Research article

Healthcare system based on Semantic Web and XML Technologies

Magdy Shayboub Ali
Computer Science Dept.,
computers and informatics,
Ismaillia, Suez Canal University
and Applied Medical Sciences
College Of Taif University,
Kingdom of Saudi Arabia (KSA)
magdy01sh@yahoo.com

Khaled M. Fouad
Computer Science Dep. -
Community College - Taif
University, Kingdom of Saudi
Arabia (KSA)
kmfi25@yahoo.com

Yasser Nada
Computer Science Dep., computers
and information systems, Taif
University, Kingdom of Saudi
Arabia (KSA)
y_nada@yahoo.com

_____

Abstract

In the paper we used semantic web technologies for managing medical information during a diagnostic process is studied. Following a steady international move towards optimization of health care delivery, the latest development in information technology has drawn the health care industry decision makers' attention. The introduction of proper information technology innovations within the health care processes should provide the necessary optimization. In this manner we propose an approach to manage medical data during the whole diagnostic process using the semantic web technologies. The purpose of the Semantic Web is to bring structure to the content of Web pages allowing software agents to carry out intelligent tasks for the user. This opens a new set of opportunities that can be utilized to improve health care management on a personal and health care provider level. The aim of this research in progress is to identify the needs and match them to the services possible with the Semantic Web. In this paper, we present an ontology-based framework that successfully combines both XML and Semantic Web Services technologies to enable the integrated access to biological data sources. The main goal is the seamless integration and application of these technologies in such a way that their deficiencies are overcome and their utility maximized. **Copyright © <ins>www.acascipub.com</ins>, all rights reserved.**

Keywords: web technology, healthcare system, medical information, diagnostic process

_____

I. Introduction

The aim of the Semantic Web is to make the Web as intelligent as possible [1]. By describing relationships between objects and properties of objects the Web is no longer seen as links between web pages, but as one large database with information.
Together with virtual communities a patient can in future be provided with an individualized, easy to use, encompassing health care management program that adds value to his/her life.
Adoption of XML and related technology is a must for our intelligent agent system because of the following reasons [2]:

• Standardization of our system
• Widespread exchange of data among different systems
• Making the system semantic to both man and machine
• Development of true platform independence

Today, XML is invading the world of computers and occupying most of its fields. It is widely spreading over the internet, networks, information systems, software and operating systems, DBMS, search tools, web development and services, communication protocols and other fields. As a result, XML data are floating within and between different applications and systems all over the internet and intranets. Due to the huge amount of XML structured data being circulated, controlling XML data becomes imperative for various purposes and aims [3]. The system must be semantic by the usage of XML based indexing and XQuery based searching [4]. So that it become sociable to human and machine. All these characteristics will make the system under intelligent category [2].

In almost all human activities [5] we can discern three stages: observation, reasoning, and action. These stages in human activity play a role not only in daily life, but also in patient care, management, and research. Also in health care the same three stages can be seen in the so-called diagnostic-therapeutic cycle: 1) observation, 2) diagnosis, and 3) therapy. The semantic web turned out to be very appropriate for the integration of information resources by semantic annotation of data. Furthermore, the idea of semantic web allows automatic, intelligent inferring on knowledge, retrieved from these information resources.

The "Semantic Web" can be thought of an extension of the present web, as an additional machine-processable layer of data beneath the visible layer of human-readable information. In the research we will provide some examples and review the prospects of the Semantic Web for the field of knowledge management and knowledge translation in consumer health informatics for improving access to information, and for addressing questions around the quality of health information on the web.

One of the major obstacles for integration efforts in bioinformatics is that relevant information is widely distributed, both across the internet and within individual organizations. Besides, it can be found in a variety of storage formats, including structured and semi-structured ones. Biological data formats vary ranging from flat files to sophisticated relational databases. Because of growth of applicability and information interchange among the application ranges, XML standard has been chosen by many applications, for its flexibility and being industry standard. The Semantic Web provides a common framework that enables for data integration, sharing and reuse from multiple sources. Particularly, the use of ontologies for domain knowledge representation can help bioinformatics in solving the heterogeneity problem.

In semantic web, knowledge is represented as graphs, written down in XML-based language called RDF (Resource Description Framework). RDF is dealing with URIs (another W3C standard for naming resources globally unique). Advanced use of semantically annotated data can only be accomplished by using ontologies represented as RDFS or OWL documents. RDF (Resource Description Framework) and the Web Ontology Language (OWL) are used to explicitly represent the meanings of the resources described on the Web and how they are related. These specifications, called ontologies, describe the semantics of classes and properties used in Web documents.

XML Query Language or XQuery definitions is nearing completion as the next generation functional language for querying XML documents by the sincere standardization body, XQuery Working Group, which is formed by the industry leaders and researchers all over the world. Integration of XQuery on XML documents results in powerful query implementation which is fast and efficient. Moreover, since more and more wrapping services [6] are coming up with XML based views on underlying diversified sources with different file formats, using XQuery on those system for searching XML nodes will be of great efficiency.

In this paper, we present an ontology-based framework that successfully combines both Agent and Semantic Web Services technologies to enable the integrated access to biological data sources[7]. The main goal is the seamless integration and application of these technologies in such a way that their deficiencies are over come and their utility maximized.

Related works

Tran and Wataru [8] have extracted health care information from web pages with specific sources dedicated for only healthcare information. The following information which is considered as semantic elements in web pages is extracted using proposed "semantic elements extracting algorithm":

♦ Concepts: Named entities about every parts of human body,
♦ Names of Diseases: words or phrase of disease names,
♦ Descriptions: any kind of words or phrases that describe entities above. The word "description" now refers to any words that semantically relate to entities.
♦ Pairs of Concept and Description: any combination of pair Concept and Description. This is the truth because some concepts always go with only some description. And this pair makes a full meaning.

They proposed "New semantic elements learning algorithm" that tries to find out new semantic elements from web pages and render a suggestion list which is then used as a supported materials for domain users to upgrade ontology.

It was proposed [8] an approach to integrate XML sources and handle queries in the integrated system by using a bidirectional query translation algorithm. We choose to use the global-as-view (GaV) approach for the integration of XML sources. For the sake of simplicity, we use RDFS to model the ontologies (instead of DAML+OIL or OWL). This choice limits the expressiveness of the model, so that some axioms (e.g., disjunction of two concepts) cannot be used. They transform the heterogeneous XML sources into local RDF-based ontologies (defined using the RDFS space), which are then merged into the RDF global ontology. This transformation process encodes the mapping information between each concept in the local RDF ontology and the path to the corresponding element in the XML source.

