PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
UNIVERSITY MOHAMED BOUDIAF –of M'SILA

Faculty: Mathematics and Informatics
Department: Computer Sciences
N°: ……………………………..…

Domain: Mathematics and Informatics
Field: Computer Sciences
Sub-Field: Software Engineering

A Dissertation in Fulfillment

For the Requirements of the Degree of Master

By: DJENAOUI Abdelatif

SUBJECT

Creation of an Ontology from a Database To assist Medical Diagnosis

Defended publicly on: 07/06/2017

Board of Examiners:

Dr. BOURAHLA Mustapha	**University of M'sila**	**Chairman**
Dr. KADRI Said	**University of M'sila**	**Supervisor**
Ms. BOUZAAROURA Ahlam	**University of M'sila**	**Examiner**

Promotion: 2016/2017

PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

UNIVERSITY MOHAMED BOUDIAF –of M'SILA

Faculty: Mathematics and Informatics

Department: Computer Sciences

N°: ………………………..…

Domain: Mathematics and Informatics

Field: Computer Sciences

Sub-Field: Software Engineering

جامعة محمد بوضياف - المسيلة

Université Mohamed Boudiaf - M'sila

A Dissertation in Fulfillment

For the Requirements of the Degree of Master

By: DJENAOUI Abdelatif

SUBJECT

Creation of an Ontology from a Database To assist Medical Diagnosis

Defended publicly on: 07/06/2017

Board of Examiners:

Dr. BOURAHLA Mustapha	**University of M'sila**	**Chairman**
Dr. KADRI Said	**University of M'sila**	**Supervisor**
Ms. BOUZAAROURA Ahlam	**University of M'sila**	**Examiner**

Promotion: 2016/2017

To the soul of my father "عراس"

Table of Contents

List of Tables:

List of Figures:

List of Acronyms

A

B

C

D

Introduction:

Our life becomes more complicated day after day, because of their ramifications and fine details in all fields. It becomes a duty for human to study and keep abreast of these developments, in order to solve the problems associated with them, and to know their causes and diagnosis.

Diagnosis is the most important stage in the process of finding answers to the questions that we face. Where the precise identification of the causes of an accident or phenomenon is the first step in understanding, interpreting, solving or resolving the problem and avoid it in the future.

So, what does mean diagnosis?

In **Oxford Dictionary** Diagnosis is: « The identification of the nature of an illness or other problem by examination of the symptoms. ».

This term is strongly linked to the field of **medicine**, which pass through the identification of symptoms leading to knowledge of the disease, and it has a great importance because of its relation with human life, health and living conditions.

However, the fast development of technologies used in medicine has been accompanied by an evolution in the types of diseases and their many names and precise details, it became imperative for physicians to resort to computers to do some of the most difficult tasks, such as precision surgical procedures, and the interpretation of the results of complex analyzes and others.

The most important means used in this field are: Expert Systems, Intelligent Agents and Semantic web, «The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation » [1].

This last concept is based on “Ontology” which is « the branch of metaphysics dealing with the nature of being » [https://en.oxforddictionaries.com/definition/ontology], its form is: «A set of concepts and categories in a subject area or domain that shows their properties and the relations between them. ».

The subject of this research is: Creation & use of ontologies to help on medical diagnosis, our choice of this subject has been among several reasons, including the following:

- The rapid development of this field (Medical Diagnosis).
- The increasing needs to accelerate the researches in the field of semantic web and ontologies.
- Its great importance in our life (attached to the human health).
- Our desire to give an addition to this area.

So, our main problematic can be specified as follows:

Link the symptoms ontology with the diseases ontology to give physicians assistance in decision-making.

In the way of solving this problem, we have divided our manuscript into four chapters:

- Chapter One: entitled: "Medical Diagnosis", where we explained the terms in detail and gave the theoretical framework of the subject
- Chapter Two: entitled: "From Databases to Ontologies". In this chapter, we showed the differences between databases and ontologies in the storing of information.
- Chapter Three: entitled: "Medical Diagnosis Using Ontologies", in this chapter; we talked about anterior researches and the approaches used in medical diagnosis.
- Chapter Four: entitled: "Design & Implementation of '**EasyDiag'** System", this chapter is the result of our work, which is the development of a new system called '**EasyDiag'**, and the explanation of our approach.
- Conclusion and suggestions.

Chapter I

Medical Diagnosis

1. Preface:

Since the early centuries, the diagnosis is known in many civilizations as ancient Egypt and Greece, but its concept had evolved over the time, until it became in the current state of progress.

2. Diagnosis:

2.1. Definition:

This term is used largely over the world, in many disciplines:

- Biology: "scientific determination; a description that classifies a group or taxon precisely."
- Medicine: "the process of determining by examination the nature and circumstances of a diseased condition".
- In General: "a determining or analysis of the cause or nature of a problem or situation" [2].

