization of multiple documents", *Information Processing and Management*, Vol. 40, No. 6, pp. 919-938.

Raschke, R.A., Gollihare, B., and Wunderlich, T.A. (1998), "A computer alert system to prevent injury from adverse drug events: development and evaluation in a community teaching hospital", *Journal of the American Medical Informatics Association*, Vol. 280, No. 15, pp. 1317-1320.

Rector, A.L., Nowlan, W.A., Kay, S., Goble, C.A., and Howkins, T.J. (1993), "A framework for modelling the electronic medical record", *Methods Informatics Medical,* Vol. 32, No. 2, pp. 109-119.

Richesson, R. L., and Krischer, J., Data standards in clinical research: gaps, overlaps, challenges and future directions. *The Journal of American Medical Informatics Association* , Vol. 4, No. 6, pp. 687-696.

Riou, C., Pouliquen, B., and Le Beux, P. (1999), "A computer-assisted drug prescription system: the model and its implementation in the ATM knowledge base", *Methods Informatics Medical*, Vol. 38, No. 1, pp. 25-30.

Rose, A.F., Schnipper, J.L., Park, E.R., Poon, E.G., Li, Q., and Middleton, B. (2005), "Using qualitative studies to improve the usability of an EMR", *Journal of Biomedical Informatics,* Vol. 38, No. 1, pp. 51–60.

Rosenbloom, S.T., Geissbuhler, A.J., and Dupont, W.D. (2005), "Effect of CPOE user interface design on user-initiated access to educational and patient information during clinical care", *Journal of the American Medical Informatics Association*, Vol. 12, No. 4, pp. 458-473.

Rossille, D., Laurent, J.F., and Burgun, A. (2005, "Modeling a decision-support system for oncology using rule-based and case-based reasoning methodologies", *International*

Journal of Medical Informatics, Vol. 74, No. 2-4, pp. 299–306.

Ruiz, J.G. and Hagenlocker, B. (2006), "E-prescribing", *Virtual Mentor: Ethics Journal of the American Medical Association*, Vol. 8, No. 6, pp. 381-386.

Salton, G. (1989), *Automatic text processing: the transformation, analysis, and retrieval of information by computer*, Addison-Wesley, Reading, M.A.

Sarbadhikari, S.N. (1995), "Medical Informatics – Are the doctors ready?", *Journal of the Indian Medical Association*, Vol. 93, pp. 165–166.

Sarbadhikari, S.N. and Pal, S.K. (2002), "Automated Techniques for Identifying Depression from EEG", In, Leondes CT, Ed, *Handbook of Computational Methods in Biomaterials, Biotechnology and Biomedical Systems*, Kluwer Academic Publishers, Vol. 4, Chapter 3, pp. 51–81.

Schiff, G.D. and Rucker, T.D. (1998), "Computerized prescribing: building the electronic infrastructure for better medication usage", *Journal of the American Medical Association*, Vol. 279, No. 13, pp. 1024-1029.

Schmidt, R., Montani, S., Bellazzi, R., Portinale, L., and Gierl, L. (2001), "Cased-based reasoning for medical knowledge-based systems", *International Journal of Medical Informatics*, Vol. 64, No. 2-3, pp. 355-367.

Sequist, T.D., Cullen, T., Hays, H., Taualii, M.M., Simon, S.R., and Bates, D.W. (2007), "Implementation and use of an electronic health record within the Indian health service", *Journal of the American Medical Informatics Association*, Vol. 14, No. 2, pp. 191-197.

Shan, Y., Jeacocke, D., Murray, D.W., and Sutinen, A. (2008), "Mining medical specialist billing patterns for health service management", In J. F. Roddick, J. Li, P. Christen, & P. Kennedy (Eds.), *Proceedings of the Conferences in Research and Practice in Information Technology*.

Shaw, M. and Gentry, J. (1998), "Using and expert system with inductive learning to evaluate business loans", *Financial Management*, Vo. 17, No. 3, pp. 45–56.

Shepherd M. (2007), "Challenges in health informatics", in *Proceedings of the 40th Hawaii International Conference on System Sciences 2007*.