In [2], an agent based intelligent sequence search engine is proposed, which implemented XML, XSL & XQuery to facilitate easy information exchange and sequence management [9]. Stardardised design methodology suitable for intelligent agent systems was used for designing the proposed system [10]. Intelligent monitoring of the biological data sources was implemented guarantying retrieval of sequence from the fastest data source. To deal with minimal storage and faster exchange of huge sequence information a sequence compression technology was used. Hong Su Elke and others [11] focus on SQO specific to XML stream processing. The distinguishing feature of pattern retrieval on XML streams is that it solely relies on the token-by-token sequential traversal. There is no way to jump to a certain portion of the stream. We however can use schema constraints to expedite such traversal by skipping computations that do not contribute to the final result. Ontology as a data model has been widely used in many fields in computer sciences. In information extraction field ontology-based IE method is considered as an advanced technique compared to traditional ones due to the high achievement results. Taking some existing IE system into considerations, we much focus on two main factors: semantic elements extraction and ontology enhancement performance.

About semantic extraction, M. Abulaish at el [12] introduced a method to extract relation between entities in biological texts. The document is applied with part of speech tagger to assign parts of speech to different works and then to be converted to
binary structure taking all words that is tagged verbs as a key node. The system can extract entities and relation between them (El V E2) by recording all occurrences of combination and make a fuzzy biological relations.
Then extracted relations are to be used to enhance the ontology. Enhanced ontology increases the effectiveness of information extraction process and also guides the user to pose queries in a more focused way.

The same method is introduced in [13] by L.Dey et al. they propose a system that can extract imprecise descriptions in wine domain. Wine has color, taste, flavor properties but sometime there are some descriptive words along with properties to describe more about attributes such as level of taste, flavor. In order to extract imprecise information, the system n tries to.

Services Available on Three Current Health Care Web Sites

A short overview follows of the services currently available on three health care sites with divergent content.

www.Health24.com (June 2008)

The aim of the Health24.com site is to inform and educate visitors on health related matters and help them to find health care service providers. It is a South African site with a majority stake held by Media24 (a Naspers Group publishing company).

Funding is mainly through online advertising by medical service and product providers, banks, and retail magazines. They go to great lengths to ensure visitors that their content is not determined by commercial interests. The content of the site is continually updated by a team of medical specialists and journalists. It covers a wide range of topics, namely, common medical conditions, diet, fitness, pregnancy, parenting, sex, oral health, mind health, and family health for men, woman, children, teens, and pets. The site offers narratives of personal experiences, research articles, graphics and videos, forums, ask an expert (with over 40 experts), quizzes, interactive health tools, competitions, and surveys. Users can subscribe to daily tips and a newsletter.

There are also links to medical aid schemes, gyms, and other medical services. They collect information about a visitor to enable them to provide information regarding a product, service, or event. They also allow the syndication of content by other companies on their own websites, newsletters and intranets.
Health24.com claim to have over half a million visitors of which nearly 75% are in the age group 25 to 49 and 70% are woman. Just over half are parents. A user survey conducted in 2006 indicated that 77% of the Health24.com readers consult the Internet before visiting a health professional.

www.COJJOworkshops.co.za (June 2008)

COJJOworkshops is a South African site with a subsection that targets specific health problems. It offers information, education, and support in the form of online and mobile workshops. The online services are integrated with other related services, especially from COJJOhealth.co.za and COJJOaware.co.za. In the workshops the uniqueness of each participant is recognized in terms of personality, culture, and circumstances in life and, therefore, the problems that they have. The value for participants lies in accurate and timely information sharing, teamwork, facilitation, and coaching in groups that includes the patient, family/caregivers, professional health care providers, health service providers, medical schemes, manufacturers, and employers. Organisations, pharmacies, clinics, hospitals, or medical practices can request the development of new workshops with focused content for a service, setting or local community. Typical workshops are "Living with chronic Health Problems", "Building your Health Care Team", and "Managing Health Care Costs".

The related site COJJOhealth has a coaching service that allows the user to ask, via e-mail, telephone or other means, medically related questions that are based on the information provided. Full subscribers have access to a secure patient and healthcare team portal. The content is aggregated from various sources and includes a daily news service. In depth information focuses on diabetes, weight management, diet, exercise, medication, smoking cessation, and complications. Their target audience is referrals from health care practitioners, service providers, employers, strategic partnerships, local associations, advertising, and health care campaigns. Advertising is customised content related. Disclosure of information is only to subsidiaries and associated companies, or to entities, to improve the service. Aggregated information can be made available.

www.google.com/health (June 2008)

Google/health is a recent site that adds another dimension to personal health care. After signing up for free, it allows you to not only search for disease related information, but also to store and manage all your health information in one place. Members enter their profile in terms of conditions, medications, and allergies. The information can be imported from participating doctors, hospitals, labs, and pharmacies and is available 24/7 to a doctor during a consultation, or to hospitals in cases of emergency. With all information in one place interactions between medications, allergies and conditions can be traced, a second opinion can be asked for, and prescriptions can be refilled online. Google has no financial relationships with any company and it is free for the user to connect and share their health information with who they want. Google guarantees security and privacy.

Unfortunately it is limited to patients residing in the USA. Google/health also provides information on ailments such as symptoms, treatment, causes, test and diagnosis, prognosis, prevention, complications, and when to contact a doctor. Illustrations are provided as well as a list of links to web pages with related news, scholar results, related groups, and related search trends. Medical practitioners and hospitals can be searched for by specialty or location with the help of Google maps[14]. In last work [5], it was concentrated mainly on the diagnostic process within the diagnostic therapeutic cycle. The diagnosis can be defined as the "description of a health problem in terms of known diseases" and the diagnostic process as a "set of actions needed to obtain the diagnosis". We propose a methodology based on the semantic web technologies for handling the medical data within the diagnostic process.

In this paper [15], we propose an approach to integrate XML sources and handle queries in the integrated system by using a bidirectional query translation algorithm. We choose to use the global-as-view (Ga V) approach for the integration of XML sources. For the sake of simplicity, we use RDFS to model the ontologies (instead of DAML+OIL or OWL). This choice limits the expressiveness of the model, so that some axioms (e.g., disjunction of two concepts) cannot be used.

II. Knowledge Base and Domain Ontology

1. Knowledge Base

The core of our proposed system is its knowledge base, which encapsulates the human expertise. In practice, the system is only as good as its knowledge base. Hence, the first stage in constructing the current expert system was knowledge acquisition, in which relevant information relating to our domain was extracted, refined and structured so that it could be used in the reasoning process. Our knowledge contains domain facts and the rules. Our knowledge is represented in XML format.

After the step of the domain identification and knowledge acquiring from a participating expert of Jaundice, a model for representing the knowledge must be developed. Numerous techniques for handling information in the knowledge-base are available; however, most systems utilize rule-based approaches. The knowledge engineer, working with the expert, must try to define the possible best structure. Other commonly used approaches include decision trees, blackboard systems and object oriented programming.