And Oxford Dictionary defines diagnosis: "The distinctive characterization in precise terms of a genus, species, or phenomenon". [3]

Cambridge Dictionary: "a judgment about what a particular illness or problem is, made after examining it". [4]

So in global **the diagnosis is**: "Identification of a condition, disease, disorder, or problem by systematic analysis of the background or history, examination of the signs or symptoms, evaluation of the research or test results, and investigation of the assumed or probable causes. Effective prognosis is not possible without effective diagnosis".

2.2. Difference between Diagnosis & Diagnostic:

Diagnostic is the diagnosis procedure, which consists on repeatedly gathering new information about the case to be diagnosed, so that the physician can narrow down the list of possible diseases that the patient suffers from. In this part of the diagnosis the physician must work with several health care elements as the available means of assessment of the patient condition, the signs and symptoms, the feasible diseases, syndromes and social issues, as well as with the health care interventions.

Once the diagnostic procedure arrives to an acceptable evidence of the patient ailments, the physician determines the condition which constitutes the diagnosis and saves it as a part of the patient's medical record. [5]

3. Medical Diagnosis:

"The identification of the nature of an illness or other problem by examination of the symptoms" [3].

3.1. Definition:

Medical diagnosis is defined as the identification of a disease by investigating a patient's signs, symptoms and history, which provides a solid basis for the treatment and prognosis of the individual patient. [6]

3.2. Types: [7]

3.2.1. Clinical diagnosis:

Diagnosis based on signs, symptoms, and laboratory findings during life.

3.2.2. Differential diagnosis:

The determination of which one of several diseases may be producing the symptoms.

3.2.3. Medical diagnosis:

Based on information from sources such as findings from a physical examination, interview with the patient or family or both, medical history of the patient and family, and clinical findings as reported by laboratory tests and radiologic studies.

3.2.4. Physical diagnosis:

Diagnosis based on information obtained by inspection, palpation, percussion and auscultation.

3.3. Good diagnostic practice [8]:

3.3.1. Clinical assessment :

History and physical examination should be carried out by medically qualified personnel using appropriate diagnostic instruments such as stethoscope, otoscope and sphygmomanometer.

The clinician should identify the patient, understand the major complaint and ask questions in order to explore the underlying cause of the complaint and make a clinical diagnosis. A major concern is to obtain the correct history of the patient and to clearly understand the complaints and their duration. Information about the duration of symptoms is often unreliable. If pain or other symptoms have persisted for days, months or years it may be difficult to obtain an accurate history.

The physical examination should include a general examination and examination of the affected body system. The physical examination should take the privacy of the patient into consideration.

3.3.2. Diagnostic investigation:

There are three phases in the process of diagnostic investigation:

- The pre-analytical phase.
- The analytical phase.
- The post-analytical phase.

The pre-analytical phase comprises the time and all processes for the preparation of a patient for a diagnostic investigation to the moment when the investigation is made. The analytical phase comprises the time and all processes of a diagnostic investigation. The post-analytical phase comprises the time and all processes for reporting the results of the diagnostic investigation to the person who then undertakes the medical management of the patient.

3.3.3. Diagnostic imaging :

Contains several examinations: Ultrasound, Conventional X-ray, Computed tomography, Magnetic resonance imaging…etc.

3.3.4. Laboratory services and laboratory investigations:

A proper request for laboratory investigations necessitates a knowledge of the pathology of a disease as well as the technical aspects of laboratory analysis. Clinicians should know beforehand the value of the information that they can expect from a laboratory with respect to the clinical situation of the patient. Surprisingly, even in a large general hospital with an intensive care unit, it was found that 82% of all laboratory results were within normal range.

4. Approaches to clinical problem-solving [8]:

Clinical problem-solving is a process of finding out what is wrong with a patient, starting with the patient's presenting complaint. The stages of clinical problem-solving should be followed step by step, in order to arrive at the best possible diagnosis and to plan appropriate management.

The sequence of events used in making a diagnosis is:

- History-taking
- Physical examination
- Selection of laboratory tests and interpretation of results
- Use of diagnostic facilities, e.g. X-ray, ultrasound.

The first step to make a diagnosis of a disease is the exploration of the medical history, followed by the physical examination of the patient.

The second step is the ordering of diagnostic investigations.

4.1. Medical history-taking:

The medical history consists of:

- Complaints or symptoms presented by the patient
- Further information obtained from inquiry by the clinician.

The time needed to gather a medical history depends on the nature of the problem. The physician should aim to spend at least 10 minutes with every patient, but complicated problems may require longer.

A medical history proceeds as follows:

- history of present illness
- systemic inquiry (review of systems)
- past history
- social history
- family history
- Drug history.

