T.C.

ISTANBUL ALTINBAS UNIVERSITESI

GRADUATE SCHOOL OF SCINCES ENGINEERING

Detecting and Classifying Drug Interaction using Data mining Techniques

Baraa Taha Yaseen

Master of Information Technology

Thesis Supervisor

Asst.Prof.Dr.SEFER KURNAZ

Istanbul (2018)

INTRUSION DETECTION MODEL BASED ON DATA MINING AND MACHINE LEARNING

by

Baraa Taha Yaseen

Information Technology

Submitted to the Graduate School of Science and Engineering

in partial fulfillment of the requirements for the degree of

Master of Science

ALTINBAŞ UNIVERSITY

2018

This is to certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Asst.Prof.Dr.SEFER KURNAZ

Supervisor

Examining Committee Members (first name belongs to the chairperson of the jury and the second name belongs to supervisor)

Academic Title Name SURNAME	Faculty, University	____________
Academic Title Name SURNAME	Faculty, University	____________
Academic Title Name SURNAME	Faculty, University	____________

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Academic Title Name SURNAME

Head of Department

Academic Title Name SURNAME

Director

Approval Date of Graduate School of Science and Engineering: ____/____/____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Baraa Taha Yaseen

ACKNOWLEDGEMENTS

My thanks and appreciation to my supervisor Asst.Prof.Dr.SEFER KURNAZ to his attention, support, patience to me throughout the period of this study as well as my thanks to my family, who was supported for me all the time and also to my friends who help me throughout these years.

ABSTRACT

DETECTING AND CLASSIFYING DRUGS INTERACTION USING DATA MINING TECHNIQUES

Baraa Taha Yaseen

M.S., Information Technology, Altınbaş University

Supervisor: Asst.Prof.Dr.SEFER KURNAZ

Date: 2018

Page: 57

Nowadays, there are thousands of approved drugs that can be used for treating people who have medical problems. Therefore, drug Warnings and Precautions section is intended to identify and describe a discrete set of adverse reactions and other potential safety hazards that are serious or are otherwise clinically significant because they have implications for prescribing decisions or for patient management. In this thesis, proposed tool takes the patient personal information, such as gender, age, pregnancy, the current diseases that the patient suffers from, the current drugs, and the newly diagnosed disease as input and returns the suggested drugs that suit the current patient state. The proposed tool consists of two main phases: data acquisition and data processing. In data acquisition phase, drug information has been collected from different sources including DINTO, DrugBank, RxNorm, and FDA. Moreover, a number of test cases have been designed for evaluating the performance of the proposed tool in different scenarios. The test cases have shown the efficiency with accuracy 96% and effectiveness of the proposed tool.

KEYWORDS: Data Mining, Drug Interaction, Classification.

ÖZET

VERİ MADENCİLİĞİ TEKNİKLERİNİ KULLANARAK İLAÇ ETKİLEŞİMİNİN DÜZENLENMESİ VE SINIFLANDIRILMASI

Baraa Taha Yaseen

M.S., Bilgi Teknolojisi, Altınbaş Üniversitesi

Danışman Asis.Prof Sefer KURNAZ

Tarih: 2018

Sayfa: 57

Günümüzde, tıbbi sorunları olan kişilerin tedavisinde kullanılabilecek binlerce onaylı ilaç bulunmaktadır. Bu nedenle reçetelerde bulunan , Ilaç Kullanım Uyarıları ve Önlemler bölümü, hasta yönetimine etkileri olduğu için, ciddi ya da başka bir şekilde klinik olarak önemli olan farklı istenmeyen yan etkiler ve diğer potansiyel güvenlik tehlikelerini tanımlamak ve açıklamak için tasarlanmıştır. Bu tezde, önerilen araç hastanın cinsiyet, yaş, hamilelik, hastanın yaşadığı mevcut hastalıklar, mevcut ilaçlar ve yeni teşhis edilen hastalıklar gibi hastanın kişisel bilgilerini temel olarak alır ve mevcut hastaya uygun önerilen açıklamaları, bilgileri ve ilaçları verir. Önerilen araç iki ana aşamadan oluşmaktadır: Veri toplama ve veri işleme. Veri edinme aşamasında, ilaç bilgisi DINTO, DrugBank, RxNorm ve FDA dahil olmak üzere farklı kaynaklardan toplanmıştır. Ayrıca, önerilen aracın performansını farklı senaryolarda değerlendirmek için bir dizi test çeşidi tasarlanmıştır. Test çeşitleri verimliliği % 96 doğrulukla ve önerilen aracın etkinliği ile göstermiştir.

ANAHTAR KELİMELER: Veri Madenciliği, İlaç Etkileşimi, Sınıflandırma.

Table of Contents

LIST OF TABLES

<u>Pages</u>

LIST OF FIGURES

Pages

1. INTRODUCTION

1.1 DRUG INTERACTION

The field of pharmacology has not denoted a foreigner to the conception of consciousness bases. RxNorm promotes information transfer concerning medication in clinical ways [1], regulating styles for clinical medicines and treatment patterns. SNOMED CT is the complete ontology of clinical indications for maintaining fitness information operations [2], including various classes of medications and drug-related information. MeSH[3], a controlled dictionary thesaurus for indexing Med-Line articles, includes a significant amount of pharmacological data, including pharmacological effects of drugs.

On the other hand, Adverse drug reaction (ADR) is unit of the essential conditions of mortality and morbidity in the United States, accounting for over 700,000 crisis department visits and 120,000 patients each year. It is estimated to impose an additional annual cost of $3.5 billion on the healthcare system and is a significant public health problem that can potentially be avoided [4]. Adverse drug interactions (ADIs) have been referenced as dormant causes of illness morbidity as well as raised pharmaceutical costs and malpractice claims [5].

Drug-drug interactions (DDIs) denote probable conflicting medication reactions having a significant influence on illness protection and healthcare payments [6]. DDI is a situation when individual drug affects the level or activity of another. Discovery of DDI is essential for both inmate protection and effective health care administration [7]. DDIs are grouped into five main classes: no interaction, advice, effect, mechanism, and int [8, 9].

Although there is a high number of medication databases and semi-structured sources (such as Stockley [6] and DrugBank [10]) with knowledge about DDIs, these databases are inadequate, and the proportion of their content is restricted, so it is tough to select real clinical importance to every interaction. However, although several ontologies exist for the pharmacology field, none treats all groups correlated to drug-drug interactions. In fact, few philosophies are concentrating in the description of DDI information, such as Drug Interaction Ontology (DIO) [11], PK ontology [12], and Drug Interaction Knowledge Base (DIKB) [13].

Furthermore, the essential service most DDIs knowledge bases provide, whatever the representation method they adopt, is to index drug combinations observed to associate in clinical cases or reported as such in clinician-submitted case reports. One main weakness of this strategy is that it forces drug-interaction and the instruments that employ them, to treating interacting drug pairs [14].

Additionally, ADR can occur due to the potential for medications to communicate negatively with a diversity of illnesses (drug-disease cooperation). For instances, any beta-blockers used for heart infection or high blood pressure can worsen asthma and cause it hard for bodies with diabetes to recognize when their blood sugar is too low. Any medications used to handle a cold may worsen glaucoma. However, this important issue has rarely been addressed. Also, some drugs should not be prescribed for people of certain ages or pregnant women. Therefore, patient physical information is essential during the drugs prescription process [5].

To the best of our knowledge, there is no an interactive tool that can help physicians during the drugs prescription process that considers drug-drug interactions, drug-disease interaction, and patient physical information. In this thesis, we intend to cover this gap.

1.2 MOTIVATION

The primary motivations of this work can be summarized in two points. First, many drugs have serious adverse reactions that can jeopardize human lives or can cause severe medical problems. Therefore, it is very critical for physicians knowing the different warnings and precautions associated with each drug during the drugs prescription process. However, remembering the warnings and precautions of an increasing number of drugs by the physicians is challenging. Second, despite of the obvious importance of drug Warnings and Precautions in prescribing decisions and patient management, there is currently no single complete source for drug warnings and precautions.

Additionally, the success of Data Mining techniques in uncovering and inferring the hidden patterns from huge amounts of data in many fields including the medical field encourages us to use it in inferring DDIs, the serious adverse reactions, and the clinically significant reactions associated with drugs based on the drug warnings and precautions and the current state's information of the patient which include age, sex, pregnancy, current diseases, current drugs, new prescribed drugs, and so on.

1.3 PROBLEM STATEMENT

Adverse drug reactions are major sources of medical problems and consequences that increase both of risk factors and healthcare costs. Therefore, drugs should be recommended to the patients carefully. However, the availability of various drugs makes prescribing the suitable drug by physicians a very challenging task, because they should consider the patients' physical information such as age, gender, pregnancy, etc. and patient's history record such as his current diseases and drugs. Additionally, physicians should know drug warnings and precautions which very difficult task is due to the large number of drugs. Another difficulty for physician is the scattering of drug information on the web among different databases that may entail incomplete drug guidance details.

1.4 RESEARCH AIMS

This thesis aims to sketch and implement an interactive semantic ontology-based (semantic ontology implemented using PHP Language with input drugs information and output specific drugs based on a query. Semantic concept will attend to increase the creation of search based on the regulation, communication and warning of drugs rather than keyword or drug names and helps in finding correct information. The proposed tool takes the patient personal information, such as gender, age, pregnancy, the current diseases that the patient suffers from, the current drugs, and the newly diagnosed disease as input and returns the suggested drugs that suit the current patient state. Additionally, the performance of the proposed tool should be analyzed using a number of test cases that can verify the validity of the proposed tool. Toward achieving these objectives, a set of steps have to be followed as shown below.

1. Reviewing and investigating the approaches that have been employed for drug information.
2. Collecting drug information from highly trustable sources to be using in the proposed system
3. Designing and implementing the proposed tool that can help physicians in prescribing drugs to their patients without jeopardizing their safety.
4. Evaluating the performance of the proposed tool.
5. Releasing the recommendations for the future work.

1.5 THESIS CONTRIBUTIONS

The principal purpose of this thesis is designing and implementing an interactive tool that takes into account the different factor that can lead to adverse drug interactions that can help physicians in drug prescription. <u>This tool can recommend the suitable drugs for the newly diagnosed disease</u>, keeping in mind the current diseases that the patient suffers from, the current drugs, and patient personal information such as age, gender, pregnancy. We passed through two phases toward building the proposed tool: data acquisition and data processing.

In the data acquisition phase, drug information is collected from a number of sources such as drug bank and RxNorm. In the data processing phase, given the information of the current patient, a hybrid of data mining techniques (Decision Tree, Apriori) are applied to get the proper drugs for the current case keeping in mind the different risk factors.

1.6 ORGANIZATION OF THESIS

We held the rest of thesis as follows: Chapter 2 contains an overview about the basic terminologies and concepts required to get better understanding for the suggested work. Chapter 3 presents some of research efforts which have been done in the research field. Chapter 4 presents a detailed description for the proposed framework. Chapter5 provides a number of test cases that have been designed to evaluate the proposed tool. Finally, Chapter 6 presents the conclusion.

2 BACKGROUND

This chapter presents brief description for all the concepts that necessary to understand the proposed work. It starts with presenting an overview on knowledge engineering. Then, ontology and ontology languages are overviewed. Then, the concept of web service is briefly described. Finally, the chapter is concluded by explain the main idea behind the apriori algorithm.

2.1 KNOWLEDGE ENGINEERING

In a few years ago, the community of formalisms, reasoning mechanisms, and mechanisms to operationalize Knowledge-based Systems (KBS) focus on Artificial Intelligence (AI). Typically, the improvement efforts were limited to the achievement of small KBSs to investigate the utility of the various methods. Although specific readings contributed somewhat hopeful issues, the transfer of this technology into profitable use to produce big KBSs failed in many cases. Similarly, as the software disaster occurred in the demonstration of the control Software Engineering the unsatisfactory situation in constructing KBSs made clear the necessity for more methodological procedures [15]. Therefore, the purpose of the further power Knowledge Engineering (KE) is applying the process of building KBSs from an art into an engineering method [15].

The discipline of KE developed out of the early work on expert systems in the seventies. With the increasing prevalence of knowledge-based systems, there also arose a demand for a methodical strategy for developing such schemes, related to methodologies in main-stream software engineering. Over the years, the regulation of knowledge engineering has grown into the development of theory, techniques, and mechanisms for increasing knowledge-intensive utilization. In other words, it presents supervision about when and how to utilize appropriate awareness presentation techniques for solving specific dilemmas [15].

KE is an organization method that includes combining consciousness into computer systems to resolve complicated dilemmas usually expecting a high level of individual expertise. It often requires five separated steps in shifting human experience into some structure of knowledge-based systems (KBS), namely knowledge retrieval, knowledge validation, knowledge description, inferring, an interpretation and explanation.

In this section, we are focusing on the knowledge representation methods. But before proceeding with describing the different types of knowledge representation methods, the different types of knowledge are overviewed first.

2.1.1 Types of Knowledge

Organizing knowledge is essential to discover methods to describe it [24]. In the scope of design and engineering, the experience can generally be grouped along different dimensions. Each class has its basis [16].

- **Implicit vs. explicit**

Explicit knowledge leads to an awareness that is transmittable in symmetrical, regular language. Implicit knowledge is strongly rooted in operations, background, and association in a particular context. It consists of the cognitive factor and scientific factor.

- **Individual vs. social**

Individual knowledge is produced by and exists in the individual whereas social learning is designed by and exists in the common operations of a group.

This thesis is essentially concerned with pharmaceutical experience which can be separated into two classes: declarative information and procedural information. Declarative information involves hypotheses and decisions. Decisions are statements about the world that are either "true" or "false." These hypotheses may be related by Boolean speculators such as "and," "or," and "and not" to form sentences.