Knowledge representation has been defined as "A set of syntactic and semantic conventions that make it possible to describe things. The syntax of a representation specifies a set of rules for combining symbols to form expressions in the representation language. The semantics of a representation specify how expressions so constructed should be interpreted (i.e. how meaning can be derived from a form). In the proposed system, the knowledge representation methodology uses XML format. Where, two elements of knowledge, facts and model rules are represented using XML format. The overall knowledge structure is shown in figure 1.
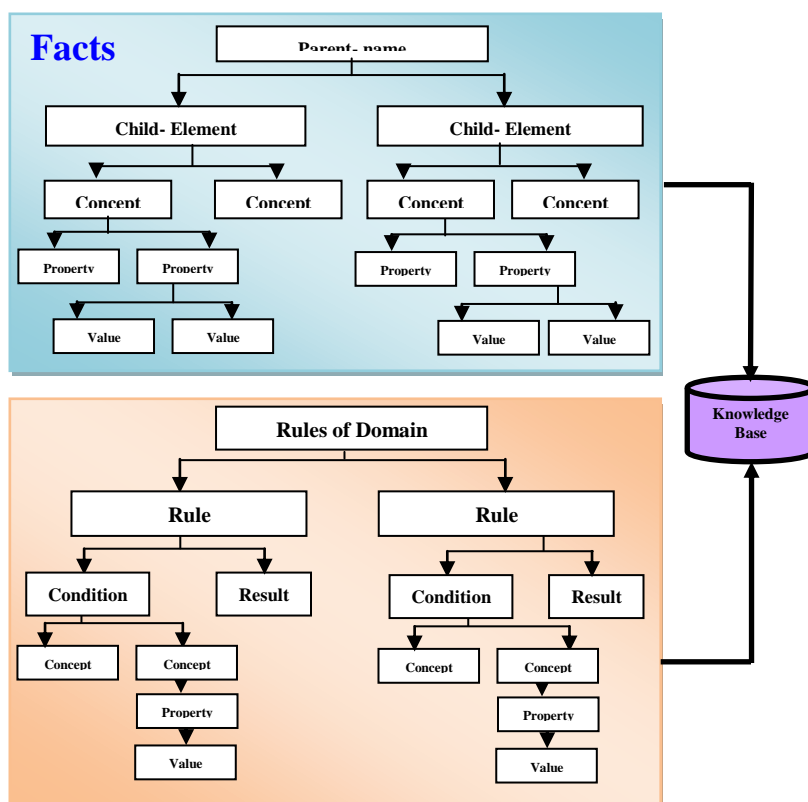

Figure1: Overall knowledge structure

The sample of the developed facts in our knowledge is shown in figure 2. The sample of the developed rules is shown in figure 3.

```
<DiagTestVal>
        <ConceptVal Name="Total Bilirubin"/>
                <ValueVal Val="Normal / Increased"/>
                <ValueVal Val="Increased"/>
        </ConceptVal>
        <ConceptVal Name="Conjugated Bilirubin"/>
                <ValueVal Val="Increased"/>
                <ValueVal Val="Normal"/>
        </ConceptVal>
        <ConceptVal Name="Unconjugated Bilirubin"/>
                <ValueVal Val="Increased"/>
                <ValueVal Val="Normal / Increased"/>
                <ValueVal Val="Normal"/>
        </ConceptVal>
        .
        .
</DiagTestVal>
```

Figure 2: Sample of developed facts in our knowledge

In our domain knowledge, the facts is represented as shown in figure 2 where the concept "Total Bilirubin" is a one of the diagnostic test for "Jaundice" and the concept has possible values are " Normal / Increased " and " Increased ". The property of each concept here is default as "Value".

```
<DiagConcept>
<"ResultConcept Name="Prehepatic" NoTrueFinding="1">
</"TestConcept Cpt="Total Bilirubin" Val="Normal / Increased>
</"TestConcept Cpt="Conjugated Bilirubin" Val="Increased>
</"TestConcept Cpt="Unconjugated Bilirubin" Val="Increased>
</"TestConcept Cpt="Urobilinogen" Val="Increased>
</"(TestConcept Cpt="Urine Color" Val="Normal (urobilinogen>
</"TestConcept Cpt="Stool Color" Val="Normal>
</"TestConcept Cpt="Alkaline Phosphatase Levels" Val="Normal>
</"TestConcept Cpt="Alkaline Transferase and Aspartate Transferase Levels" Val="Normal>
</"TestConcept Cpt="Conjugated Bilirubin in Urine" Val="Not Present>
<ResultConcept/>
<ResultConcept Name="Hepatic" NoTrueFinding="4">
</"TestConcept Cpt="Total Bilirubin" Val="Increased>
</"TestConcept Cpt="Conjugated Bilirubin" Val="Normal>
</"TestConcept Cpt="Unconjugated Bilirubin" Val="Normal / Increased>
</"TestConcept Cpt="Urobilinogen" Val="Normal / Increased>
</"(TestConcept Cpt="Urine Color" Val="Dark (urobilinogen+conjugated bilirubin>
</"TestConcept Cpt="Stool Color" Val="Normal>
</"TestConcept Cpt="Alkaline Phosphatase Levels" Val="Increased>
</"TestConcept Cpt="Alkaline Transferase and Aspartate Transferase Levels" Val="Increased>
</"TestConcept Cpt="Conjugated Bilirubin in Urine" Val="Present>
<ResultConcept/>
.
.
.
```

Figure 3: Sample of developed Rules in our knowledge

The knowledge can be formulated as shown in the following simple statements: IF the 'traffic light' is green THEN the action is go, as for example: IF the 'traffic light' is red THEN the action is stop. These statements represented in the IF- THEN form are called production rules or just rules. The term 'rule' in artificial intelligence, which is the most commonly type of knowledge representation, can be defined as IF-THEN structure that relates given information or facts in the IF part to some action in the THEN part. A rule provides some description of how to solve a problem. Rules are relatively easy to create and understand. Any rule consists of two parts: the IF part, called the antecedent (premise or condition) and the THEN part called the consequent (conclusion or action). The basic syntax of a rule is: IF <antecedent> THEN <consequent> . The rules in XML format have a different structure with the previous meaning but in different format. Sample of rule built in the proposed system is shown in Figure 3; it can be interpreted as following:

1. <DiagConcept>; represents the root in the domain of the Jaundice.
2. The node of "*ResultConcept*" represents a rule consequent and has attribute "*Name*" its value takes the consequent as " *Prehepatic* ".
3. The child nodes "TestConcept" represent the decision rule for each part in Jaundice diagnosis that has two attributes are " *Cpt* ", and "*Val*". For example the antecedent of rule is " *Total Bilirubin = Normal / Increased* " .
4. The attribute "*NoTrueFinding*" represents the number of rule antecedent selected.