When patients complain of pain, useful questions to delineate the problem are:

- The site (where is the pain?)
- Radiation (does it spread elsewhere?)
- Severity (how bad is it? Does it stop you carrying out normal activities?)
- Character (what is the pain like? Stabbing, burning, dull, etc?)
- Aggravating and relieving factors (what brings the pain on? What relieves the pain?)
- Duration of the pain (how long have you had it?).

4.2. General examination [8]:

Every patient must have a general examination performed including examination of the head, eyes (conjunctivae), mouth, neck, the skin and the extremities, all of which may provide clues to the cause of the patient's illness. The observations of the general examination should also be entered into the pre-printed record form. A complete general examination includes the following:

- General appearance (cleanliness, emaciation or obesity, marasmus, degree of illness).
- General mood, level of consciousness, ability to walk or sit (in a child), gait, posture.
- Obvious deformities.

- Vital signs: temperature, pulse rate, respiration rate, blood pressure.
- Body weight.

4.3. Respiratory system [8]:

The respiratory system is examined with the patient sitting up, or with a child sitting on its mother's lap. The patient should remove the shirt completely for the examination of the chest. Examine both the anterior and posterior parts of the chest; there is more posterior lung area available for examination than anterior.

- Inspection: shape and movement of the chest.
- Character of respirations.
- Palpation.
- Percussion: resonant (normal); hyper-resonant, dull, stony dull.
- Auscultation.

4.4. Cardiovascular system [8]:

The cardiovascular system is examined with the patient lying on his/her back on a couch, propped up at a 45° angle. A child can be sitting on its mother's lap.

4.5. Abdominal system [8]:

The abdomen is examined with the patient lying flat on his/her back on a couch, with arms at the sides and knees slightly flexed. Ask the patient to relax and breathe gently.

4.6. Genitourinary system [8]:

Wear examination gloves and have all instruments, a light or torch, and sterile swabs ready to hand.

4.7. Skin [8]:

Typical features on the skin very often provide the first information on a disease in a patient. The features may be caused by an infection (bacterial, viral, fungal, protozoan), metabolic disorder, malignant disease or exposure to harmful agents (allergens, toxins).

5. Examination of clinical symptoms and signs.

5.1. Symptoms:

5.1.1. Definition:

"any feeling of illness or physical or mental change that is caused by particular disease" [4].

5.1.2. Other Definition:

"A physical or mental feature which is regarded as indicating a condition of disease, particularly such a feature that is apparent to the patient" [3].

5.1.3. Medical Definition [7]:

Symptoms of greatest significance to the health care provider, establishing the identity o f the illness.

The symptoms shown in the temperature, pulse, and respiration.

- **Dissociation symptom** anesthesia to pain and to heat and cold, without impairme nt of tactile sensibility.
- **Objective symptom:** one perceptible to others than the patient, such as pallor, rapid pulse, rapid respiration, or restlessness.
- **Presenting symptom:** the symptom or group of symptoms about which the patie nt complains or from which he seeksrelief.
- **Signal symptom:** a sensation, aura, or other subjective experience indicative of an impending epileptic or other seizure.
- **Subjective symptom:** one perceptible only to the patient, such as pain, pruritus, or vertigo.

5.2. Medical symptoms types [9] :

There are three main types of symptoms:

- Chronic symptoms - long lasting or recurrent symptoms. These are often seen in diabetes, asthma, and cancer.

- Relapsing symptoms - symptoms which had occurred in the past, disappeared, and then come back. For instance in depression, multiple sclerosis, and also cancer.
- Remitting symptoms - when symptoms improve, and sometimes go away completely.

5.3. Signs [10]:

A medical sign is an objective feature indicating a medical fact or characteristic that is detected by a physician, nurse, or medical/laboratory device during the examination of a patient.

Sometimes, a sign may not be noticed by the patient, or not seem relevant to them, but it is meaningful for the physician. Below are some examples of specific signs:

- High blood pressure - this may indicate a cardiovascular problem, a reaction to medication, an allergy, as well as many other possible conditions or diseases.
- Clubbing of the fingers - this may be a sign of lung disease.

5.4. The observer defines whether it is a sign or symptom [10]:

Some say that it does not matter what the sign or symptom is, what matters is who observes it. For example, a rash could be a sign, symptom, or both:

- If the patient notices the rash, it is a symptom.
- If the doctor, nurse, or anyone other than the patient notices the rash, it is a sign.
- If both the patient and doctor notice the rash, it is both a sign and a symptom.

6. Differential Diagnosis [11]:

The differential diagnosis is the medical art of distinguishing one disease from another, looking alike. During the era of adulteration it has become difficult to give a set of symptoms, signs and investigative findings to exclude those diseases which simulate the actual disease of the patient, i.e. the diagnosis. All this depends on the clinical acumen—capability of the physician.