2.1.2 Knowledge Design Methods

Producing awareness and making it explicit is one of the critical roles of knowledge description. Sowa [17] represents knowledge form as a multidisciplinary case that merges techniques from logic, philosophy, and calculation. The foundations of knowledge description lie in the study of affected intelligence, but its influence has stretched to many disciplines. Bernard and Xu [18] recommend a citation structure for the variety and the characterization of awareness that refers to all types of experience. As shown in Figure 2.1, Owen and Horváth group knowledge designs into five sections: virtual, symbolic, pictorial, linguistic, and algorithmic. Figure 2.1 shows some of the many description models.

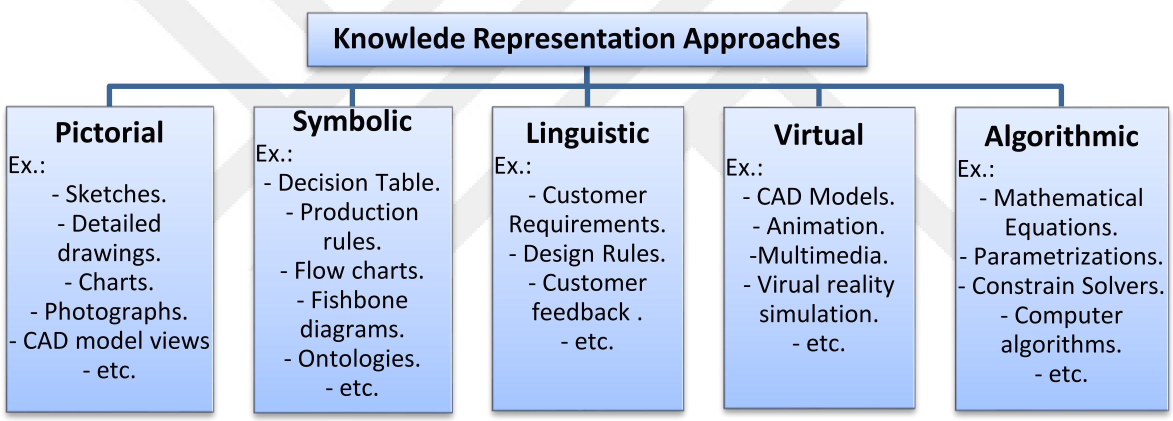


Figure 2.1: The different knowledge representation approach.

2.2 DATA MINING

Generally, there many data mining techniques that have been adopted in healthcare such as classification, clustering, association, and regression as shown in figure 1. A brief description about each one of them is provided next.

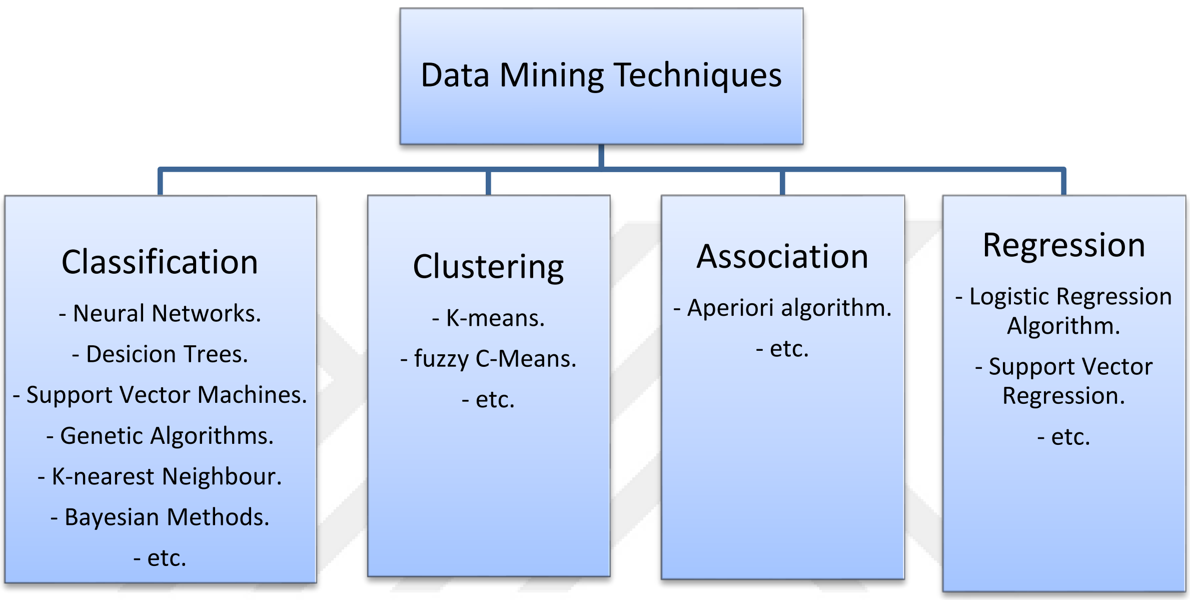


Figure 2.2: The different data mining techniques used in healthcare management.

2.2.1 Classification Techniques

Classification breaks data units into distinct groups. The categorization procedure foretells the aim class for each data points. For example, an inmate can be grouped as “great risk” or “below risk” inmate with their illness model using data organization approach. It is a supervised training procedure having identified class divisions. Binary and multilevel are the two arrangements of classification. In a binary arrangement, only two possible conditions such as, “high” or “low” risk patient may be considered while the multiclass strategy has more than two purposes for example, “large,” “moderate” and “fading” risk inmate [19].

Classification contains two footsteps. The initial step is designing structure, which is applied to explain the training records of a database. The additional step is designing method where the assembled model that used for classification. The efficiency of the categorization is assessed rendering to the rate of test units or test dataset that are properly classified [20]. There are a large group of techniques which have been used in healthcare supervision to complete the coordination process which includes: J-48, SVM, K-nearest neighbor, neural networks, Bayesian methods, etc.

2.2.2 Clustering

Clustering is the initial assignment in data-mining and a standard method for arithmetical data interpretation used in various areas, including pattern identification, image processing, data retrieval, and bioinformatics [21]. Clustering is an unsupervised training method that is distinct from categorization. Grouping is unlike to distribution as it has no predefined levels. In collecting large database are divided into the structure of small separate subgroups or clusters. Clustering partitioned the data points based on the correlation measure [19]. Clustering procedures discover obtaining the data such that objects in the equivalent cluster are more related to each other than other groups. Several clustering procedures used in healthcare administration such as K-means, Fuzzy C-Means, etc.

2.2.3 Association

Association Rule Mining developed much later than machine learning and is directed to greater control from the investigation field of databases. Although association rule mining was first presented as a market basket analysis agent, it has since enhanced one of the most important means for performing unsupervised exploratory data analysis over a comprehensive range of investigation and investment fields, including biology and bioinformatics . Usually, Association is one of the essential approaches of data mining that is used to find out the familiar patterns, new relationships among a set of data items in the data repository [19]. [20]. Apriori algorithm is one of the association algorithms that is widely adopted in healthcare control.

2.2.4 Regression

Regression is applied to get out duties that demonstrate the association between various variables. An analytical form is created using training dataset. In analytical modeling, two sets of variables called dependent variable, and an independent variable which usually represented using ‘Y’ and ‘X.’ Regression is widely used in pharmaceutical area for foretelling the diseases[19]. Many regression methods used in healthcare control.

2.3 WEB SERVICES CONCEPTS

In the proposed system, we heavily depended on web services to collect the domain knowledge. In this section, an overview is provided on the basic concepts of web services.

Earlier methods such as such as the Remote Method Invocation (RMI), Common Request Broker Architecture (CORBA), and Distributed Component Object Model (DCOM) were applied to generate consumer and server forms. These are utilized in highly linked distributed schemes, in which both server and consumer are reliant on each other. Adaptability and protection dilemmas arise while we employ these methods. This class of transactions will prevent by firewall and proxy servers. Deeply linked schemes are mostly utilized for intranet purposes because platform and technology employed are previously known or identical on both server and client side. Consequently, we are applying a web service which is a conventional way to appropriate services over the internet. The clients don’t possess any earlier experience of the web services before they apply it. Consequently, web service is program autonomous and loosely linked [22]. Like its predecessors, Web Services is a collection of models and a programming approach for participating data among various software applications. Furthermore, Web services are a standardized way to categorize services on the Internet [22]

.

Web Services can be categorized into pair principal classes: RESTful and SOAP-based Web Services. The architectural style of this organization used in the implementation process [22].

SOAP attains for Simple Object Access Protocol. It is an OOP approach that determines a conventional rule applied for transferring XML-based information. It is illustrated as protocol designation for transferring structured data in the developing of Web Services in machine interfaces. The designation describes an XML based case for moving information, and the protocol specifies a set of controls for transforming platform-specific data models into XML descriptions [22].

Representational State Transfer (REST) means a resource-oriented technique, and it signifies described by Fielding in as a structural form that includes of a collection of scheme guidelines that determine the appropriate behavior for applying web patterns such as HTTP and URIs. Although REST is basicaly described in the circumstances of the Web, it is becoming a popular implementation method for generating web services. RESTful is developed with Web models (URI, HTTP, and XML) and REST sources. REST policies include connectivity, addressability and stateless. RESTful applied to determine particular actions that operated on URL sources. Both SOAP and RESTful are utilized for performing Web Services. Nevertheless, individually has its separate characteristics and weaknesses that make it more or less fitting for several kinds of application as shown in Table 2.1 [22].

Table 2.1: Comparison between SOAP-based Web Services and RESTful-based Web Services

Criteria	**SOAP-based WS**	**RESTful-based WS**
Server/ Client	Tightly coupled	Loosely coupled
URI	One URI representing the service endpoint	URI for each resource
Transport Layer Support	All	Only HTTP
Caching	Not Supported	Supported
Interface	Non-Uniform Interface (WSDL)	Uniform Interface
Context aware	Client context aware of WS behavior	Implicit web service behavior
Data Types	Binary requires attachment parsing	Supports all data types directly
Method Information	Body Entity of HTTP	HTTP Method
Data Information	Body Entity of HTTP	HTTP URI
Describing Web Services	WSDL	WADL
Expandability	Not Expandable (No hyperlinks)	Expandable without creating new WS (using xlink)
Standards used	SOAP specific standards (WSDL, UDDI, WS-Security)	Web standards (URL, HTTP methods, XML, MIME Types)
Security/Confidentiality	WS-security standard specification	HTTP Security

2.4 AN OVERVIEW ON APRIORI ALGORITHM

Data Mining is an important research field that aims to find useful information within large amounts of data. Nowadays, Data Mining becomes popular in the healthcare field as an efficient analysis method. It is needed for detecting hidden and important patterns from medical data. In the health industry, DM techniques have many uses such as discovering the committing of fraud in health insurance, providing the medical solutions for the sick people with low cost and detecting the causes of diseases and determining the possible medical treatment methods [19].

Generally, there are many data mining approaches that have been employed in the healthcare sector, such as clustering, classification, regression and association. In the proposed work, two popular data mining approaches have been employed cooperatively for recommending the suitable drugs for a certain patient namely, apriori algorithm and decision tree algorithm. In this section provide an overview on the apriori algorithm while decision tree algorithm is covered in the following section.

In the proposed work, we have employed a well-known association algorithm, the apriori algorithm, to discover the associations between the drugs and the patient’s drug history (current diseases and the drugs that are currently taken by the patient). A brief description for this algorithm is given below.

Apriori procedure is one of the most powerful procedure to mine the recurrent element sets of Boolean connotation rules. This procedure is method based on two-phases occurrence, the strategy of connotation rule mining procedure can be decayed into two sub-matters [23]:

1) Discover all the element sets with the support larger than the lowest support, which is called recurrent item;

2) Based on recurrent elements, all the connotation guidelines will be produced, and for individually regular recurrent set A, all the nonempty subgroup a of A will be originated, if the proportion of support (A)/support(a) ≥ min-confidence, to produce the connotation directions A-a. That is, from the recurrent elements obtained in the initial step, to exploit the guidelines with confidence not fewer than lowest confidence min-confidence the user quantified.

We can describe the realization method as ensues: Initially, Apriori method determine the common collections L with elements of 1, then from L to produce the claimant collections Cz with elements of 2, examine the performance database D, determine the support to determine Lz, and so on. Once commonly collections are created from the database, the algorithm can generate powerful correlation rules directly. We can discuss the steps of the algorithm as follows: Apriori method designated couple sub-processes which are gen() and subgroup(). Gen() function returns a claimant, then utilize the Apriori resources (all non-empty subsets of common element sets must also be expected) to eliminate those applicants of the non-frequent element sets. Formerly produced all of the candidates, we will examine the database D, and for specific action, apply the Subset () procedure to recognize all the applicant subsets, and make an aggregate calculation for several of these applicants. Lastly, all applicants matched the smallest support form common item set L [23].

2.5 AN OVERVIEW ON DECISION TREES

DT (Decision Tree) is an example of the various dominant procedures in information detection and data mining [19]. DT method is a kind of data mining methods that have sources in standard analytical controls such as linear regression. DT also participate sources in the related discipline of cognitive science that designed neural networks. A DT is a classifier denoted as a recursive separation of the situation space. A DT includes of a source node, leaf nodes that indicate any classes, inner nodes that describe examination conditions as shown in Figure 2.3.