2. Domain Ontology

The term "Ontology" [16] is becoming frequently used in many contexts of database and artificial intelligence researches. However, there is not a unique definition of what an ontology is [17,18]. An initial definition was given by Tom Gruber: "*an ontology is an explicit specification of a conceptualization*" [17]. However, this definition is general and remains still unsatisfied for many researchers. In [19] Nicola Guarino argues that the notion of "conceptualization" is badly used in the definition. We note that many real-world ontologies already combine data instances and concepts [20]. The definition [16] differs from this point of view as we show later . Informally, it was defined an ontology as an intentional description of what is known about the essence of the entities in a particular domain of interest using abstractions, also called concepts and the relationships among them. Ontologies [21] are designed for being used in applications that need to process the content of information, as well as, to reason about it, instead of just presenting information to humans. They permit greater machine interpretability of content than that supported by XML, and OWL, by providing additional vocabulary along with a formal semantics. Because of the intrinsic complexity of the concepts involved, the medical domain is one of the most active ones in defining and using ontologies.
The ontology in our system is focused in our medical domain that is "Jaundice diseases". Jaundice [22], is a yellowing of the skin, conjunctiva (clear covering over the sclera, or whites of the eyes) and mucous membranes caused by increased levels of bilirubin in the human body. When red blood cells die, the heme in their hemoglobin is converted to bilirubin in the spleen and in the hepatocytes in the liver. The bilirubin is processed by the liver, enters bile and is eventually excreted through feces. Consequently, there are three different classes of causes for jaundice. Pre-hepatic or hemolytic causes, where too many red blood cells are broken down, hepatic causes where the processing of bilirubin in the liver does not function correctly, and post-hepatic or extrahepatic causes, where the removal of bile is disturbed.

Figure 4 shows part of our ontology for "Jaundice diseases". Jaundice diseases [23] are divided into three types, these type is shown as follows:
Prehepatic: Pre-hepatic jaundice is caused by anything which causes an increased rate of hemolysis (breakdown of red blood cells).
Hepatic: Hepatic (in hepatocellular jaundice there is invariably cholestasis) jaundice causes include acute hepatitis, hepatotoxicity, Gilbert's syndrome.
Post-hepatic: Post-hepatic jaundice, also called obstructive jaundice, is caused by an interruption to the drainage of bile in the biliary system. The most common causes are gallstones in the common bile duct, and pancreatic cancer in the head of the pancreas.
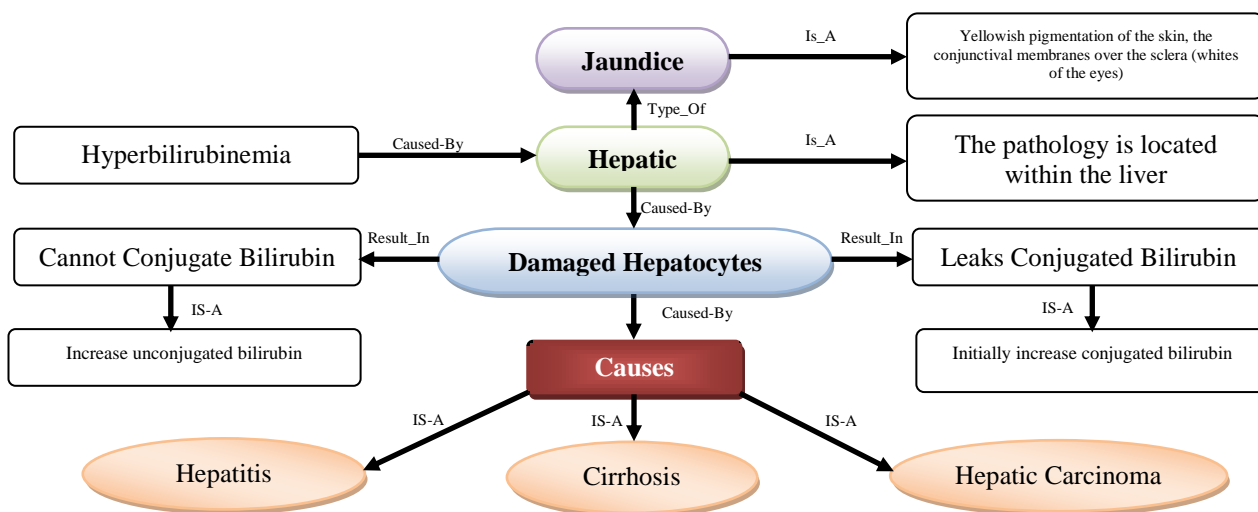

Figure 4: Part of domain ontology

This part of domain ontology shows these points:

- "Jaundice" is a yellowish pigmentation of the skin, the conjunctival membranes over the sclera (whites of the eyes).

- Hepatic is type of Jaundice.

- Hepatic is the pathology is located within the liver.

- Hepatic is caused by damaged hepatocytes.

- Hyperbilirubinemia is caused by Hepatic.

- Damaged Hepatocytes results in leaks conjugated bilirubin and cannot Conjugate bilirubin.

- Leaks conjugated bilirubin is a Initially increasing of conjugated bilirubin.

- Cannot Conjugate bilirubin is a Increasing of unconjugated bilirubin.

- Damaged hepatocytes caused by one of diseases: hepatitis, cirrhosis, or hepatic carcinoma

Our domain ontology is represented in Web Ontology Language (OWL) [24]. The Web Ontology Language (OWL) [25] describes classes, properties, and relations among these conceptual objects in a way that facilitates machine interpretability of Web content. OWL is the result of the Web Ontology Working Group (now closed) and descends from DAML+Oil, which is in turn an amalgamation of DAML and OIL.

OWL is defined as a vocabulary, just as are RDF and RDF Schema, but it has a richer semantics. Hence, an ontology in OWL is a collection of RDF triples, which uses such vocabulary. The definition of OWL [25] is organized as three increasingly expressive sublanguages:

OWL Lite offers hierarchies of classes and properties, and simple constraints with enough expressive power to model thesauri and simple ontologies. However, it imposes limitations on how classes are related to each other.

OWL DL increases expressiveness and yet retains decidability of the classification problem. OWL DL offers all OWL constructs, under certain limitations

OWL Full is the complete language, without limitations, but it ignores decidability issues.

The major points [25] behind the requirements that the W3C Consortium defined for ontology description languages for the Semantic Web are sown in table 1.

Table 1: The major points behind the requirements

Items	Details
The design of the language should be compatible with XML, in the sense that	- An ontology should have an XML serialization syntax. - An ontology should use the XML Schema data types, where applicable
The design of the language should follow description logic, in the sense that	- The language should be based on the notions of concept (or class), role (or property), and individual. - The language should support expressions thereof
The design of the language should support the definition of ontology vocabularies, in the sense that	- An ontology should be identified by a URI reference. - The classes, properties, and individuals of an ontology should be identified by URI references.
The design of the language should facilitate	- The development of ontologies in a distributed fashion. - The definition of different versions of the same ontology. - The reuse of previously defined ontologies

Our domain otology is implemented by protégé tool [26,27]. The part of owl code that is extracted from protégé is shown in figure 5, and figure 6 shows the extracted graph of entire domain ontology.

```
<"owl:Class rdf:ID="Intrahepatic>
<rdfs:subClassOf>
</ "owl:Class rdf:ID="Obstructive>
<rdfs:subClassOf/>
<owl:Class/>
<"owl:Class rdf:ID="Hemolytic_Anemia>
<rdfs:subClassOf>
</ "owl:Class rdf:ID="Hemolytic_Causes>
<rdfs:subClassOf/>
<owl:Class/>
<"owl:Class rdf:ID="Drugs>
<rdfs:subClassOf>
</ "owl:Class rdf:ID="Compression_Causes>
<rdfs:subClassOf/>
<owl:Class/>
<"owl:Class rdf:ID="Compression>
<rdfs:subClassOf>
</ "owl:Class rdf:ID="Cholestasis>
<rdfs:subClassOf/>
<owl:Class/>
<"owl:Class rdf:ID="Destruction_Of_Erythrocyte>
<rdfs:subClassOf>
</ "owl:Class rdf:ID="Hemolytic>
<rdfs:subClassOf/>
<owl:Class/>
.
.
```