Shi, W. and Branden, J.A. (2005), "How to combine CBR and RBR for diagnosing multiple medical disorder cases", in *Proceedings of the 6th international conference on case-based reasoning*, Chicago, IL, USA (pp. 477–491).

Shiffman, R.N., Brandt, C.A., Liaw, Y., and Corb, G.J. (1999), "A design model for computer-based guideline implementation based on information management services", *Journal of the American Medical Informatics Association*, Vol. 6, No. 2, pp. 99-103.

Shortliffe, E.H., Fagan, L.M., Wiederhold, G., and Perreault, L.E. (2000), *Medical informatics: Computer applications in health care and biomedicine*, Springer Verlag.

Sim, I., Gorman, P., Greenes, R.A., Haynes, R.B., Kaplan, B., Lehmann, H., and Tang, P.C. (2001), "Clinical decision support systems for the practice of evidence-based medicine", *Journal of the American Medical Informatics Association*, Vol. 8, No. 6, pp. 527-534.

Sitting, D.F. and Stead, W.W. (1994), "Computer-based physician order entry: The state of the art", *Journal of the American Medical Informatics Association*, Vol. 1, pp. 108-123.

Smith, A.J. and Tasioulas, T. (2002), "Education on prescribing can be improved", *British Medical Journal*, Vol. 325, No. 7367, pp. 776.

Spenceley S.E., Warren J.R., Mudali S.K., and Kirkwood I. (1997), "Intelligent data entry for physicians by machine learning of an anticipative task model", in *Proceedings of the Fourth Australian and New Zealand Intelligent Information System Conference (ANZIIS-96)*, pp. 64–67.

Stanton, L.A., Peterson, G.M., Rumble, R.H., Cooper, G.M., and Polack, A.E. (1994), "Drug-related admissions to an Australian hospital", *Journal of Clinical Pharmacy and Therapeutics*, Vol. 19, pp. 341–347.

Sternberg, R.J. and Horvath, J.A. (1999), *Tacit knowledge in professional practice*, Erlbaum, Mahwah, NJ.

Su, C.T., Yang, C.H., Hsu, K.H., and Chiu, W.K. (2006), "Data mining for the diagnosis of type II diabetes from three-dimensional body surface anthropometrical scanning data", *Computers & Mathematics with Applications*, Vol. 51, No. 6-7, pp. 1075-1092.

Sun, Z. and Finnie, G.R. (2004), *Intelligent techniques in E-commerce: a case based reasoning perspective*, New York: Springer-Verlag.

Susan, E.G. and Warren, J.M. (2000), "Statistical modelling of general practice medicine for computer assisted data entry in electronic medical record systems", *International Journal of Medical Informatics*, Vol. 57, No. 2-3, pp. 77-89.

Tague-Sutcliffe, J.M. (1996), “Some perspectives on the evaluation of information retrieval systems”, *Journal of the American Society for Information Science,* Vol. 47, No. 1, pp. 1-3.

Tam, K.W.T., Kwok, K.H., Fan, Y.M.C., Tsui, K.B., Ng, K.K., Ho, K.Y.A., Lau, K.T., Chan, Y.C., Tse, C.W.C., and Lau, C.M. (2008), “Detection and prevention of medication misadventures in general practice”, *International Journal for Quality in Health Care*, Vol. 20, No. 3, pp. 192–199.

The National Alliance for Health Information Technology (2008), *Report to the office of the national coordinator for health information technology on defining key health information technology terms.*

Thompson, M.L. (1997), “Characteristics of information resources preferred primary care physician”, *Bulletin of the Medical Library Association*, Vol. 85, No. 2, pp. 187-192.

Thornett, A.M. (2001), “Computer decision support systems in general practice”, *International Journal of Information Management,* Vol. 21, No. 1, pp. 39-47.

Toussi, M., Lamy, J.B., Toumelin, P.L., and Venot A. (2009), “Using data mining techniques to explore physicians therapeutic decisions when clinical guidelines do not provide recommendations: methods and example for type 2 diabetes”, *BMC Medical Informatics and Decision Making*, Vol. 9, pp. 28.