Knowledge of the differential diagnosis expedites the process of reaching a correct diagnosis thus avoiding lot of suffering to the patient.

Chapter 2:

From Databases to Ontologies

1. Preface:

Many organizations nowadays face the problem of accessing existing data sources by means of flexible mechanisms that are both powerful and efficient, those mechanisms have different formats and methods, and each one has its advantages and disadvantages.

The user must choose the convenient mechanism to satisfy his needs, according to some characteristics like: power, speed, protection, semantic…etc.

In this chapter, we will explain this characteristics for the two famous mechanisms of data storing: Databases & Ontologies.

2. Data versus Information [12]:

2.1. Data (especially computer data):

The presentation of facts, information or concepts which are created in a computer readable form or are translated into such a form.

2.2. Information:

Information is a useable answer to a concrete question. Something is information if a specific question is answered and that answer increases the understanding of the questioner and enables him to come closer to a specific objective.

Information has the following aspects:

- Structured and syntactic
- Semantic (as regards content)
- Pragmatic (relevant to applications)

2.3. Relationship between Data and Information :

The terms data and information are often used interchangeably and in the wrong context.

Therefore a list of distinguishing features is presented below:

- Semantic aspects of data are often coded. This codes need to be defined and interpreted after conventions previously agreed.

- Generally, information needs to be reconstructed or derived from data (e.g. the average rainfall of the month July over the last 10 years).
- Normally, data do not contain aspects relevant to applications (e.g. it is not possible to derive information for applications like tax, development, flood risk, etc. from the coordinates of a parcel of land).

3. Data Bases:

3.1. Definition [13]:

A database is an integrated collection of logically related records or files consolidated into a common pool that provides data for one or more multiple uses.

One way of classifying databases involves the type of content, for example: bibliographic, full-text, numeric, and image. Other classification methods start from examining database models or database architectures.

The data in a database is organized according to a database model. The relational model is the most common.

3.2. Types of DBs [13]:

3.2.1. <u>Relational Data Bases :</u>

A relational database matches data using common characteristics found within the data set. The resulting groups of data are organized and are much easier for people to understand.

For example, a data set containing all the real-estate transactions in a town can be grouped by the year the transaction occurred; or it can be grouped by the sale price of the transaction; or it can be grouped by the buyer's last name; and so on.

Such a grouping uses the relational model (a technical term for this is schema). Hence, such a database is called a "relational database."

The software used to do this grouping is called a relational database management system. The term "relational database" often refers to this type of software.

Relational databases are currently the predominant choice in storing financial records, manufacturing and logistical information, personnel data and much more.

Strictly, a relational database is a collection of relations (frequently called tables).

3.2.2. Data Warehouse :

A data warehouse stores data from current and previous years — data extracted from the various operational databases of an organization. It becomes the central source of data that has been screened, edited, standardized and integrated so that it can be used by managers and other end-user professionals throughout an organization.

3.2.3. Real-Time Data Base :

A real-time database is a processing system designed to handle workloads whose state may change constantly. This differs from traditional databases containing persistent data, mostly unaffected by time.

For example, a stock market changes rapidly and dynamically. Real-time processing means that a transaction is processed fast enough for the result to come back and be acted on right away. Real-time databases are useful for accounting, banking, law, medical records, and scientific data analysis. As computers increase in power and can store more data, real-time databases become integrated into society and are employed in many applications.

3.2.4. NoSQL Data Bases:

The next generation of database systems is known as NoSQL databases and document oriented databases.

NoSQL databases are often very fast, do not require fixed table schemas.

Examples of NoSQL systems: MongoDB and Oracle NoSQL Database.

3.3. Data Base Management System (DBMS) [13]:

A Database Management System (**DBMS**) consists of software that organizes the storage of data. A DBMS controls the creation, maintenance, and use of the database storage structures of organizations and of their end users. It allows organizations to place control of organization-wide database development in the hands of Database Administrators (DBAs) and other specialists. In large systems, a DBMS allows users and other software to store and retrieve data in a structured way.

There are Database Management Systems (DBMS), such as:

Microsoft SQL Server, Oracle, Sybase, dBase, Microsoft Access, MySQL from Sun Microsystems (Oracle), and DB2 from IBM…etc.