Figure 5: Part of OWL code

Figure 6: Domain ontology graph

III. System Architecture

The Semantic Web provides a mechanism for adding meaning to data, essential in healthcare when the same clinical information can have many different representations. Combining the Semantic Web and agent in relation to the healthcare domain results in semantic Web services for healthcare, which will ultimately enable the automated interpretation of clinical data. Semantic Web for healthcare open new possibilities in knowledge management, clinical decision support, and application integration. Semantic Web have the potential to support an advanced healthcare environment, offering new applications and services to health networks. Primary issues of concern are data security and confidentiality.

Our system architecture is based on two basic elements that are semantic web technology and agents. The architecture is shown in figure 7.
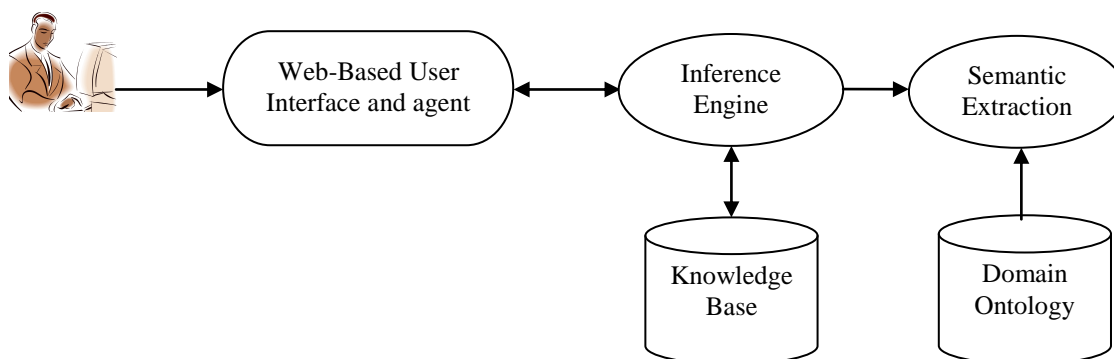

Figure 7: System Architecture

The proposed methodology envisages an integration of many data resources, containing the necessary knowledge. In this manner the integrated information resources can be seen as an inter-connected database. To allow an easy access to integrated knowledge resources, our technical solution enables one to use well-known queries, like SQL or XQuery, to access this integrated data universally.

The whole query processing [28] consists of 5 steps: (1) label the xml document, (2) perform pattern matching of the entire twig over the label stream, (3) extract the output part from each label match of the entire twig, (4) map the label of the output part to the output value in xml document, (5) post-process job such as redundant result elimination and result grouping.

There are three benefits of using semantic information in query processing:

First, by analyzing the semantics of both the query and XML document, we can avoid unnecessary computation on redundant data in XML document smartly. Data redundancy is a very common and inevitable issue in XML document, because XML provides the ease for an ordinary user to generate XML data, but most of them may not have good schema design knowledge.

Second, by carefully analyzing the components in predicate part and output part of a twig query, it is possible to save unnecessary computation. In particular, if the predicate part involves a multi-valued attribute of an object, once the first match of this attribute is found, we can return the output value without finding the remaining matches, since all of them contribute to the same query output value.

Third, the semantic information can help return a more user-expected and meaningful query result. The query answer returned by strictly enforcing the semantics of XPath and XQuery may not be the expected one.
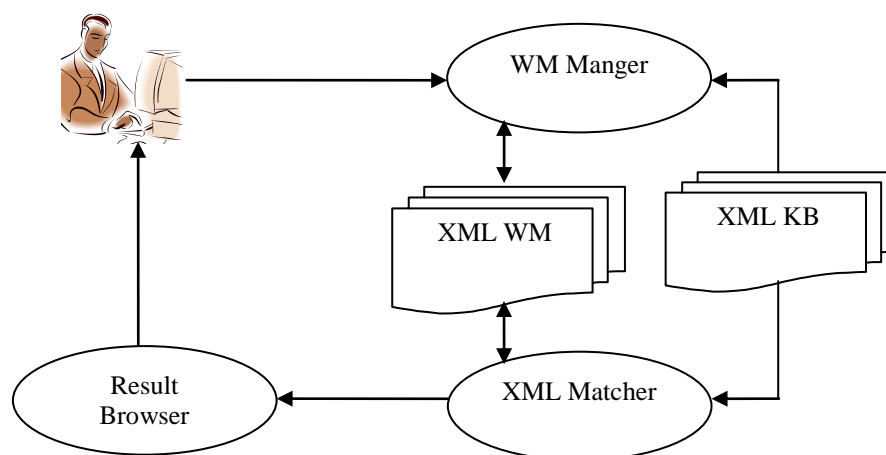
- **Inference Engine**

The entire control and operation of the system is done by the inference engine; that is developed using C#; which handles the knowledge in format of XML to get the result from the XML file (Knowledge base) that stores the knowledge rules. Figure 8 shows the block diagram of inference engine. The main roles of the inference engine are summarized as: It applies the expert domain knowledge to what is known about the present

situation to determine new information about the domain. The inference engine is the mechanism that connects the user inputs in the form of answers to the questions to the rules of knowledge base and further continues the session to come to conclusions. This process leads to the solution of the problem. The inference engine also identifies the rules of the knowledge base used to get decision from the system and also forms the decision tree. The inference was shown in [29].

The inference mechanism consists of three main components namely: working memory manager (WM manager), XML matcher, and result browser. Figure 8 depicts the main components of the inference mechanism, while Figure 9 illustrates the inference engine flowchart.

- *WM Manager Component*: The WM manager interacts with the user-friendly interface to get the concepts and its properties as well as the values of those properties b y using communication model. The user-friendly interface permits the user to edit his complaints easily. This complaint is considered as a user finding. When the user selects concept, property, and value to be entered in the working memory, the WM manager creates an XSL query statement, which represents these findings.

- *XML Matcher Component*: In this methodology the rule is succeeded when all its child nodes are existed in working memory. This is achieved when the attribute 'ExistInWM'of every child node is set to "Yes". So, the matcher gets those succeeded rules by comparing the value of the attribute 'NoTrueFindings' of every parent node and the number of the child nodes in this rule and select the matched one. The succeeded rules are store in the result store for later use by display result component.