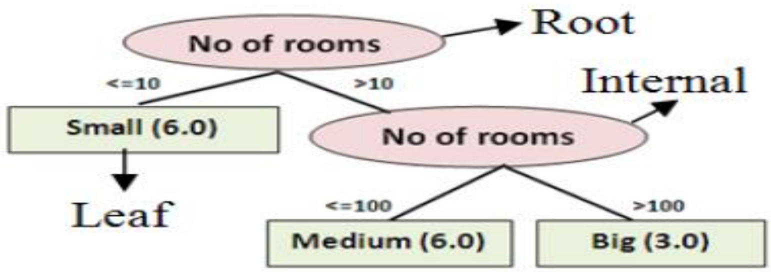


Figure 2.3: An example for decision trees

A DT recursively divides the cases into branches based on some analytical methods, such as Gini index and information gain to define the property and corresponding threshold of that attribute that separates the input instances into two or more combinations. This method is recursively executed at each leaf node until the tree structure is completed. The primary purpose of the dividing method is to concern a variable-threshold pair which maximizes the correlation in the generated subgroups. ID3, C4.5, and C5 decision tree algorithms are employing Entropy-based information gain, while CART and CHAID decision tree algorithms are using Gini index and Chi-squared test, respectively.

Decision trees have a number of advantages including being easy to understand by the end user, it can handle a variety of input data: nominal, numeric and textual, it can handle noisy datasets or missing values, it can achieve high accuracy, and it can be implemented data mining packages over a variety of platforms.

2.6 CHAPTER REVIEW

In this section, all the theories related to the proposed work have been overviewed and described briefly. The next chapter covers some research efforts which have been done in the field.

3 RELATED WORKS

3.1 INTRODUCTION

This chapter presents some of works that are relevant to the proposed work. It starts with defining biomedical ontologies. Then, an overview is provided on a set of well-known biomedical ontologies. Also, it covers the most significant drug database and drug-drug interactions. Finally, the chapter is concluded with presenting some recent approaches for extracting and classifying drug-drug interactions.

3.2 DRUG DATABASES

Due to the important effect of DDIs on patient safety and healthcare cost, as a common type of adverse reactions, huge research efforts have been performed to represent DDI information. In this section, we cover some of these efforts.

Recently there are a number of publicly available databases and semi-structured sources that contain drug information, including DDIs information, such as DrugBank [10], Stockley [6], RxNorm [1], Drug Interactions Facts [9].

DrugBank is a wide ranged online database which contains inclusive pharmacological and biochemical information about drugs, which argues their mechanisms of treatment and their objectives. It is developed, sustained and enhanced by extensive literature surveys demonstrated by domain-specific trained biocurators [10].

Stockley's Drug Interactions will continue to be the world's most imposing and inclusive international reference book on drug interactions. It discusses interactions between therapeutic drugs, herbal medicines, proprietary medicines foods, drinks, and drugs of abuse and also provides complete details of the clinical indication for interactions being discussed, an assessment of their clinical significance, and clear guidance on managing the interaction in practice [6].

RxNorm offers a vocabulary for normalized names of clinical drugs. It was originally built to cover all prescribed medications in the USA. It includes each drug's active ingredient, dose, interactions, and strengths [1].

Drug Interactions Facts help health experts by offering a quick and a more realistic communication screening instrument from covering 18,000 monographs. In just moments, possible communications can be immediately analyzed by (class, generic drug, or even trade name). It presents information on the opening, austerity, and documentation of clinically essential interactions; it also involves a discussion of their consequences, mechanism, and administration [8,9].

3.3 AN OVERVIEW ON DRUG-DRUG EXTRACTION APPROACHES

In this section, we cover some of research efforts that have been performed in the field of Drug-Drug interactions (DDIs) extraction and classification.

In [24], the authors presented a new feature-based kernel method to extract and classify drug interactions described in biomedical literature. Like many previous works, their method consists of two steps. First, they detect interacting drug pairs, and then they classify each extracted pair into one of four interaction categories. To perform the first step, they have enhanced an existing feature-based system by adding new features, correction patterns, and trigger words. Then, they used a binary classifier (LIBSVM classifier is used with RBF kernel) to detect interacting drug pairs. To perform the second step, they have built a new feature-based kernel classifier that exploit the lexical field particularity of each interaction type. This classifier is composed of 4 binary classifiers (SVM classifiers with radial basis function (RBF) and linear kernel) work sequentially. When evaluated on the DDIExtraction 2013 challenge corpus, their system achieved an F1- score of 71.14%.

In [25], the writers investigated the aggregate of pairing various machine-learning techniques to obtain DDI: (i) a feature-based approach adopting an SVM with a collection of attributes derived from texts, and (ii) a kernel-based approach mixing three different kernels. Investigations attended on the DDIExtraction2011 challenge corpus show that our method is helpful in selecting DDIs with 0.6398 F1 scores.

In [26], the writers performed a language convolutional neural network (SCNN) based Drug Interaction selection process. In scheme, a new expression inserting, arrangement expression surrounding, is planned to manipulate the syntactic knowledge of a decision. Then the status and part of conversation attributes are added to increase the inserting of the specific expression. Later, auto-encoder is intended to encode the common bag-of-words function as the dense real value vector. Lastly, a mixture of inserting-based convolutional attributes and common characteristics are supplied to the softmax classifier to obtain DDIs from biomedical literature. Experimental results on the DDIExtraction 2013 corpus show that SCNN captures a better achievement (an F-score of 0.686) than other state-of-the-art approaches.

In [27], the writers included a scheme promoted to select DDI for drug specifying combinations observed in biomedical documents. This approach relies massively on deep syntactic parsing to represent the relationships among drug remarks. In explaining the DDI extraction operation, they assessed the compatibility of both text-based and database obtained characteristics for DDI discovery. For machine learning, they examined both SVM and RLS approaches, with particular investigations for defining the optimal factors and training strategy. Their scheme has produced an achievement of 62.99% F-score on the DDI Extraction 2011 task.

In [28], the producers formed a corpus of Federal medicine Administration recommended medicine container supplement records that have been manually interpreted for pharmacokinetic DDIs by a pharmacologist and a medicine information specialist. Then, they estimated three various machine learning algorithms (SVM, and J48) for their experience to 1) recognize pharmacokinetic DDIs in the package insert corpus and 2) analyze pharmacokinetic DDI records by their modality (i.e., whether they report a DDI or no interaction between medication pairs). Investigations determined that an SVM algorithm implemented sufficient on both assignments with an F-measure of 0.859 for pharmacokinetic DDI description.

Table 3.1: Summary of some methodologies relevant to the proposed work.

Ref	Methodology	Technique	Accuracy
[24]	classify drug interactions described in biomedical literature	SVM classifiers with radial basis function (RBF) and linear kernel	71 %.
[25]	Extract drug interaction between pair of drugs and classify related drugs	SVM classifier	63%
[26]	Encode the traditional bag-of-words feature to extract DDIs from biomedical literature	syntax convolutional neural network (SCNN)	68%
[27]	Binary classification of drug pairs as interacting or non-interacting, and the second stage for further classification of interacting pairs into one of four interacting types	both stages were Support Vector Machines	64%
[28]	system developed to extract drug-drug interactions (DDI) for drug mention pairs found in biomedical texts	decision tree (J48) + SVM	85%

3.4 CHAPTER REVIEW

This chapter has introduced an overview on some research efforts that have been done in the fields that are relevant to the proposed work. In the next chapter, we provide an interactive tool for representing the various aspects of drug interactions.

4 METHODOLOGY FRAMEWORK

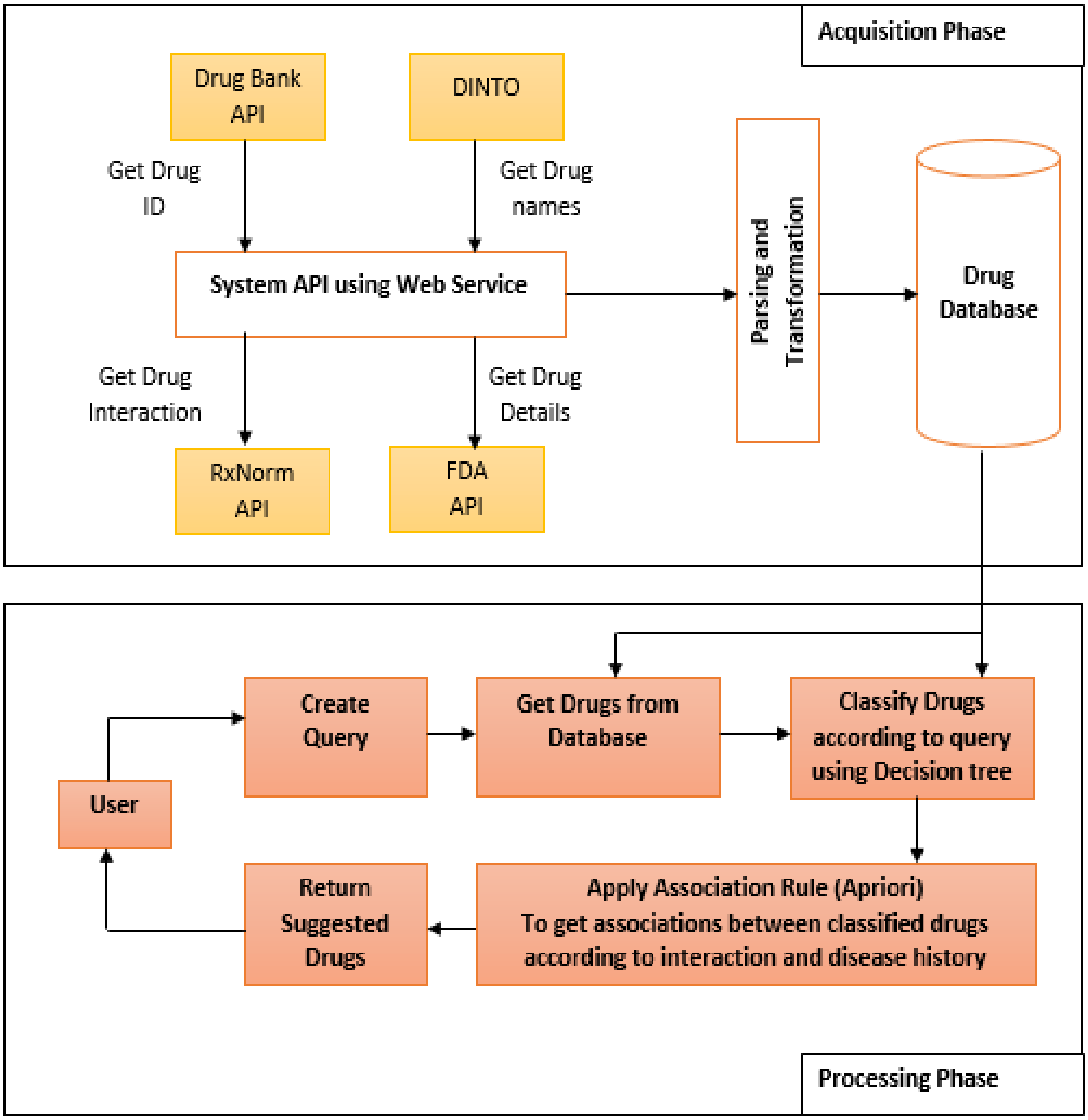


Figure 4.1: Proposed Framework

4.1 DRUG INFORMATION ACQUISITION

In this section, drug information acquisition, which is the first phase of the proposed tool, is described in detail. For easier reading, Figure 4.1 has been divided into two figures (Figure 4.2 and Figure 4.11), which are included in the corresponding sections. As shown in Figure 4.2, this phase consists of a set of steps ended with building the drug database.

- **Specification**: in which we intend to build integrated drug interaction by using information from different resources.

- **Acquisition**: in which the drug information is collected from different sources and stored in a relational database.

In this section, drug information acquisition is described in detail. For easier reading, Figure 4.2 and Figure 4.3 are included in the corresponding sections. As shown it consists of a set of steps ended with building the drug interaction database.

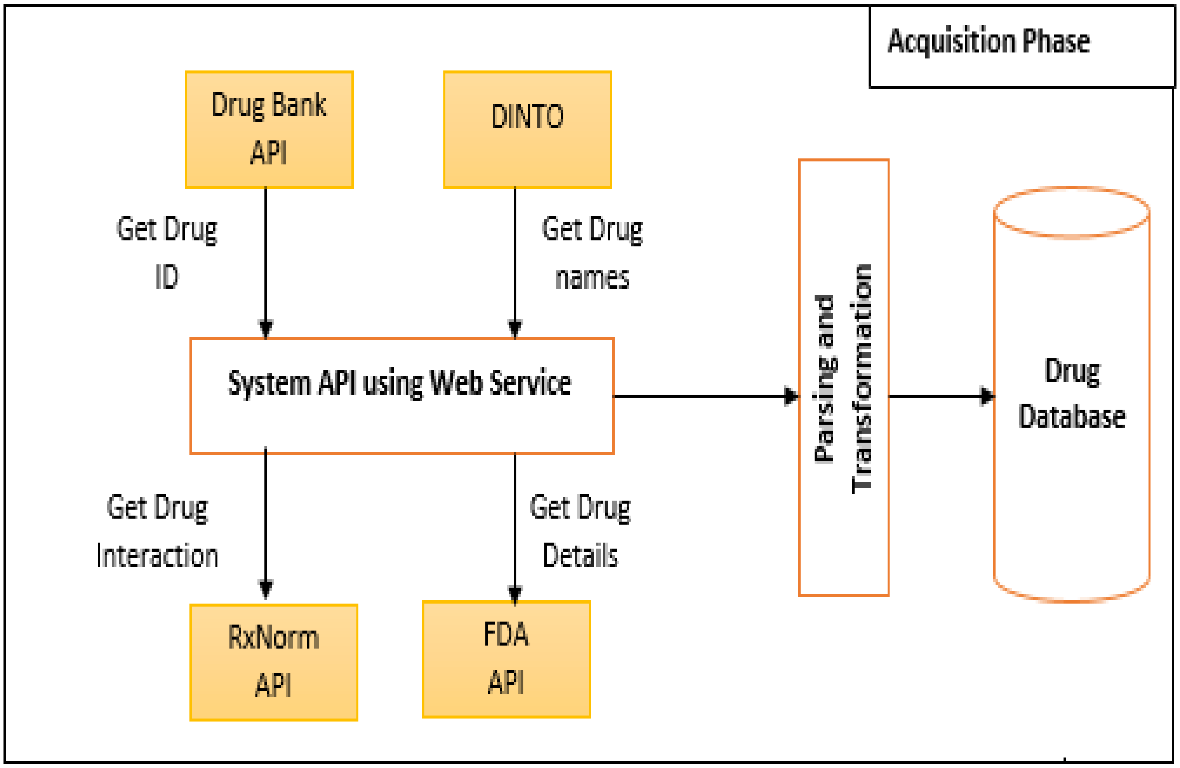


Figure 4.2: Drug information acquisition and representation phase

Drug information has been collected from different sources, including DINTO, DrugBank, RxNorm, and FDA. The Flowchart of drug information acquisition is shown in Figure 4.3.