3.4. Characteristics of a DBMS [14]:

Traditionally, data was organized in file formats. DBMS was a new concept then, and all the research was done to make it overcome the deficiencies in traditional style of data management. A modern DBMS has the following characteristics:

- Real-world entity: A modern DBMS is more realistic and uses real-world entities to design its architecture.
- Relation-based tables: DBMS allows entities and relations among them to form tables. A user can understand the architecture of a database just by looking at the table names.
- Isolation of data and application: A database system is entirely different than its data. A database is an active entity, whereas data is said to be passive, on which the database works and organizes. DBMS also stores metadata, which is data about data, to ease its own process.
- Less redundancy: DBMS follows the rules of normalization, which splits a relation when any of its attributes is having redundancy in values.
- Consistency: Consistency is a state where every relation in a database remains consistent.
- Query Language: DBMS is equipped with query language, which makes it more efficient to retrieve and manipulate data.
- ACID Properties: DBMS follows the concepts of Atomicity, Consistency, Isolation, and Durability (normally shortened as ACID).
- Multiuser and Concurrent Access: DBMS supports multi-user environment and allows them to access and manipulate data in parallel
- Multiple views: DBMS offers multiple views for different users. A user who is in the Sales department will have a different view of database than a person working in the Production department.

- Security: Features like multiple views offer security to some extent where users are unable to access data of other users and departments.

3.5. Further Advantages of a DBMS [12]:

- Scalable: Describes the ability of a system to adapt to different amounts of data and numbers of users while maintaining the same performance. In a scalable system those can be increased through the extension of the computing capacity without any programming work.

- Expandability: New applications or other additions like new user interfaces can be realized without interfering with already working applications.

- Standardized and Combined Features: Features for data definition, data organization and data integrity can be standardized and combined.

- More Efficient System Operation and Development: It is possible to have more efficient features for maintaining system and further developing of the system.

- Access Control: Access to a database or parts of it (like single tables) can be controlled or restricted for each user but also for user groups.

- Recovery and Back-up: A database management system has features which allow to recover data sets, for example, after a failed transaction or a system crash. Additionally, a back-up from the whole system can be made to store in a save place.

- Quicker Application Development: As all the data handling is 'delegated' to the database management system it is possible to develop new applications quick and flexible.

3.6. Disadvantages of a DBMS [12]:

- Danger of a Overkill: For small and simple applications for single users a database system is often not advisable.

- Complexity: A database system creates additional complexity and requirements. The supply and operation of a database management system with several users and databases is quite costly and demanding.

- Qualified Personnel: The professional operation of a database system requires appropriately trained staff. Without a qualified database administrator nothing will work for long.

- Costs: Through the use of a database system new costs are generated for the system itself but also for additional hardware and the more complex handling of the system.

- Lower Efficiency: A database system is a multi-use software which is often less efficient than specialized software which is produced and optimized exactly for one problem.

4. Ontologies:

4.1. Definition [15]:

4.1.1. In Philosophy:

Fundamental branch of metaphysics – Studies "being" or "existence" and their basic categories – Aims to find out what entities and types of entities exist.

4.1.2. In Information Science

- An ontology is an engineering artefact consisting of:
    - A vocabulary used to describe (a particular view of) some domain.
    - An explicit specification of the intended meaning of the vocabulary. Often includes classification based information
    - Constraints capturing background knowledge about the domain.
- Ideally, an ontology should:
    - Capture a shared understanding of a domain of interest.
    - Provide a formal and machine able to be handled model.

4.2. Ontology Classifications [16]:

Several classifications of ontologies have been presented in the literature. Each of them focused on different dimensions in which ontologies can be classified. We will focus here on one of these classifications: ontologies based on the scope of the described objects. (Figure 1).