- *Result Browser Component*: Result Browser component gets the value of the attribute disorder for every rule in the result store, and pass value to the user interface by using communication model to display it to the user.


```
RuleId = Get_RuleId(Finding)
For Each RuleId in RulesId
        NoTrueFindings=Get_NoTruFindings(RuleId)
        NoTrueFindings=NotrueFindings+1
        Set_NotTrueFindings(RuleId, NoTrueFindings)
Next RuleId
```

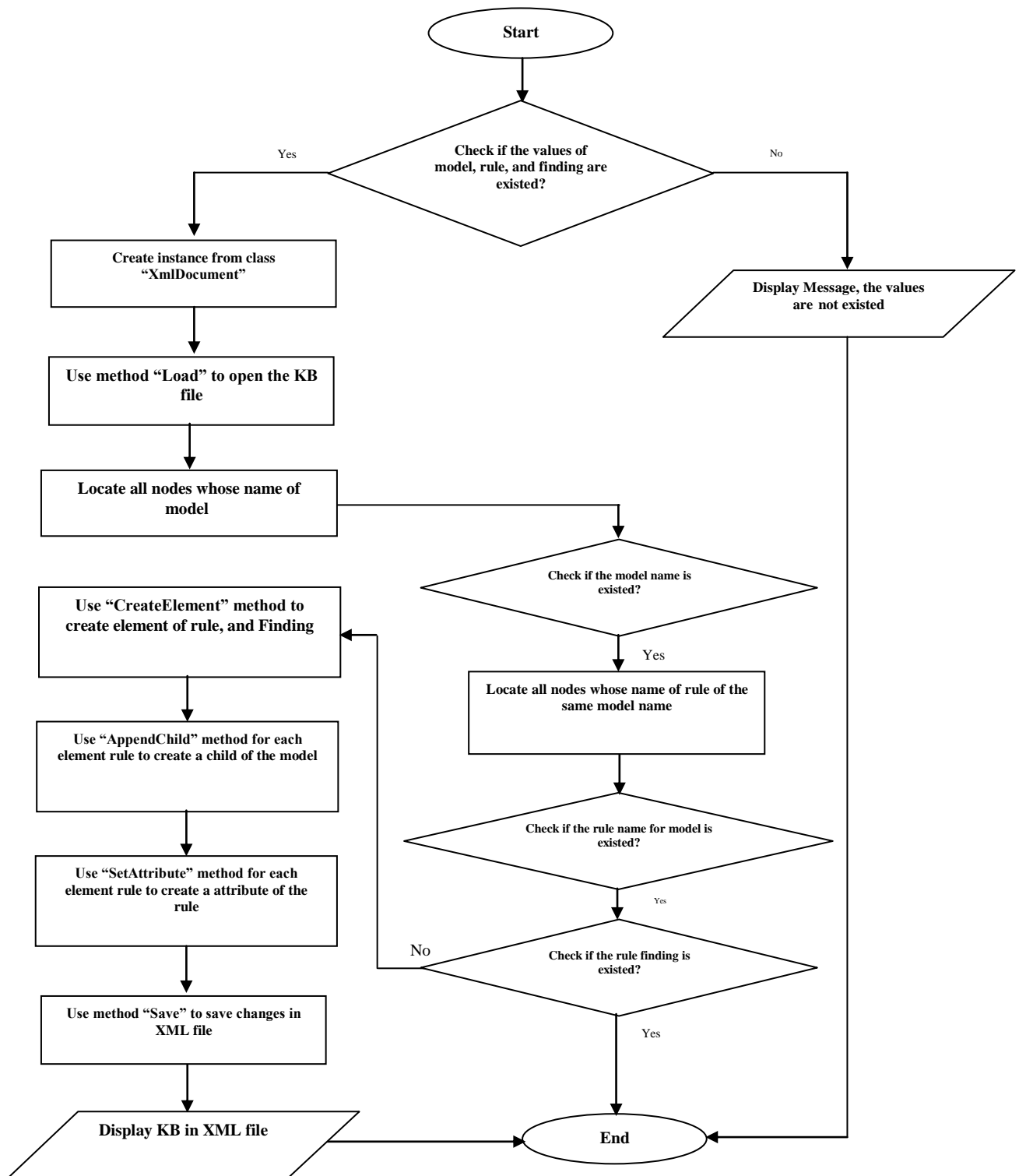**Figure 8:** Inference engine block diagram

Figure 9: Inference engine block diagram

- **Semantic Extraction**

This is a very important part of the system, in this part we use ontology to extract semantic elements from the selected data which is derived during system runtime. Semantics elements include Concepts (Cs) which are

actually Named entities of parts of body, Descriptions (annotation) of concepts (Ds), Names of diseases (Ns) and Pairs of Concept and Description (Ps). Description elements are any words or phrases that are used to describe concepts ignoring their part of speech. The Figure 9 shows the algorithm to extract semantic elements from the selected data by the system.
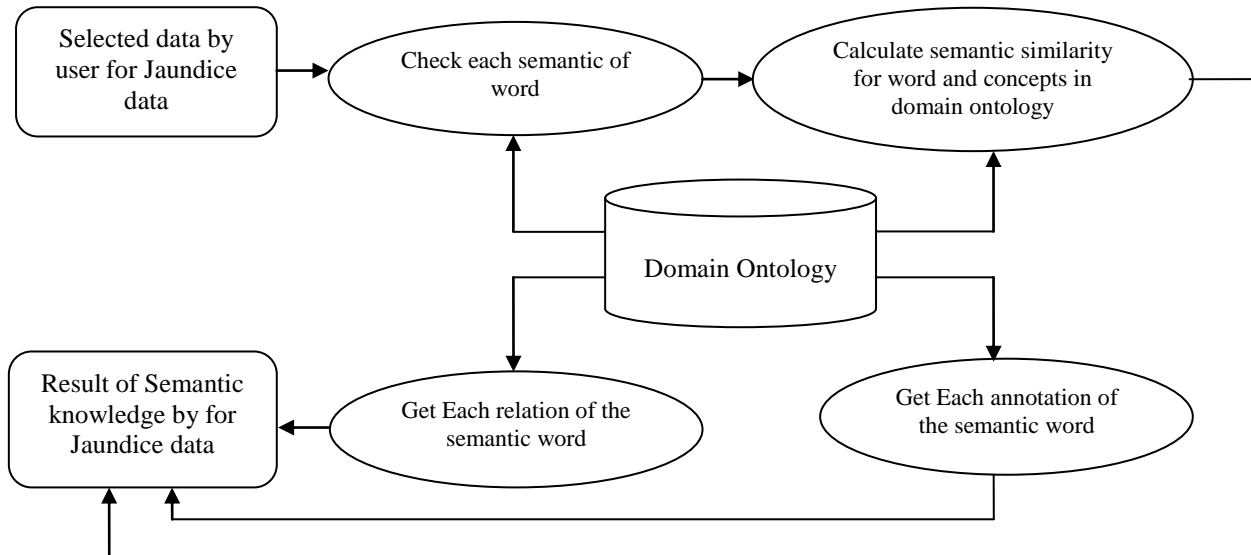