Tsipouras, M.G., Exarchos, T.P., Fotiadis, D.I., Kotsia, A.P., Vakalis, K.V., Naka, K.K., and Michalis, L.K., (2008), “Automated diagnosis of coronary artery disease based on data mining and fuzzy modeling”, *IEEE Transactions on Information Technology in Biomedicine*, Vol. 12, No. 4, pp. 447-457.

Turban, E. and Aronson, J.E. (2001), *Decision support systems and intelligent systems*, sixth Edition (6th ed). Hong Kong: Prentice International Hall.

Turner, R. (1988), “Organizing and using schematic knowledge for medical diagnosis”, *Case-based Reasoning*, pp. 435-446.

Uzoka, F.M.E. (2009), “Fuzzy-expert system for cost benefit analysis of enterprise information systems: a framework", *International Journal on Computer Science and Engineering*, Vol. 1, No. 3, pp. 254-262.

Valdes, I., Kibbe, D.C., Tolleson, G., Kunik, M.E. and Peterson, L.A. (2004), “Barriers to proliferation of electronic medical records”, *Informatics in Primary Care*, Vol. 12, No. 1, pp. 3-9.

Van Hyfte, D., Van Der Maas, A., Tjandra-Maga, T., and De Vries Robbe, P. (2001), “A formal framework of knowledge to support rational psychoactive drug selection”, *Artificial Intelligence in Medicine*, Vol. 22, No. 3, pp. 261-275.

Wager, K.A., Lee, F.W., and White, A.W. (2001), *Life after a disastrous electronic medical record implementation one clinic’s experience*, Idea Group Publishing, pp. 154-169.

Wang, S.J., Middleton, B., Prosser, L.A., Bardon, C.G., Spurr, C.D., Carchidi, P.J., Kittler, A.F., Goldszer, R.C., Fairchild, D.G., Sussman, A.J., Kuperman, G.K., and Bates, D.W. (2003), “A cost-benefit analysis of electronic medical records in primary care”, *The*

American Journal of Medicine, Vol. 114, No. 5, pp. 397-403.

Warren, J.R., Davidovic, A., Spenceley, S., and Bolton, P. (1998), “Mediface: anticipative data entry interface for general practitioners”, in *Proceedings Computer Human Interaction Conference 1998*, pp. 192-199.

Wazana, A. (2000), “Physicians and the pharmaceutical industry: is a gift ever just a gift?”, *Journal of the American Medical Association*, Vol. 283, No. 3, 373-380.

Westberg, E.E. and Miller, R.A. (1999), “The basis for using the Internet to support the information needs of primary care”, *Journal of the American Medical Informatics Association*, Vol. 6, No. 1, pp. 6-25.

Wickramasinghe, N., Gupta, J.N.D., Sharma, S.K. (2005), *Creating knowledge-based healthcare organizations*, Hershey PA: Idea Group Pub.

Wiig, K.M. (1994), *Knowledge management, The central management focuses for intelligent-acting organization*, Arlington: Schema Press.

Wilson, A.M., Thabane, L., and Holbrook, A. (2004), “Application of data mining techniques in pharmacovigilance”, *British Journal of Clinical Pharmacology*, Vol. 57, No. 2, pp. 127-134.

Wilson, J.R. and Corlett, N. (2005), *Evaluation of human work*, CRC Press, U.S.A.

Wozniak, M. (2006), “Two-stage classifier for diagnosis of hypertension type”, in: Maglaveras, N., Chouvarda, I., Koutkias, V., Brause, R. (eds.) *ISBMDA 2006*, Vol. 4345, pp. 433–440.

Wu, H.C., Luk, Y.P., and Wong, K.F. (2008), "Interpreting TF-IDF term weights as making relevance decisions", *ACM Transactions on Information Systems,* Vol. 26, No. 3, pp. 13-37.