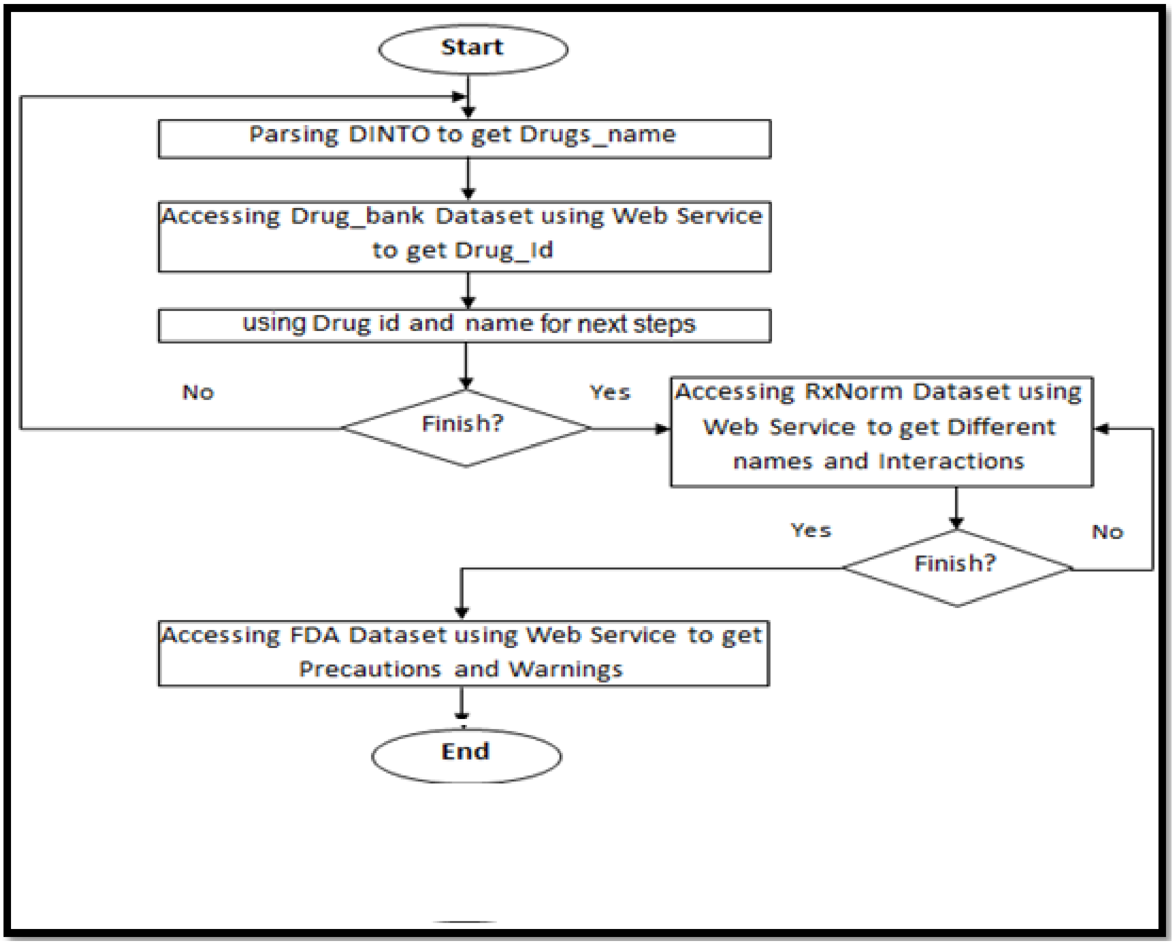
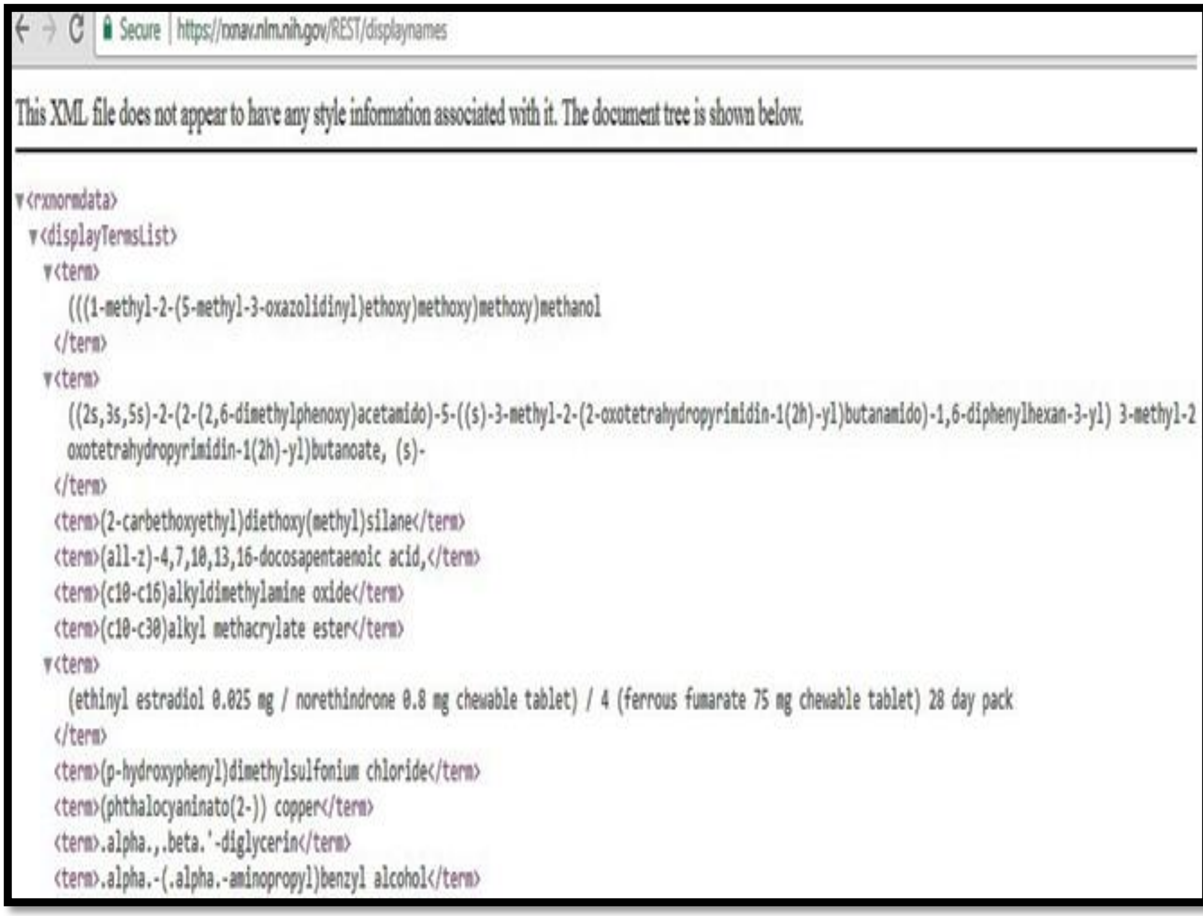


Figure 4.3: The flowchart of drug information acquisition process

The process of drug information acquisition starts with getting the drug names from DINTO. After that, for each drug name, an HTTP request is sent to DrugBank through API web service to get the global drug ID. Then, using the global drug ID and drug name, RxNorm database is searched using API web service to get the different drug names and drug interactions. National Library of Medicine has produce the RxNorm RESTful (Representational state transfer) web API which is a web service for accessing the RxNorm data.

The The web service can return the data in XML or JSON formats. The format can be specified by appending the extension (.xml or .json) to the URI before the query parameters. In order to get the different drug synonyms and drug interactions, RxCui (RxNorm drug id) must be obtained first using drug name and drug id. In our work, we obtained XML files that contain the different names for the same drug and other XML files that contain drug information. Samples of these files are shown in Figures 4.4 and 4.5, respectively.




```
▼<rxnormdata>
  ▼<displayTermsList>
    ▼<term>
       (((1-methyl-2-(5-methyl-3-oxazolidinyl)ethoxy)methoxy)methoxy)methanol
     </term>
    ▼<term>
       ((2s,3s,5s)-2-(2-(2,6-dimethylphenoxy)acetamido)-5-((s)-3-methyl-2-(2-oxotetrahydropyrimidin-1(2h)-yl)butanamido)-1,6-diphenylhexan-3-yl) 3-methyl-2
       oxotetrahydropyrimidin-1(2h)-yl)butanoate, (s)-
     </term>
     <term>(2-carbethoxyethyl)diethoxy(methyl)silane</term>
     <term>(all-z)-4,7,10,13,16-docosapentaenoic acid,</term>
     <term>(c10-c16)alkyldimethylamine oxide</term>
     <term>(c10-c30)alkyl methacrylate ester</term>
    ▼<term>
       (ethinyl estradiol 0.025 mg / norethindrone 0.8 mg chewable tablet) / 4 (ferrous fumarate 75 mg chewable tablet) 28 day pack
     </term>
     <term>(p-hydroxyphenyl)dimethylsulfonium chloride</term>
     <term>(phthalocyaninato(2-)) copper</term>
     <term>.alpha.,.beta.'-diglycerin</term>
     <term>.alpha.-(.alpha.-aminopropyl)benzyl alcohol</term>
```


Figure 4.4: A sample of an XML file that contains the different names for each drug and its information's such as drug contrition and chemical compounds and ingredients.

Figure 4.5: A sample of an XML file that contains information about a certain drug (Cymbalta drug and its other names with its Ids for each one of them in RxNorm database).

After obtaining the drug information from RxNorm database, Rxcui is used to get the drug interactions. A sample of an XML file that contains the drug interaction of a specific drug is shown in Figure 4.6. RESTful Drug Interaction API from RxNorm is a web service that is useful for accessing drug-drug interactions.

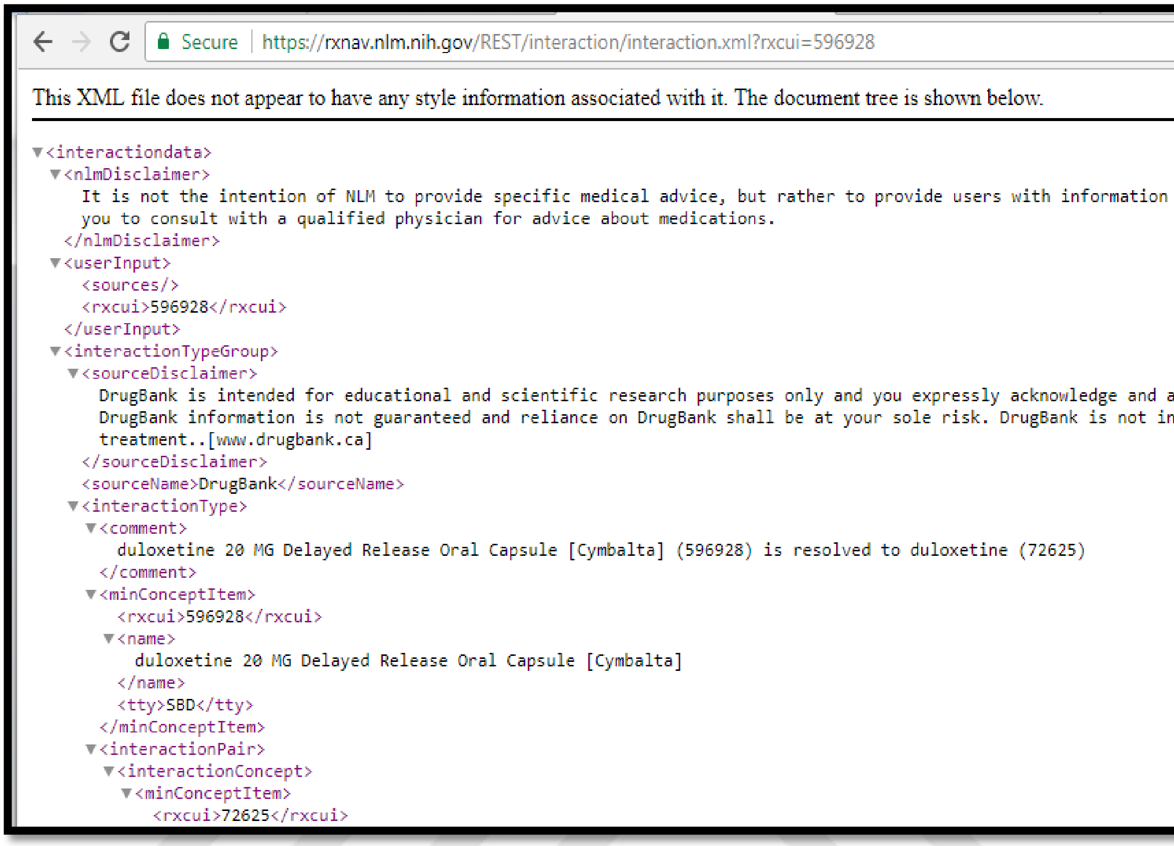


Figure 4.6: A snippet of XML file that contains Cymbalta drug and all drugs that interact with.