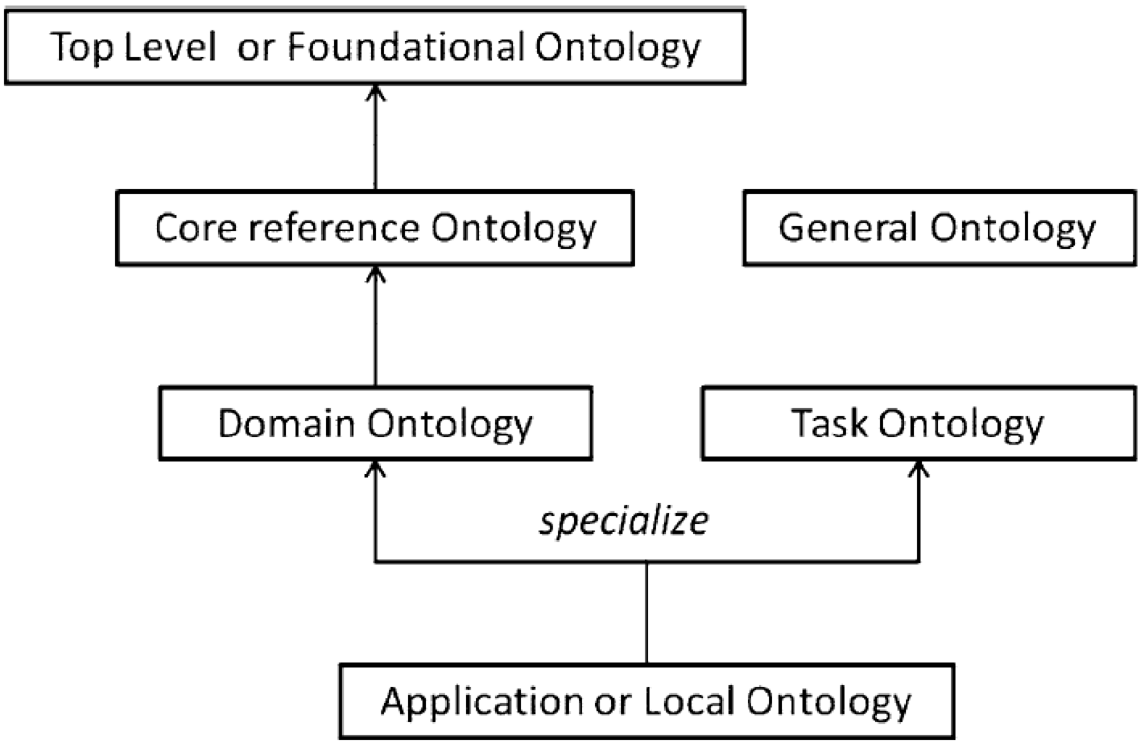


Figure 1: Ontology classification based on domain scope

4.2.1. Local Ontologies/Application Ontologies:

Local or application ontologies are specializations of domain ontologies where there could be no consensus or knowledge sharing. This type of ontology represents the particular model of a domain according to a single viewpoint of a user or a developer.

This kind of ontology is a combination of domain ontology and task ontology in order to fulfill the specific purpose of an application.

4.2.2. Task Ontologies :

The task ontology contains knowledge to achieve a task, on the other hand the domain ontology describes the knowledge where the task is applied.

4.2.3. Domain Ontologies :

Domain ontology is only applicable to a domain with a specific view point. That is to say that this viewpoint defines how a group of users conceptualize and visualize some specific phenomenon. This domain ontology could be linked to a specific application: Human disease for example.

4.2.4. Core Reference Ontologies :

Core reference ontology is a standard used by different group of users. This type of ontology is linked to a domain but it integrates different viewpoints related to specific group of users. This type of ontology is the result of the integration of several domain ontologies. A core reference ontology is often built to catch the central concepts and relations of the domain.

4.2.5. General Ontologies:

General ontologies are not dedicated to a specific domain or fields. They contain general knowledge of a huge area.

4.2.6. Upper level Ontologies (Foundational /Top level):

Foundational or top level ontologies are generic ontologies applicable to various domains. They define basic notions like objects, relations, events, processes and so on.

The most well-known foundational ontology are the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) and the Basic Formal Ontology (BFO). These two ontologies are formal and propose a different logical theory for representation of world assumption. Thus, domain or core reference ontologies based on the same foundational ontology can be more easily integrated.

4.3. Ontology Representation [17]:

Ontology is comprised of four main components: concepts, instances, relations and axioms. The present research adopts the following definitions of these ontological components:

4.3.1. A Concept: (also known as a class or a term):

Is an abstract group, set or collection of objects. It is the fundamental element of the domain and usually represents a group or class whose members share common properties. This component is represented in hierarchical graphs, such that it looks similar to object-oriented systems. The concept is represented by a "super-class", representing the higher class or so-called "parent class", and a "subclass" which represents the subordinate or so-called "child class". For instance, a university could be represented as a class with many subclasses, such as faculties, libraries and employees.

4.3.2. An Instance (also known as an individual):

Is the "ground-level" component of an ontology which represents a specific object or element of a concept or class. For example, "Jordan" could be an instance of the class "Arab countries" or simply "countries".

4.3.3. A Relation (also known as a slot):

Is used to express relationships between two concepts in a given domain. More specifically, it describes the relationship between the first concept, represented in the domain, and the second, represented in the range. For example, "study" could be represented as a relationship between the concept "person" (which is a concept in the domain) and "university" or "college" (which is a concept in the range).

4.3.4. An Axiom:

Is used to impose constraints on the values of classes or instances, so axioms are generally expressed using logic-based languages such as first-order logic; they are used to verify the consistency of the ontology.

4.4. Design criteria for ontologies [18]:

- Clarity: An ontology should effectively communicate the intended meaning of defined terms.
- Coherence: An ontology should give inferences that are consistent with the definitions.
- Extendibility: An ontology should have the capability to add new terms for special uses without revision of existing vocabulary.
- Minimal encoding bias: An ontology should be independent to the issues of implementing language.

- Minimal ontological commitments: An ontology should require the minimal ontological commitment that is sufficient to support knowledge sharing.