Figure 10: diagram of algorithm to extract semantic elements

IV. Results of Proposed System

The proposed system was evaluated with different users, including developers, and staff. The system is validated by experts in the domain of Jaundice causes and diseases. Tests of the system were carried out by the developers to make sure the system would work correctly as well as the system is web based system, another validation and evaluation for the system will be carried out through the using through the web and the feedbacks from the users will be considered for any comments and modifications. The system has six parts, the first includes the initial symptoms in our domain "Jaundice". The second part includes related findings for the initial symptoms. The third part includes the selected symptom of user. The fourth parts includes the result of our inference engine to extract the diseases class of Jaundice. The fifth part includes suspected diseases of the selected diseases class. The sixth part includes annotation data for the semantic word. Figures 10,11,12, and 13 shows the snapshots of our developed system.
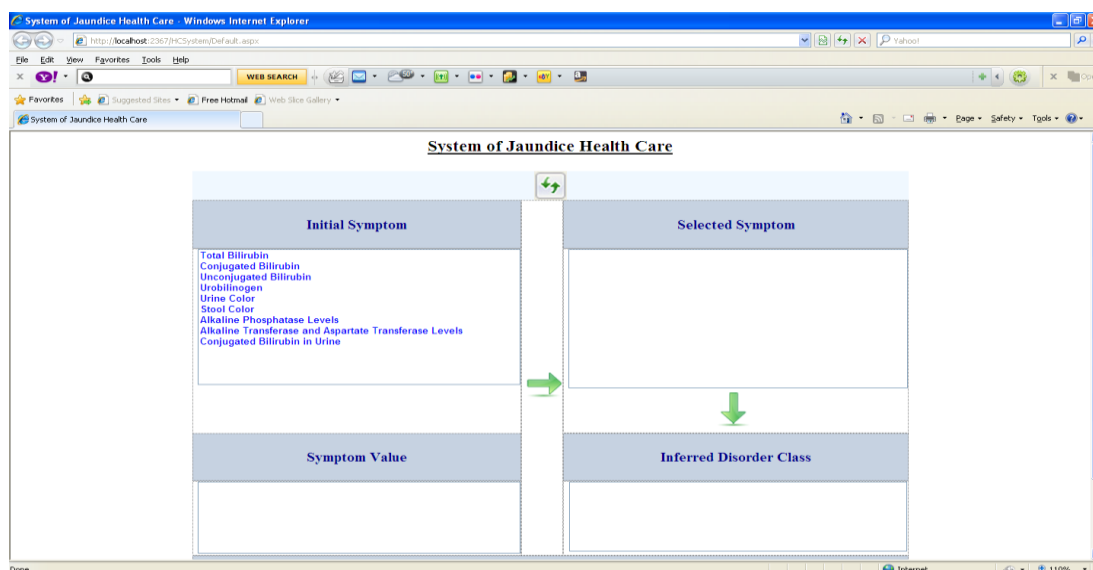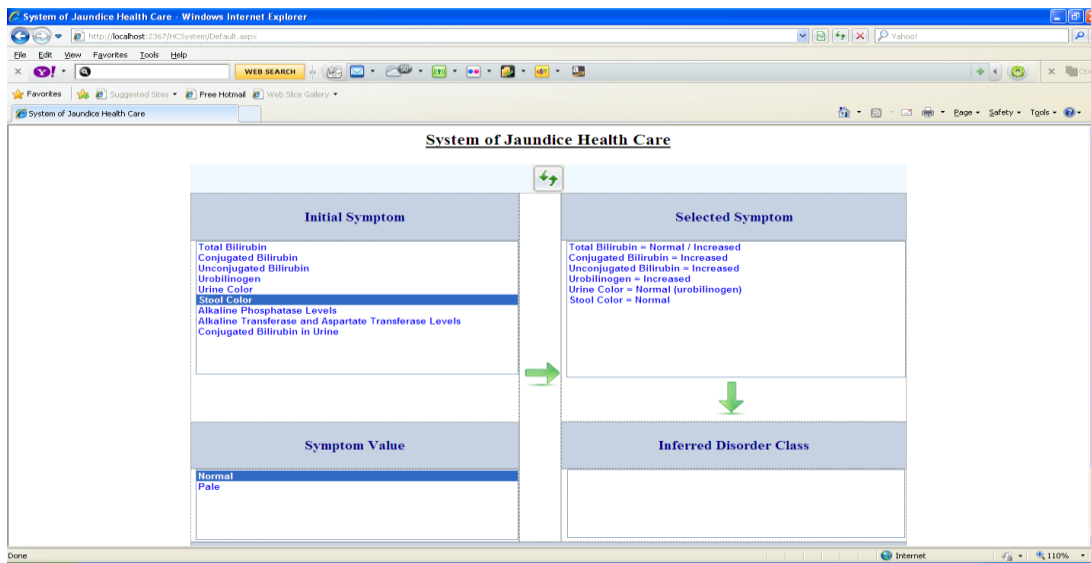

Figure 11: The interface of developed system

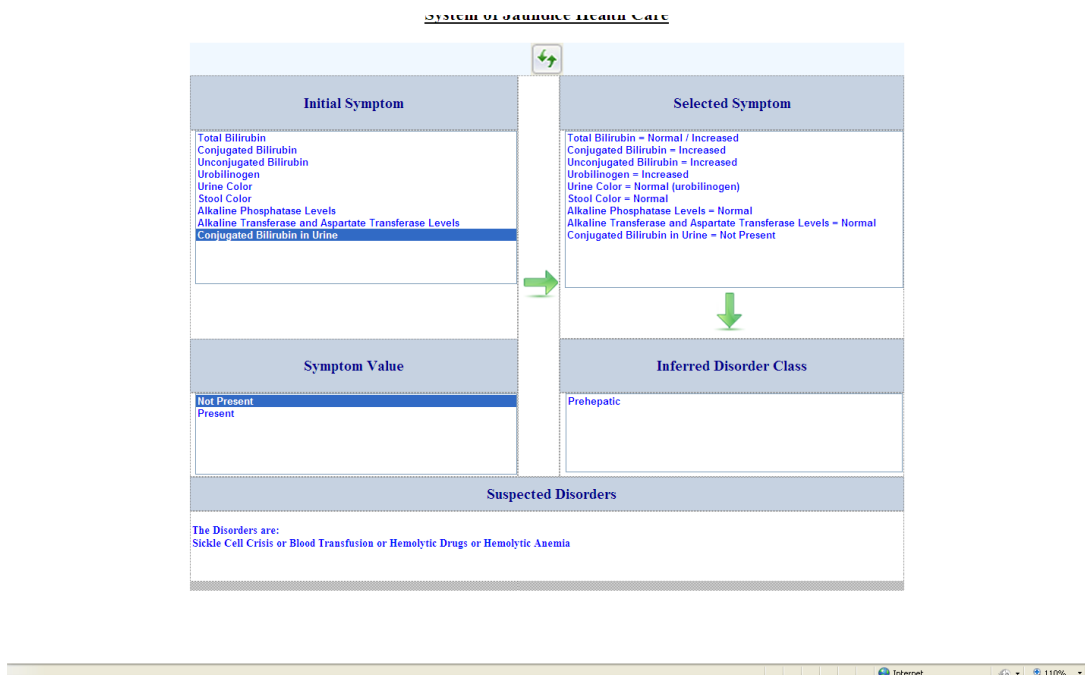**Figure 12:** The interface of developed system

Figure 13: The interface of developed system

V. Conclusion

The work presented in this paper tries to overcome the general lack of research in the area of web-based healthcare systems (WBHS) by combining the features of semantic web and agents. The paper addressed the issues associated with the analysis, design, development, and use of web-based healthcare system for the Jaundice diseases that give the user more knowledge in that domain and diagnoses tasks in Jaundice diseases . It

is the first time to develop such system; which is web based, with the new methodology supporting with semantic web and agent technology, the ontology, the knowledge representation.