After getting the XML files of the different drugs, a parsing processing is applied over these files. The parsing code is shown in Figure 4.7.

```
function parseDrugOWLFiles($files){
    $model = ModelFactory::getDefaultModel();
    $temp = fopen("temp/txt.txt",'w');
    foreach($files as $file){
        $model->load($file);
        $it = $model->getStatementIterator();
        while ($it->hasNext()) {
                $about = $it->next();
                $label = $it->next();
                $resource = $it->next();
                $id = $it->next();

                if(in_array($label->getLabelObject(),$array)) continue;
                else $array[$id->getLabelObject()] = $label->getLabelObject()

        }
    fclose($temp);
   }
    return $array;
}
```

Figure 4.7: A snippet of PHP parsing code to obtain all drug interactions information

Additionally, the drug precautions, warnings, adverse effects, indications and, usage, are obtained from USA Food and Drug Administration (FDA), FDA contains a documentation that submitted by Drug producers and suppliers about their products. It is a necessary for labeling contains a summary of the essential scientific information important for the effective and safe use of the drug. The openFDA drug product labeling API returns data from these submissions for both prescribed and over-the-counter (OTC) drugs. Which are further broken down into sections, such as indications for use (prescription drugs) or purpose (OTC drugs), adverse reactions, and so forth [89]. Endpoint's data may be downloaded in zipped JSON files. Records are represented in the same format as API calls to this endpoint. Each update to the data in this endpoint could change old records [90]. In thesis, we access FDA using by passing drug name and id that obtained from DRUGBANK in order to retrieve fields such as Boxed_warning, precautions, dosage_and_administration, and adverse_reactions. HTTP Requests with URL using specific to the drug labeling endpoint. The PHP code written for getting this drug information is shown in Figure 4.8.

```php
foreach($arrayofDrugs as $drug){
  $drugId = array_search($drug,$arrayofDrugs);
  $query = "$drug";
  $query = str_replace(' ', '%20', $query);
  $josnText = file_get_contents("https://api.fda.gov/drug/label.json?api_key=$key&search=$query:boxed_warning");
  $obj = json_decode($josnText,true);

  //$brand_name = $obj['results'][0]['openfda']['brand_name'][0];
  if($arrayofDrugs[$i]==null) continue;
  $purpose = str_replace("'","",$obj['results'][0]['purpose'][0]);
  $warning = str_replace("'","",$obj['results'][0]['warnings'][0]."***".$obj['results'][0]['keep_out_of_reach_of_
```

Figure 4.8: A snippet of PHP code for obtaining drug information from FDA

Then all this information inserted to the database by using PHP code as shown in Figure 4.9.

```php
<?php
//error_reporting(0);
set_time_limit(0);

$connection = mysql_connect("127.0.0.1","root","")or die(mysql_error());
mysql_select_db("test_drug",$connection) or die(mysql_error());


$q = "select * from drug_info";
$r = mysql_query($q,$connection)or die(mysql_error());
while($f = mysql_fetch_array($r)){
    $josnText = file_get_contents("https://rxnav.nlm.nih.gov/REST/interaction/interaction.json?rxcui={$f['rxcui']}");
    $obj = json_decode($josnText,true);
    for($i=0;$i<count($obj["interactionTypeGroup"][0]["interactionType"][0]["interactionPair"]);$i++){
        $drug1 = $obj["interactionTypeGroup"][0]["interactionType"][0]["interactionPair"][$i]["interactionConcept"][0]["sourceConceptItem"]["name"];
        $drug1_id = $obj["interactionTypeGroup"][0]["interactionType"][0]["interactionPair"][$i]["interactionConcept"][0]["sourceConceptItem"]["id"];
        $drug1_url = $obj["interactionTypeGroup"][0]["interactionType"][0]["interactionPair"][$i]["interactionConcept"][0]["sourceConceptItem"]["url"];

        $drug2 = $obj["interactionTypeGroup"][0]["interactionType"][0]["interactionPair"][$i]["interactionConcept"][1]["sourceConceptItem"]["name"];
        $drug2_id = $obj["interactionTypeGroup"][0]["interactionType"][0]["interactionPair"][$i]["interactionConcept"][1]["sourceConceptItem"]["id"];
        $drug2_url = $obj["interactionTypeGroup"][0]["interactionType"][0]["interactionPair"][$i]["interactionConcept"][1]["sourceConceptItem"]["url"];

        $interaction = $obj["interactionTypeGroup"][0]["interactionType"][0]["interactionPair"][$i]["description"];
        if($i>9) break;
        mysql_query("insert into drug_interactions (drug1_id,drug1_url,drug2_id,drug2_url,interaction) values ('$drug1_id','$drug1_url','$drug2_id','$dr
    }
}
mysql_close($connection);

?>
```

Figure 4.9: A snippet of PHP code that used to insert drug information in database.

Drug_id	rxnorm_id	Drug_name	generic_name	uses	warning_precautions	side_effect	
DB00682	11289	Warfarin Oral Tablet	Warfarin	Both -- Preventing harmful blood clots helps to reduce the risk of a stroke or heart attack. Conditions that increase your risk of developing blood clots include a certain type of irregular heart rhythm , heartvalve replacement, recent heart attack, and certain surgeries --heart attack,bleeding,heart rhythm,heart diseases	warning: Warfarin can cause very serious bleeding-- pregnancy: seek professional advice before taking this product -- precautions: blood disorders such as anemia, hemophilia, bleeding problems such as bleeding of the stomach intestines, bleeding in the brain, blood vessel disorders such as aneurysms, recent major injury surgery, kidney disease, liver disease, alcohol use, mental mood disorders including memory problems , frequent falls injuries.	Tell your doctor right away if you have any signs of serious bleeding, including: unusual pain swelling discomfort, unusual easy bruising, prolonged bleeding from cuts or gums, persistent frequent	https://img.medscapestatic.com/pi/fe
DB01551	1191	Aspirin Tablet	Aspirin	Both -- Aspirin is used to reduce fever , headache and relieve mild to moderate pain from conditions such as muscle aches, toothaches, common cold, and headaches --pain,headache	precautions: kidney disease, liver disease, diabetes, stomach problems such as ulcers, heartburn, stomach pain, aspirin-sensitive asthma , growths in the nose, gout, certain enzyme deficiencies -- pregnancy:Do not use this medication during the last 3 months of pregnancy because of possible harm to the unborn baby or problems during delivery	easy bruising bleeding, difficulty hearing, ringing in the ears, signs of kidney problems, change in the amount of urine, persistent or severe nausea vomiting, unexplained tiredness, dizziness, dark urine, yellowing eyes and skin	https://img.medscapestatic.com/pi/fe
DB00482	215927	Celebrex	Celecoxib	Both --It is used to treat arthritis, acute pain, and menstrual pain and discomfort. The pain and swelling relief provided by this medication helps you perform more of your normal daily activities -- stomach problems,Acidity	precautions: liver disease, stomach intestine esophagus problems such as bleeding, ulcers, recurring heartburn , heart disease (such as angina, heart attack , high blood pressure, stroke, blood disorders such as anemia, bleeding clotting problems , growths in the nose		https://img.medscapestatic.com/pi/fe
DB00390	3407	Digoxin	Digoxin	Both -- Digoxin is used to treat heart failure. It is also used to treat certain types of irregular heartbeat --heart failure,heart attack,heart diseases	precautions: kidney problems, thyroid problems-- warning:effects on the heartbeat -- pregnancy:this medication should be used only when clearly needed. Discuss the risks and benefits with your doctor	This drug may make you dizzy or blur your vision	https://img.medscapestatic.com/pi/fe
DB00375	104485	Colestipol HCL	Colestipol	Both -- Colestipol is used along with a	precautions: swallowing	Constipation, stomach/abdominal	https://img.medscapestatic.com/pi/fe

Figure 4.10: Snipped of database that show all information of drugs

The second phase toward building the proposed interactive tool, as shown in Figure 4.11, employs a number of techniques including selecting appropriate drugs, apriori algorithm and applying decision rules to benefit from the build ontology during the prescription process.

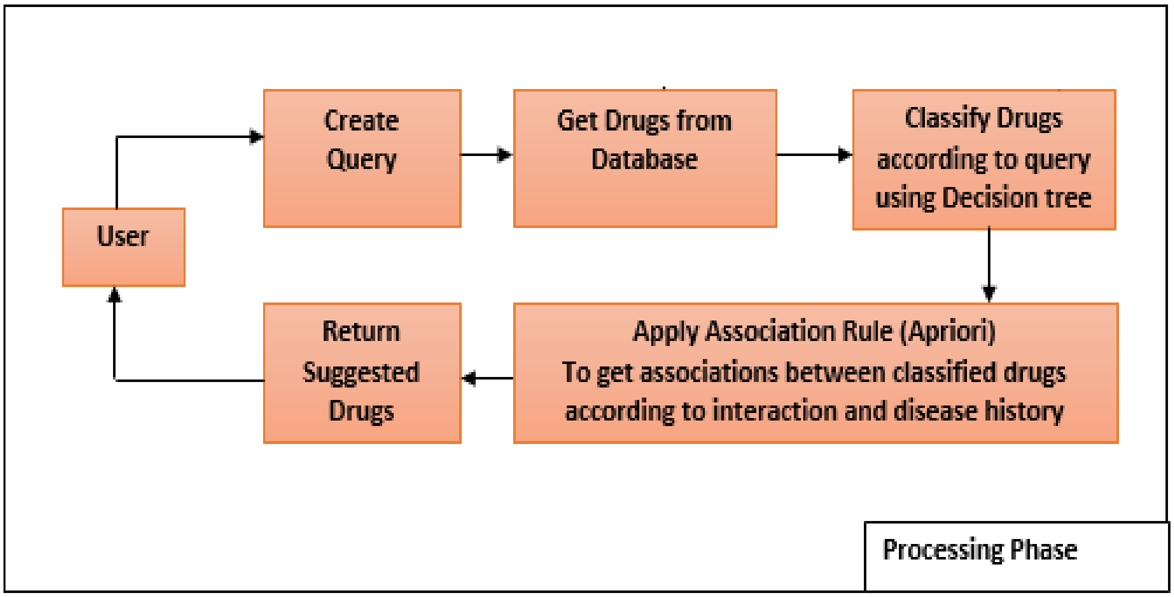


Figure 4.11: The drug processing phase

Based on a query provided by the user, which include patient physical information (age, gender, pregnancy), drug history (current diseases and the drugs that are currently taken by the patient), and the newly diagnosed disease, the proposed tool returns a list of drug which can be used for the current disease accompanied with a notification with each drug to demonstrate if there is a conflict (drug-drug interaction, drug-disease interaction, cannot be used because of the age or pregnancy of the current patient, etc) between the drug from one side, and the patient's physical information and drug history from the other side.

In this phase, a number of steps are performed. The first step is extracting all the drugs for the current case, then these drugs are evaluated to show whether they are suitable or not. In order to evaluate these drugs, they are selected based on age and gender. Then, the apriori algorithm is employed to discover the associations between the drugs and the patient's drug history (current diseases and the drugs that are currently taken by the patient). Finally, the decision tree algorithm is applied to make a decision about the candidate drugs.

5 EXPERIMENTAL RESULT

5.1 DEVICE AND TOOL CAPABILITIES

The proposed interactive tool has been implemented using a number of technologies, including Apache server version 2, XML (Structure format), PHP programming Language (Send HTTP Request and convert response to XML), and MSQL (Database version to store response query) version 5.1. The implemented database and consequently the implemented ontology contain information about 30 drugs. In the following subsections, we provide two test cases with different physical information and different drug history.

Table 5.1: Tools and device used to preform proposed framework

Metric	Values
CPU	Intel core i5
RAM	2G
Operating system	Windows 7
Programming Language	PHP v4
Server Platform	Apache server v2
Database	MYSQL Version 5

5.2 TEST CASES

5.2.1 Test Case I:

In this test case, the system is triggered by a query, as show in figure 5.1 and figure 5.2 respectively, with patient physical information (Name = “X”, Age Range = “41-65”, Gender = “Female”, Pregnancy = “No”), drug history (Diseases = “stomach upset, low blood pressure, coronary heart failure”, Drugs = “concor 5, omeprazole”), and current disease (Current Disease = “Diabetes”). The system has responded, as shown in Figure 5.1b, with one drug called “Metformin Tablet” as a suitable medication, accompanied with drug Id, usage side effect, precautions, DDIs, and drug diseases interactions.

5.2.2 Test Case II:

In this test case, the system is triggered by a query, as show in figure 5.3, figure 5.4 , figure 5.5 and figure 5.6 respectively, with patient physical information (Name = “Y”, Age Range = “21-40”, Gender = “Female”, Pregnancy = “Yes”), drug history (Diseases = “high blood pressure”, Drugs = “digoxin”), and current disease (Current Disease = “heart attack”). The system has responded, as shown in figure 5.4, figure 5.5 and figure 5.6 respectively with drugs list, including “Warfarin Oral Tablet”, “Digoxin”, and “Colestipol HCL” as a suitable medication, accompanied with drug Id, usage side effect, precautions, DDIs, and drug diseases interactions. Also, we can notice that a notification is given with the candidate drug in case of detecting drug-drug interaction, drug-disease interaction, and/ or a conflict between patient’s physical information and the candidate drug.