Xhemali, D., Hinde, C.J., and Stone, R.G. (2009), "Naive bayes vs. decision trees vs. neural networks in the classification of training web pages", *International Journal of Computer Science Issues,* Vol. 4, No. 1, pp. 16-23.

Yang, Y. and Chute, C.G. (1994), "An example-based mapping method for text categorization and retrieval", *ACM Transactions on Information Systems*, Vol. 12, No. 3, pp. 252-277.

Yin, R.K. (2009), *Case study research: design and methods,* SAGE Publications, London.

Yu, W.D. and Jonnalagadda, S.R. (2006), "Semantic web and mining in healthcare", in *Proceedings 8th International Conference on e-Health Networking, Applications and Services*, pp. 198-201.

Yusof, M.M. and Buckingham, C.D. (2009), "Medical case-based reasoning: A review of retrieving, matching and adaptation processes in recent systems", In M. H. Hamza (Eds.), *Proceedings of the Artificial Intelligence and Applications (AIA 2009),* Innsbruck, Austria.

Zhan, J.M., Loh, H.T., and Liu, Y. (2009), "Gather customer concerns from online product reviews – a text summarization approach", *Expert Systems with Applications*, Vol. 36, No. 2, pp. 2107-2115.

Zhang, J.J. (2002), "Representations of health concepts: a cognitive perspective", *Journal of*

Biomedical Informatics, Vol. 35, No. 1, pp. 17-24.

Zhou, D., He, Y., and Kwoh, C.K. (2006), “Validating text mining results on protein-protein interactions using gene expression profiles”, in *Proceedings International Conference on Biomedical and Pharmaceutical Engineering 2006*, pp. 580-585.

Zhuang, Z.Y., Churilov, L., Burstein, F., and Sikaris, K. (2009), “Combining data mining and case-based reasoning for intelligent decision support for pathology ordering by general practitioners”, *European Journal of Operational Research,* Vol. 195, No. 3, pp. 662-675.

Zrimec, T. and Wong, J. (2007), “Methods for automatic honeycombing detection in HRCT images of the lung”, in *Proceedings of the 11th Mediterranean Conference on Medical and Biomedical Engineering and Computing 2007,* pp. 830-833.

Appendix I

Questionnaire Used for System Performance Evaluation

TEMRS Evaluation Questionnaire

For the purpose of continuous improvement of the design of the Template-based Electronic Medical Record System (TEMRS), we would be grateful if you could spend a few minutes to complete this questionnaire. The information you provided would only for the above purpose, so it is highly confidential.

Date:_________________

Position: General Practitioner

Part I Frequency of Use

	Frequency				
	Never	Seldom	Sometimes	Often	Frequently
Patient Information Retrieval	□	□	□	□	□

Part II System User-friendliness

	Level of Satisfaction				
	Very Dissatisfied	Dissatisfied	Normal	Satisfied	Highly Satisfied
System Operational Design					
Variety of Attributes	□	□	□	□	□
Simplicity	□	□	□	□	□

Ease to learn	□	□	□	□	□

Part III System Maintenance

	Level of Satisfaction				
	Very Dissatisfied	Dissatisfied	Normal	Satisfied	Highly Satisfied
Variety of Attributes	□	□	□	□	□
Simplicity	□	□	□	□	□
Ease to learn	□	□	□	□	□

Part IV System Performance

	Level of Satisfaction				
	Very Dissatisfied	Dissatisfied	Normal	Satisfied	Highly Satisfied
1. Quality and accuracy					
Information Retrieved	□	□	□	□	□
Knowledge Representation Scheme	□	□	□	□	□
Template being designed	□	□	□	□	□
2. Sufficiency of content					
Information Retrieved	□	□	□	□	□
Knowledge Representation Scheme	□	□	□	□	□

Template being designed	□	□	□	□	□

3. Ease to understanding

Information Retrieved	□	□	□	□	□
Knowledge Representation Scheme	□	□	□	□	□
Template being designed	□	□	□	□	□

4. Effectiveness

Information Retrieved	□	□	□	□	□
Knowledge Representation Scheme	□	□	□	□	□
Template being designed	□	□	□	□	□

Thank you for your kind attention!