4.5. Web Semantic and Ontologies [19]:

Ontologies have become a popular research topic in many communities. In fact, ontology is a main component of this research; therefore, the definition, structure and the main operations and applications of ontology are provided. Web content consists mainly of distributed hypertext and hypermedia, and is accessed via a combination of keyword based search and link navigation.

Hence, the ontology can provide a common vocabulary, and a grammar for publishing data, and can supply a semantic description of data which can be used to preserve the ontologies and keep them ready for inference.

4.6. What Do We Represent in an Ontology?

In the context of Semantic Web, ontologies describe domain theories for the explicit representation of the semantics of the data. In other words, ontology should be seen as a right answer to provide a formal conceptualization.

Indeed, ontology must translate an explicit consensus and develop a certain level of division. It has two essential aspects to allow the operation of the resources of web by various applications or software agents. The ontologies serve then:

1) For the vocabulary, the structuring and the operation of metadatas.

2) As representation pivot for the integration of springs of heterogeneous data.

3) To describe the web departments, and generally, everywhere it is going to be necessary to press software modules on semantic representations requiring certain consensus.

4.7. Ontology Description Languages [17]:

Ontology language is the basis of ontological knowledge systems, the definition of a system of knowledge representation language specification; it not only has a rich and intuitive ability to express and use it, but the body should be easily understood by the computer, processing and applications.

On 10 February, 2004, the World Wide Web Consortium (W3C) announced its support for two Semantic Web technology standards, RDF and OWL; that is, the information resources described in semantic language specification. OWL is a standard ontology description language, built on the RDF, which is based on the XML-authoring tools, used mainly to express the needs of computer applications to deal with knowledge and information in the document.

4.8. Resource Description Framework (RDF) :

A language used to provide a standard for metadata about the resources on the Web, is capable of representing data on and exchanging knowledge over the Web. It was developed to be understood by computers, facilitating interoperability between applications. In other words, it is a framework for using and representing metadata and describing the semantics of information about web resources in a way that is accessible to machines.

RDF, which is recommended by the W3C, uses URIs to identify resources or things (the root of an ontology is called a thing). It is based upon XML, which is designed for syntax, while RDF is intended for semantics. As has been mentioned, it is a framework for describing web resources, which is why it has become a common method of describing the properties, time, information and content of web resources, so that it can be read and understood by computer applications.

RDF can be used in several applications, one of the most important being resource discovery, used to enhance search engine capabilities. It is also used to facilitate knowledge sharing and exchange in intelligent software agents and, as previously mentioned, to describe the content and content relationships available with any resource, such as a page.

The RDF model has three elements: a resource (the subject), the object and the predicate. It is possible to say that <subject> has a property <predicate> valued by <object>

4.9. Resource Description Framework Schema (RDFS) :

The Resource Description Framework Schema (RDFS) [6] has been built upon the XML and RDF models and upon syntax. RDFS offers extra facilities to encourage evolution in both the individual RDF vocabularies and the core RDF Schema vocabulary.

It provides a machine-understandable system for defining the vocabularies needed for such applications or descriptive vocabularies. In other words, it is a group of RDF resources that can be used to define or express the properties of other RDF resources which define application specific RDF vocabularies. At the same time, RDF(S) helps developers to describe classes and properties in a specific way and to specify relationships between those properties and classes, allowing combinations between classes, properties or values. In other words, RDFS is used to define RDF vocabularies.

In general, RDFS is defined in a namespace informally called 'rdfs' and identified by the URI reference http://www.w3.org/2000/01/rdfschema#.

On the other hand, RDF is defined in a namespace informally called ‘rdf’ and identified by the URI reference http://www.w3.org/1999/02/22-rdf-syntax-ns#.

4.10. Ontology Interchange Language (OIL) :

A semantic markup language for Web semantics has been built on RDF and RDF/S, this language providing modelling primitives used in frame based and DL-oriented ontologies.

OIL has many limitations, it lacks property constructors, it has no composition or transitive closure, its only property types are transitive and symmetrical, sets are the only collection type (there are no bags or lists), there is no comparison in data value, it allows only unary and binary relations, and there are neither default values nor variables.

4.11. Web Ontology Language (OWL) :

OWL, which is a language for processing web information, became a W3C recommendation in February 2004 and was built using RDF to remedy the weaknesses in RDF/S and DAML+OIL. It provides more affluent integration and interoperability of data between communities and domains.

It can be said that there is a similarity between OWL and RDF, but the former has a stronger syntax with more machine interpretability and vocabulary language than the latter. Obviously, RDF is commonly limited to binary ground predicates, and RDFS also has the limitation that it represents a subclass hierarchy and a property hierarchy, with the domain and range definitions of these properties. In other words, the language of OWL is more expressive than that of RDF(S).