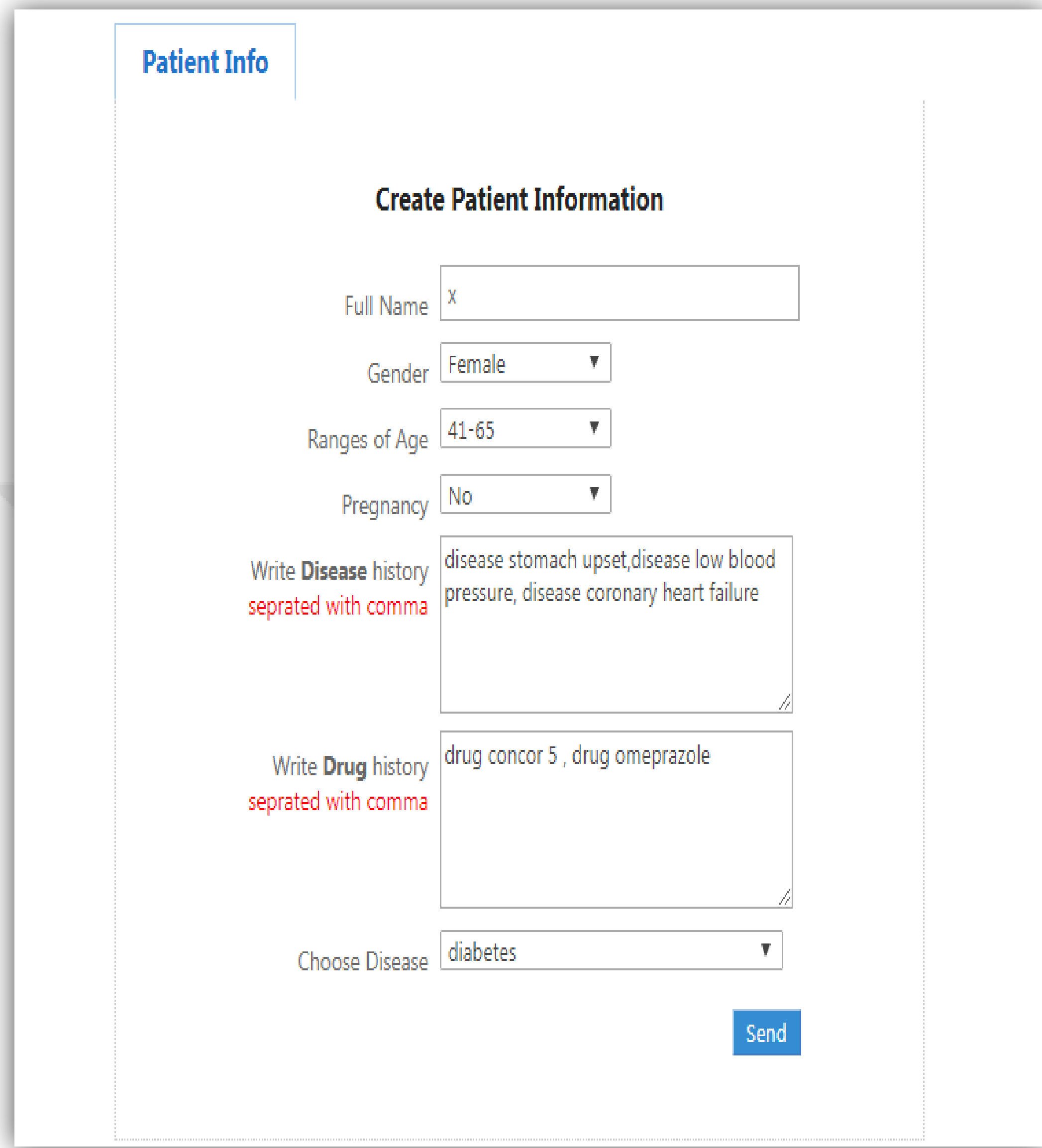


Figure 5.1: Screen Shot of Preparing the Query of Test Case 1

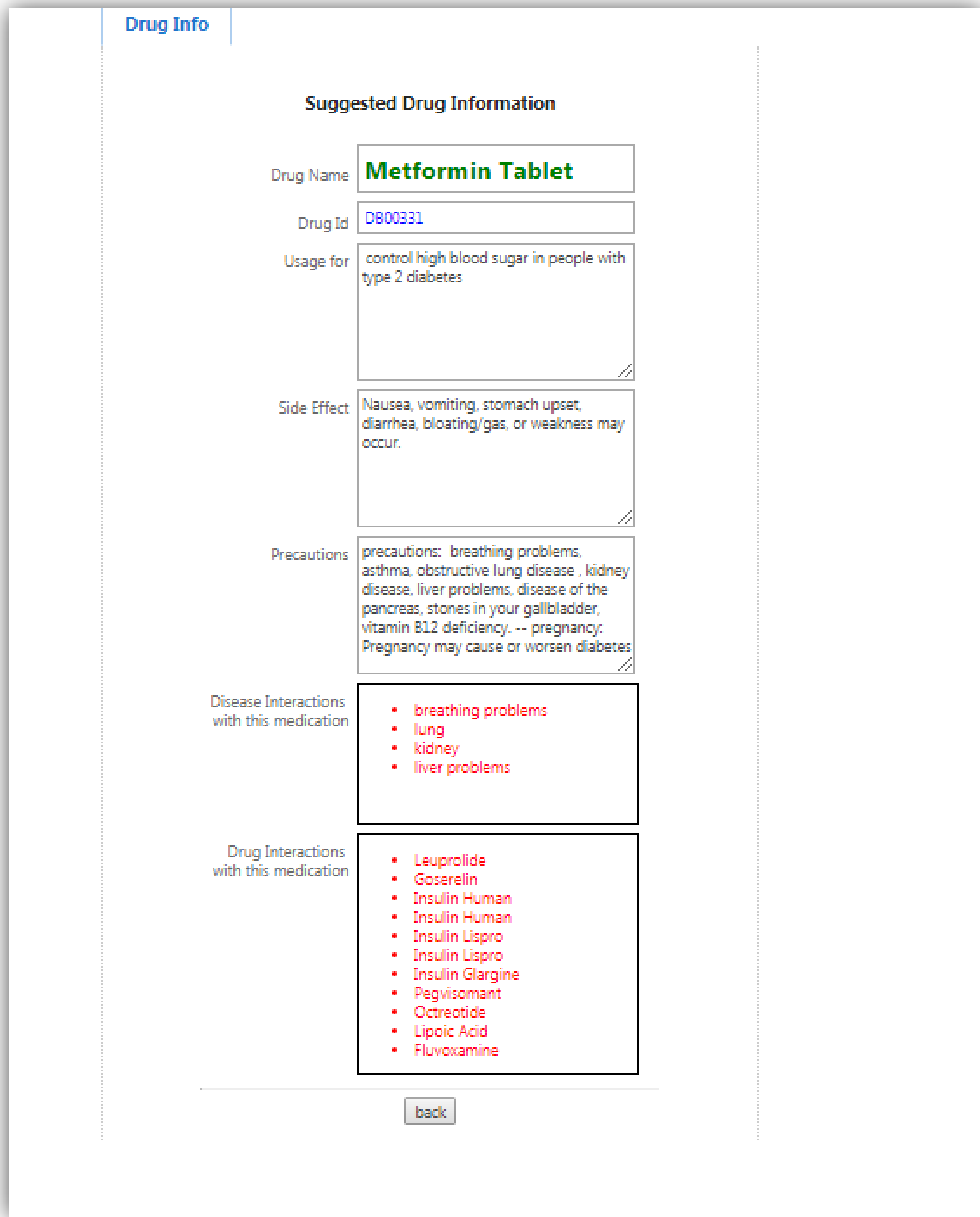


Figure 5.2: Screen shot of System Response for the query of Test Case 1

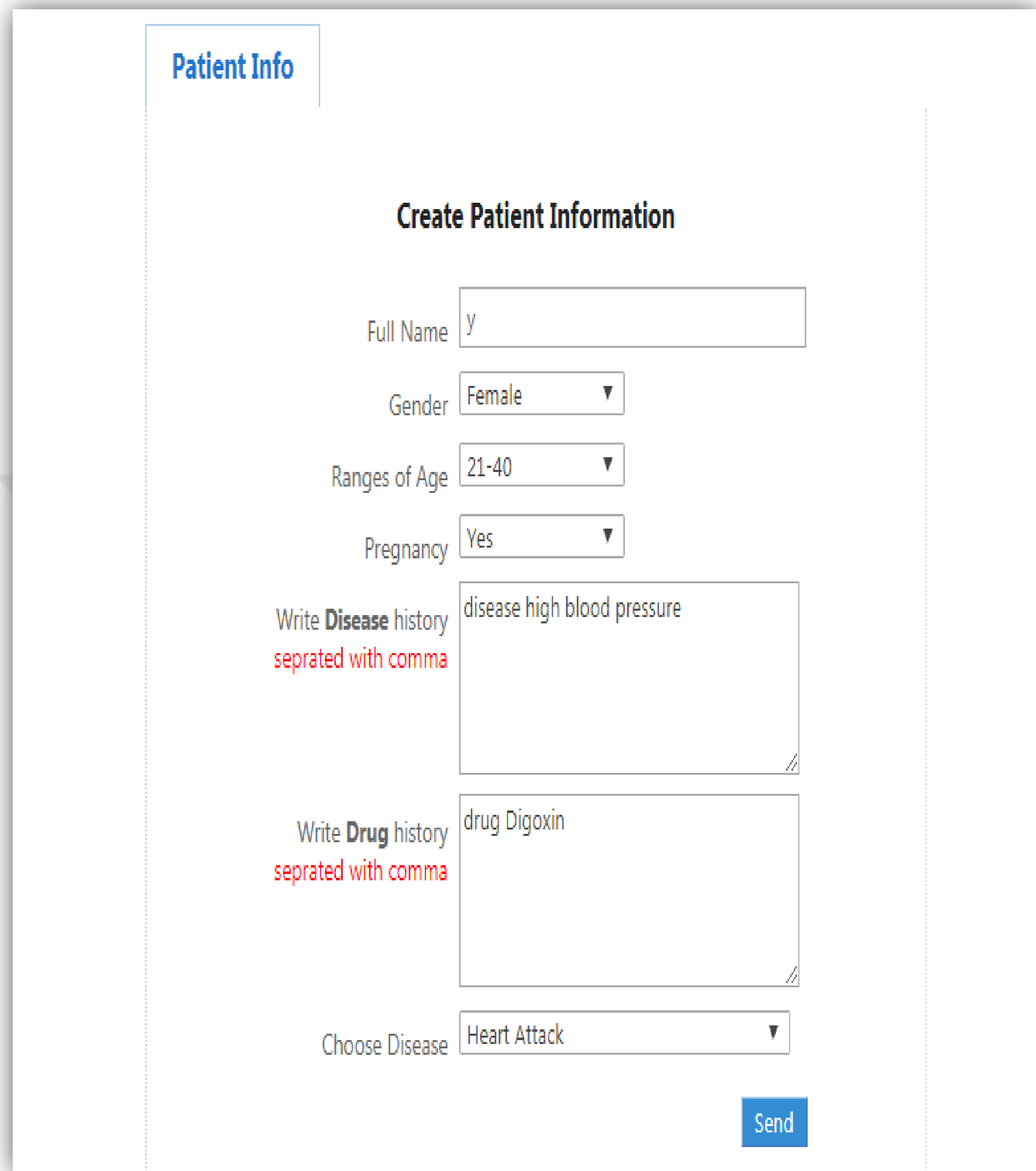


Figure 5.3: Screen Shot of Preparing the Query of Test Case 2

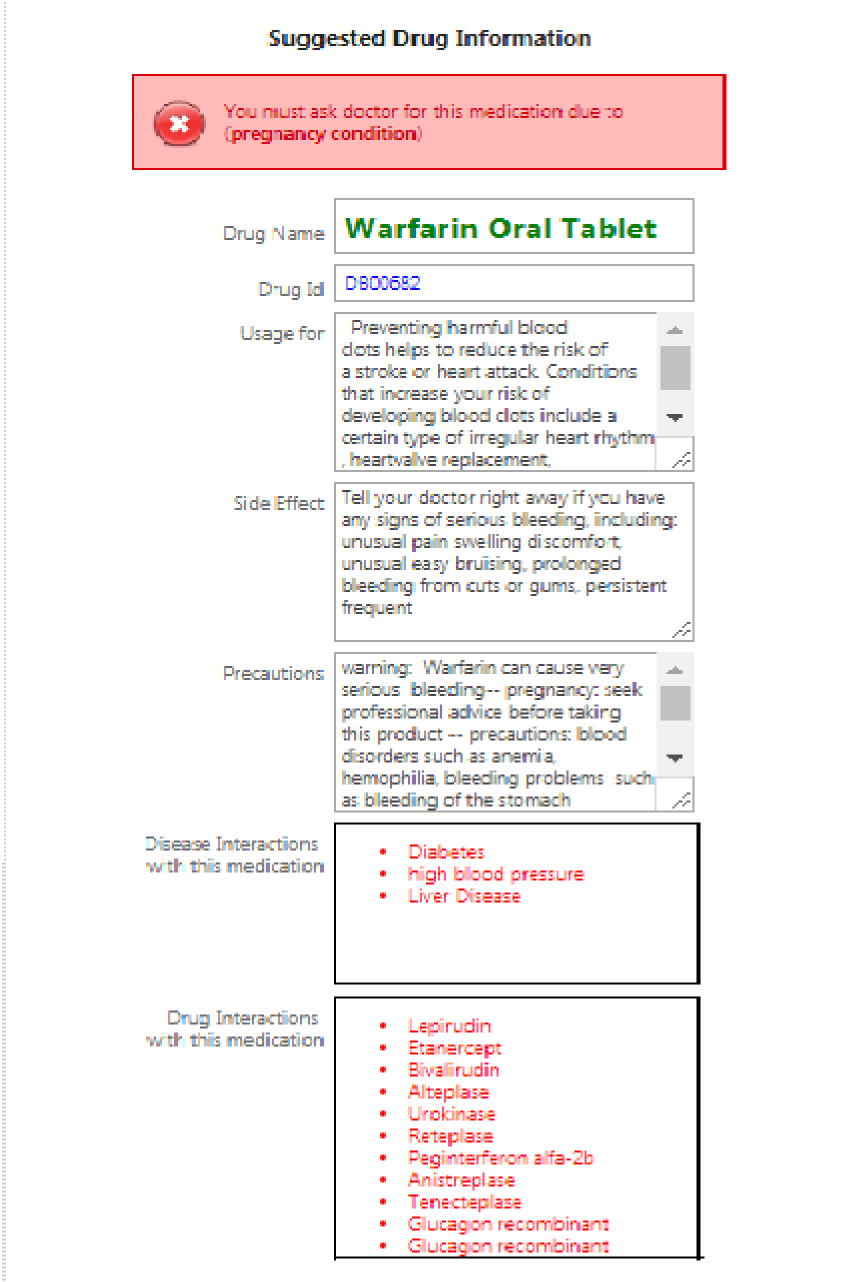


Figure 5.4: Screen shot of the First System Response for the query of Test Case 2