The work considered 27 different symptoms findings, includes three types or diseases class of Jaundice, and includes 21 different disease for entire types of Jaundice. The system introduce its ability to modify or updating and the extending of the existing diseases because the knowledge and domain ontology is represented in format that can be shared and easy to update. The system was verified, and validated by different the developers as well to be usable in the real world.

The research provides benefits to the employees and the ability to solve the contradiction in the confused problems, as well as the providing for the suggested solutions. The developed system is fully implemented to run on the web using ASP.net techniques as the main programming language, and a new server-side technology, XML-Rule-based knowledge sources and the inference mechanisms were implemented using ASP.net.

References

[1] J.C. van, K. Griffiths. (2008), Advancing Health Care Management with the Semantic Web, DOI 10.1109/BROADCOM.2008.25, 978-0-7695-3453-4/08, IEEE.
[2] F. Sufi, I. Cosic and Q. Fang. (2006), Design and Implementation of An Intelligent Agent System for Genome Data Retrieval, 81-904262-1-4/06, Research Publishing Services, IEEE
[3] Gilbert Tekli, Richard Chbeir (2010) Towards an XML Adaptation/Alteration Control Framework. 2010 Fifth International Conference on Internet and Web Applications and Services. pp 248-255, IEEE.
[4] S. Fahim, C. Irena, F. Qiang. (2004), XQuery Based Intelligent Search System for Genomic Data Retrieval, 0-7803-8894-1/04. IEEE.
[5] P. Vili, P. Luka. (2007), Managing Diagnostic Process Data Using Semantic Web, Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07), 0-7695-2905-4/07, IEEE.
[6] S. Hong ,R. Elke A.,M. Murali. (2005), Semantic Query Optimization for XQuery over XML Streams, Proceedings of the 31st VLDB Conference, Trondheim, Norway, ACM.
[7] Fensel, D., & Bussler, C. (2009). An ontology, intelligent agent-based framework for the provision of semantic web services. Expert Systems with Applications 36 (2009) 3167–3187
[8] D. Tran, K. Wataru. (2007), A Proposal of Ontology-based Health Care Information Extraction System: VnHIES, 1-4244-0695-1/07, IEEE.
[9]] F. Sufi, I. Cosic and Q. Fang, (2004), "XQuery Based Approach of Intelligent Search Agent for Genomic Data", Int. Conf. on Intelligent Sensors, Sensor Networks and Information Processing, Melbourne, Australia.
[10] L. Padgham and M. Winikoff, (2004) Developing Intelligent agent systems: a practical guide, John Wiley & Sons: Chichester, UK. ISBN 0-470-86120-7 (HB),.
[11] S. Hong ,R. Elke A.,M. Murali. (2005), Semantic Query Optimization for XQuery over XML Streams, Proceedings of the 31st VLDB Conference, Trondheim, Norway, ACM.
[12] Muhammad Abulaish, Lipika Dey (2005), "Biological Ontology enhancement with Fuzzy Relation: A Text Mining Framework" in Proceedings o the 2005 IEEEIWICIACMInternational Conference on Web Intelligence WI'05.
[13] David Taniar, Johanna Wenny Rahayu, (2006), "Web Semantic Ontology" IDEA group publishing 2006 pp.189-226
[14] J.C. van Niekerk K. Griffiths, (2008), "Advancing Health Care Management with the Semantic Web" Third International Conference on Broadband Communications, Information Technology & Biomedical Applications
[15] Isabel F. Cruz, Huiyong Xiao, Feihong Hsu, (2004), An Ontology-based Framework for XML Semantic Integration Proceedings of the International Database Engineering and Applications Symposium (IDEAS'04) IEEE.
[16] S. Wei, L. Da-Xin. (2006), Using Ontologies for Semantic Query Optimization of XML Database, R. Nayak and M.J. Zaki (Eds.): KDXD 2006, LNCS 3915, pp. 64 – 73, 2006, Springer-Verlag Berlin Heidelberg.
[17] Gruber, T. (1993), A translation approach to portable ontology specifications. In: Knowledge Acquisition, 5(2) ,199-220
[18] Chandrasekaran, B., Josephson, J., Benjamins, V. (1999), What are ontologies, and why do we need them? In: IEEE Intelligent Systems, 20-26.
[19] Guarino, N., Giaretta, P. (1995), Ontologies and knowledge bases: towards a terminological clarification. In: Knowledge Building Knowledge Sharing,ION Press. 25-32
[20] Noy, N., Hafner, C.D. (1997), The state of the art in ontology design. AI Magazine 3, 53-74

[21] V. Aida, G. Karina, S. David, B. Montserrat, (2010), Using ontologies for structuring organizational knowledge in Home Care assistance, international journal of medical informatics 7 9 (2 0 1 0) 370–387, doi:10.1016/j.ijmedinf.2010.01.012,Elsevier Ireland Ltd.

[22] http://www.medic8.com/healthguide/articles/jaundice.html.

[23] http://en.wikipedia.org/wiki/Jaundice.

[24] http://www.w3.org/TR/owl-ref/

[25] K.K. Breitman, M.A. Casanova and W. Truszkowski. (2007), Semantic Web: Concepts, Technologies and Applications , In chapter 5 "OWL" ISBN-10: 1-84628-581-X, Springer-Verlag London Limited.

[26] N. Natalya, S. Michael, D. Stefan, C. Monica, F. Ray W., and A. Mark. (2001), Creating Semantic Web Contents with Protégé-2000, 1094-7167/01, IEEE.

[27] H. Robert and W. Christine. (2007), A TOOL FOR DYNAMIC KNOWLEDGE MODELING, Proc. 6th IEEE Int. Conf. on Cognitive Informatics (ICCI'07), 1-4244-1327-31071, IEEE.

[28] B. Zhifeng, W. Huayu, and C. Bo. (2008), Using semantics in XML query processing, Proceedings of the 2nd international conference on Ubiquitous information management and communication, Suwon, Korea, ACM.

[29] B. Khaled, R. Mahmoud, El. Salwa, and T. Khaled. (2004), Knowledge Representation and Reasoning Using XML, 2nd international Artificial Intelligent, Cairo.

[30] Fensel, D., & Bussler, C. (2009). An ontology, intelligent agent-based framework for the provision of semantic web services. Expert Systems with Applications 36 (2009) 3167–3187