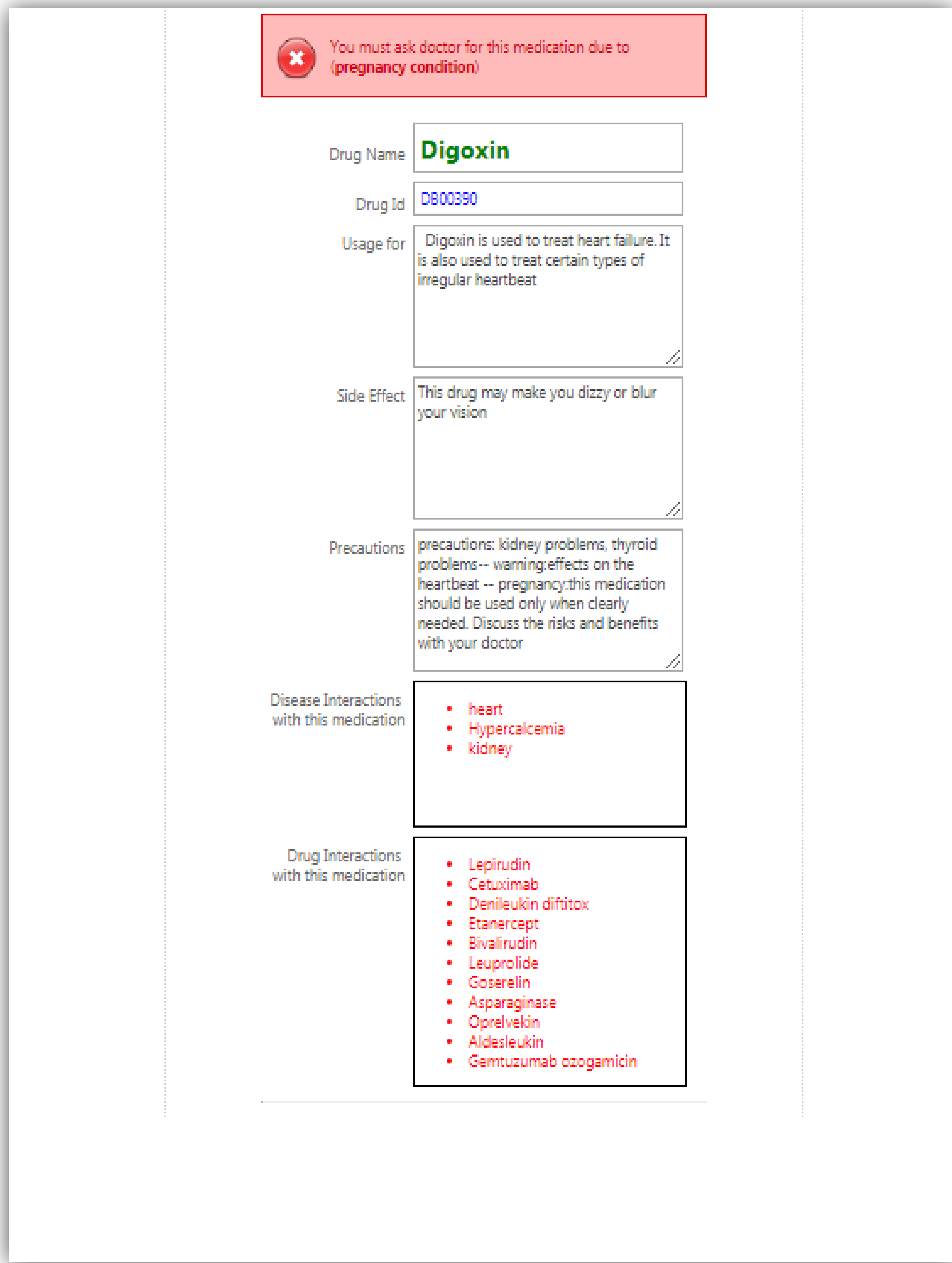


Figure 5.5: Screen shot of the Second System Response for the query of Test Case 2

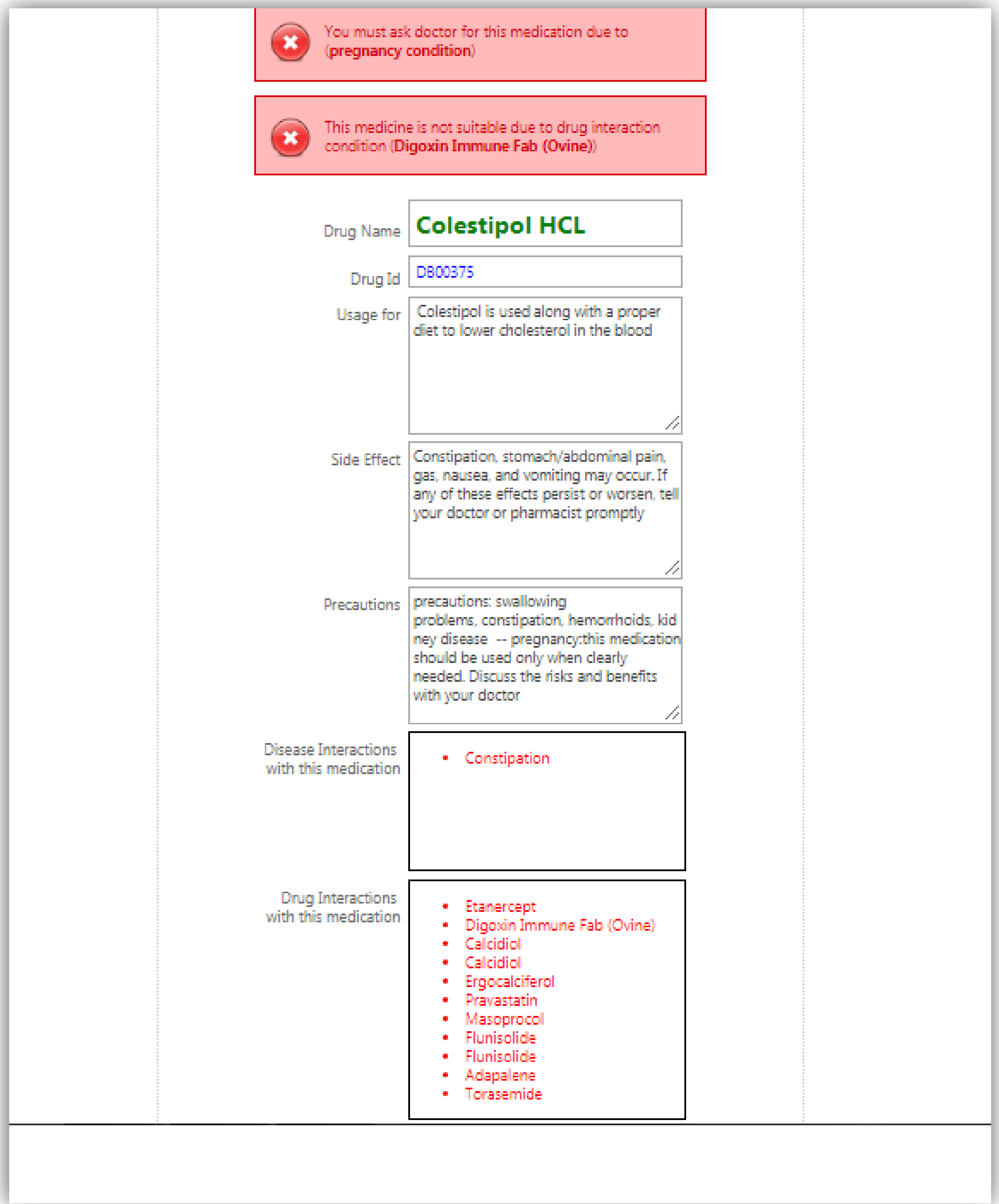


Figure 5.6: Screen shot of the third System Response for the query of Test Case 2

5.2.3 Test Case III:

In this test case, the system is triggered by a query, as show in figure 5.7 and figure 5.8, with patient physical information (Name = "Z", Age Range = "41-60", Gender = "Female", Pregnancy = "No"), drug history (Diseases = "cold", Drugs = "congestal"), and current disease (Current Disease = "headache"). The system has responded, as shown in figure 5.8 with specific drug, including "Aspirin Tablet" as a suitable medication, accompanied with drug Id, usage side effect, precautions, DDIs, and drug diseases interactions. Also, we can notice that a notification is given with the candidate drug in case of detecting drug-drug interaction, drug-disease interaction, and/ or a conflict between patient's physical information and the candidate drug.

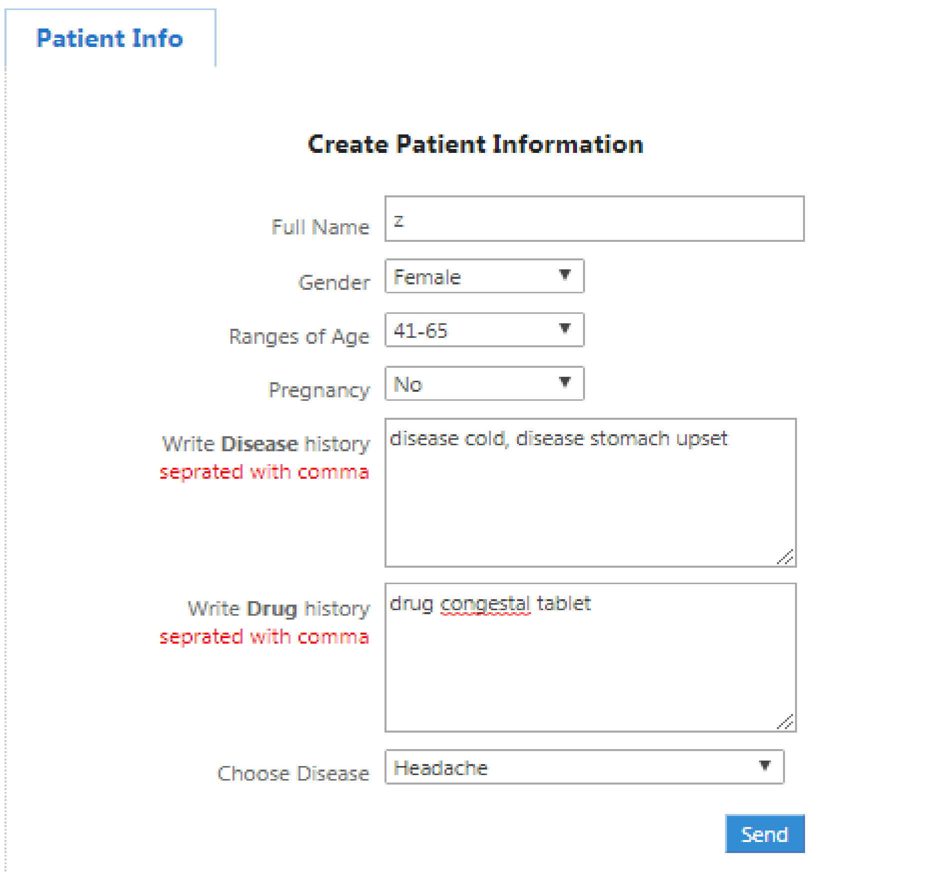


Figure 5.7: Screen Shot of Preparing the Query of Test Case 3

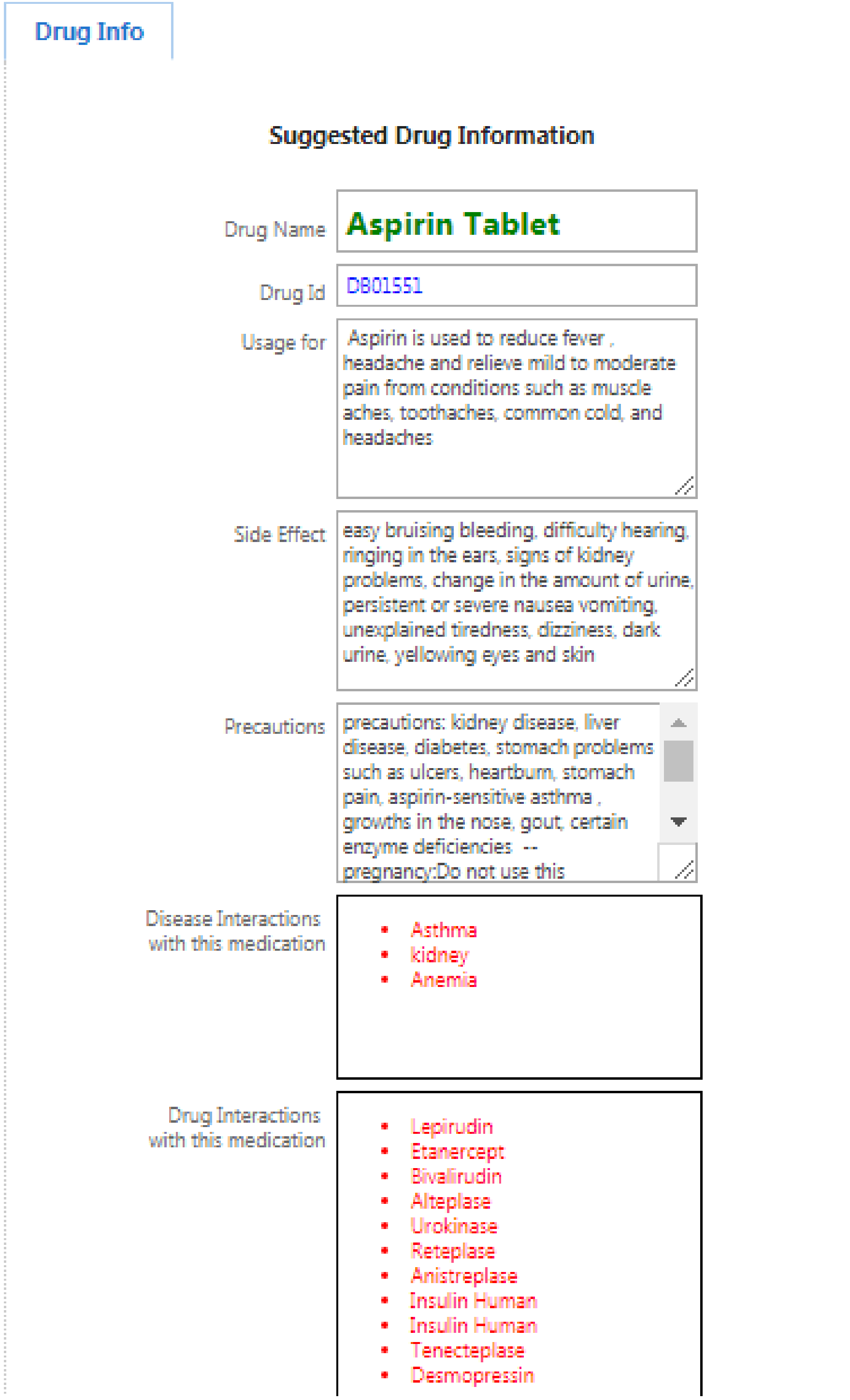


Figure 5.8: Screen shot of System Response for the query of Test Case 3

5.3 EXPERIMENTAL RESULT AND DISCUSSION

For training and testing the data sets, we use ten-fold cross validation technique. This technique splits the data set into 10 portions. Nine portions are then applied for training and the 10th fragment is applied for testing. This is then recurring, using alternative portion as the test section. Individually data portion is utilized once for testing and 9 times for training. This is then recurrent 10 times, with a novel portion being the testing piece. An average outcome is produced from the 10 runs. The accuracy of the applied procedures has been evaluated using in terms of correctly records, incorrectly records, TP rate, FP rate, precision, recall, ROC area, CPU Time, accuracy, error. We used different measures to evaluate the methods and show results in figure 5.9 adopted on the heart diseases dataset as shown below:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \quad 5.1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad 5.2)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \quad 5.3)$$

$$\text{F-measure} = \frac{\text{Precision+ Recall}}{2} \times 100\% \quad 5.4)$$

Table 5.2: Reuslt of Proposed Test Cases

Test Cases	**Recall**	**Precision**	**F − measure**	**Accuracy**
Test case I	98.2%	97.8%	98%	96.5%
Test case II	98.6%	95.6%	97.1%	96.2%

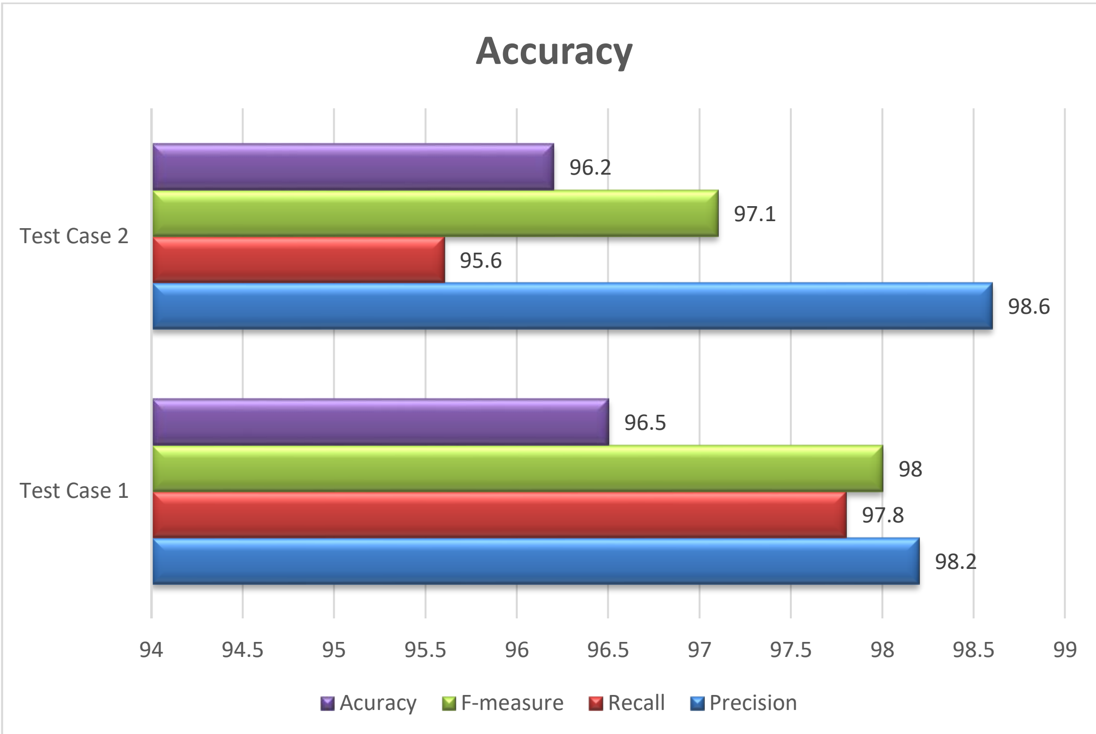


Figure 5.9: Accuracy and Error Diagram of Proposed Test Cases

6 CONCLUSION

Nowadays, there are thousands of approved drugs that can be used for treating people who have medical problems. Despite of the usefulness of drugs in treating many diseases, many of them can cause serious adverse reactions, or otherwise clinically significant adverse reactions. The effects of these adverse reactions can cost the lives of the patients. Therefore, medicine Warnings and Precautions sections are dedicated to recognize and represent a discrete assortment of adverse effects and other implied protection risks that are serious or are differently clinically crucial because they have indications for guiding decisions or for inmate administration.

In order to prescribe the suitable drugs without jeopardizing the patients' safety, physicians should know the adverse reactions for a big number of drugs. However, it is difficult to remember the warnings and precautions associated with a large number of drugs. Therefore, there was a necessity to build an automated system for representing the different types of drug interactions that can endanger the patients' safety.

In this thesis, proposed tool takes the patient personal information, such as gender, age, pregnancy, the current diseases that the patient suffers from, the current drugs, and the newly diagnosed disease as input and returns the suggested drugs that suit the current patient state. The proposed tool consists of two main phases: data acquisition and data processing. In data acquisition phase, drug information has been collected from different sources including DINTO, DrugBank, RxNorm, and FDA. All collected information is inserted into a centralized database. In data processing phase, a number of data mining methods have been applied on the constructed ontology to suggest the suitable drugs for the patients. Given patient physical information (age, gender, pregnancy), drug history (current diseases and the drugs that are currently taken by the patient), and the newly diagnosed disease, the proposed tool returns a list of drug which can be used for the current disease accompanied with a notification with each drug to demonstrate if there is a conflict.

Moreover, a number of test cases have been designed for evaluating the accuracy of the proposed tool in different scenarios. The test cases have shown the efficiency and effectiveness of the proposed tool.

REFERENCES

[1] S. J. Nelson, K. Zeng, J. Kilbourne, T. Powell, and R. Moore, "Normalized names for clinical drugs: RxNorm at 6 years," *Journal of the American Medical Informatics Association,* vol. 18, pp. 441-448, 2011.

[2] *Drug bank homepage. Accessed 02 June 2018.* Available: https://www.nlm.nih.gov/healthit/snomedct/index.html

[3] "Mesh on demand homepage. Accessed 13 Feb 2017.."

[4] M. Pirmohamed, A. M. Breckenridge, N. R. Kitteringham, and B. K. Park, "Adverse drug reactions," *Bmj,* vol. 316, pp. 1295-1298, 1998.

[5] M. Liu, M. E. Matheny, Y. Hu, and H. Xu, "Data mining methodologies for pharmacovigilance," *ACM SIGKDD Explorations Newsletter,* vol. 14, pp. 35-42, 2012.

[6] e. S. s. Stockley HI, "Drug Interactions. 5th ed. London," *Chicago: The Pharmaceutical Press,* 2014.

[7] S. V. Doubova, H. Reyes-Morales, L. del Pilar Torres-Arreola, and M. Suárez-Ortega, "Potential drug-drug and drug-disease interactions in prescriptions for ambulatory patients over 50 years of age in family medicine clinics in Mexico City," *BMC health services research,* vol. 7, p. 147, 2007.

[8] K. Baxter and C. Preston, *Stockley's drug interactions* vol. 495: Pharmaceutical Press London, 2010.

[9] S. L. D. Tatro, "The Authority on Drug Interactions," *MO: Wolters Kluwer Health,* 2010.

[10] "Drug bank homepage. https://www.drugbank.ca/. Accessed 02 June 2018.."

[11] K. S. S. Yoshikawa, and A. Konagaya, "Drug interaction ontology (DIO) for inferences of possible drug-drug interactions," *Studies in health technology and informatics,* vol. 107, pp. 454–8, 2014.

[12] H.-Y. Wu, S. Karnik, A. Subhadarshini, Z. Wang, S. Philips, X. Han, *et al.*, "An integrated pharmacokinetics ontology and corpus for text mining," *BMC bioinformatics,* vol. 14, p. 35, 2013.

[13] R. Boyce, C. Collins, J. Horn, and I. Kalet, "Computing with evidence: Part I: A drug-mechanism evidence taxonomy oriented toward confidence assignment," *Journal of biomedical informatics,* vol. 42, pp. 979-989, 2009.

[14] M. Herrero-Zazo, I. Segura-Bedmar, P. Martínez, and T. Declerck, "The DDI corpus: An annotated corpus with pharmacological substances and drug–drug interactions," *Journal of biomedical informatics*, vol. 46, pp. 914-920, 2013.

[15] S. V. R.Sivasankari, "An Effective Re-clustering Based Submit Selection Using MST and Heuristic Algorithm," *Journal INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET),* pp. pp. 1310 - 1314, 2016.

[16] S. Vegas, N. Juristo, and V. R. Basili, "Maturing software engineering knowledge through classifications: A case study on unit testing techniques," *IEEE Transactions on Software Engineering,* vol. 35, pp. 551-565, 2009.

[17] J. F. Sowa, *Knowledge representation: logical, philosophical, and computational foundations* vol. 13: Brooks/Cole Pacific Grove, CA, 2000.

[18] A. Bernard and Y. Xu, "An integrated knowledge reference system for product development," *CIRP annals*, vol. 58, pp. 119-122, 2009.

[19] D. Tomar and S. Agarwal, "A survey on Data Mining approaches for Healthcare," *International Journal of Bio-Science and Bio-Technology*, vol. 5, pp. 241-266, 2013.

[20] S. Patel and H. Patel, "Survey of data mining techniques used in healthcare domain," *International Journal of Information*, vol. 6, 2016.

[21] S. Nagarajan and R. Chandrasekaran, "Design and implementation of expert clinical system for diagnosing diabetes using data mining techniques," *Indian Journal of science and Technology*, vol. 8, pp. 771-776, 2015.

[22] S. Mumbaikar and P. Padiya, "Web services based on soap and rest principles," *International Journal of Scientific and Research Publications*, vol. 3, pp. 1-4, 2013.

[23] R. Agarwal and R. Srikant, "Fast algorithms for mining association rules," in *Proc. of the 20th VLDB Conference*, 1994, pp. 487-499.

[24] A. Raihani and N. Laachfoubi, "EXTRACTING DRUG-DRUG INTERACTIONS FROM BIOMEDICAL TEXT USING A FEATURE-BASED KERNEL APPROACH," *Journal of Theoretical & Applied Information Technology*, vol. 92, 2016.

[25] A. B. Abacha, M. F. M. Chowdhury, A. Karanasiou, Y. Mrabet, A. Lavelli, and P. Zweigenbaum, "Text mining for pharmacovigilance: Using machine learning for drug name recognition and drug–drug interaction extraction and classification," *Journal of biomedical informatics*, vol. 58, pp. 122-132, 2015.

[26] Z. Zhao, Z. Yang, L. Luo, H. Lin, and J. Wang, "Drug drug interaction extraction from biomedical literature using syntax convolutional neural network," *Bioinformatics*, vol. 32, pp. 3444-3453, 2016.

[27] J. Björne, A. Airola, T. Pahikkala, and T. Salakoski, "Drug-drug interaction extraction from biomedical texts with svm and rls classifiers," *Proceedings of DDIExtraction-2011 challenge task*, pp. 35-42, 2011.

[28] G. G. Richard D. Boyce , Hendrik Harkema . , "Using SVM and Decision Tree to identify pharmacokinetic drug-drug interactions," *Conference: Proceedings of the Workshop on Biomedical Natural Language Processing*, 2018.