To cope with the limitations of RDF, RDFS and DAML+OIL, W3C defined OWL. Indeed, OWL is an extension of RDFS; in other words, it builds on RDF and RDFS, using XML syntax; overall, OWL uses the RDF meaning of classes and properties. W3C classifies OWL into three sublanguages, each of which is intended to supply different aspects of these incompatibilities. These are OWL Lite, OWL DL and OWL Full.

OWL Lite is the simplest version of OWL and provides a classification hierarchy and simple constraints; it permits only the expression of relationships with maximum cardinality equal to 0 or 1, thus being designed for easy implementation. In this sublanguage, there is some restriction of OWL DL to a subset of language constructors, with some limitations such as an absence of explicit negation or union. The disadvantage of this sublanguage is restricted expressiveness.

OWL DL is so called because it uses Description Logic to represent the relations between objects and their properties. Indeed, it provides maximum expressiveness while preserving the completeness of computational properties. OWL Lite is a sublanguage of OWL DL.

The sublanguage OWL Full provides highest expressiveness and the syntactic freedom of RDF but without preserving guarantees on computational complexity. OWL Lite and OWL DL are sublanguages of OWL Full.

5. Databases versus Ontologies:

5.1. Important differences in semantics [20] :

	Database:	**Ontology**
World assumption	Closed world assumption (CWA) Missing information treated as false	Open world assumption (OWA) Missing information treated as unknown
Name assumption	Unique name assumption (UNA) Each individual has a single, unique name	No UNA Individuals may have more than one name
Behavior	Schema behaves as constraints on structure of data Define legal database states	Ontology axioms behave like implications (inference rules) Entail implicit information

Table 1: Important differences in semantics

5.2. Query answering:

5.2.1. OWA example N°1:

given facts/data:	
Individual: Ali **Facts**: **hasFriend** Omar **hasFriend** Mohamed **hasToy** SuperMario **Individual**: Monsif	
Query: Is Monsif a friend of Ali?	
Database response	**Ontology response**
No	**Don't know** OWA (didn't say Monsif was not Ali's friend)

Table 2: : Open World Assumption e.g. N°1

5.2.2. UNA example N°1:

given facts/data:	
Individual: Ali **Facts**: **hasFriend** Omar **hasFriend** Mohamed **hasToy** SuperMario **Individual**: Monsif	
Query: How many friends does Ali have?	
Database response	**Ontology response**
2	**at least 1** No UNA (Omar and Mohamed may be 2 names for same person)

Table 3: Unique Name Assumption e.g. N°1

5.2.3. OWA example N°2:

given facts/data:	
Individual: Ali **Facts**: **hasFriend** Omar **hasFriend** Mohamed **hasToy** SuperMario **Individual**: Monsif **DifferentIndividuals**: Omar Mohamed	
Query: How many friends does Ali have?	
Database response	**Ontology response**
2	**at least 2** OWA (Ali may have more friends we didn't mention yet)

Table 4 Open World Assumption e.g. N°2

5.2.4. UNA example N°2:

given facts/data:	
Individual: Ali **Facts**: **hasFriend** Omar **hasFriend** Mohamed **hasToy** SuperMario **Types**: **hasFriend only** Omar **or** Mohamed **Individual**: Monsif **DifferentIndividuals**: Omar Mohamed	
Query: How many friends does Ali have?	
Database response	**Ontology response**
2	**2!**

Table 5: Unique Name Assumption e.g. N°2

5.2.5. Insert new Facts/Data example:

Insert new facts/data:	
Individual: Maria **Individual**: Dora **Types**: Puppet **Facts**: **isToyOf** Maria	
Response from DBMS?	**Response from Ontology reasoner?**
Update rejected: constraint violation **Range of hasPuppet is Human;** **Maria is not Human (CWA)**	**Infer that Maria is Human** **(range restriction)** **Also infer that Maria is a Girl** (only a Girl can play with Dora Puppet)

Table 6: Insert new Facts/Data example

5.3. DB/Ontology query answering comparison [20]:

DB Query Answering	Ontology Query Answering
Schema plays no role – Data must explicitly satisfy schema constraints	**Ontology axioms play a powerful and crucial role** – Answer may include implicitly derived facts – Can answer conceptual as well as extensional queries • E.g., Can a Muggle have a Phoenix for a pet?
Query answering amounts to model checking – I.e., a "look-up" against the data	**Query answering amounts to theorem proving** – I.e., logic al entailment
Can be very efficiently implemented – Worst case complexity is low (log space) w.r.t. size of data	**May have very high worst case complexity** – E.g., for OWL, NP-hard w.r.t. size of data (upper bound is an open problem) – Implementations may still behave well in typic al cases

Table 7: Query answering comparison

5.4. When to Use an Ontology or DB [20]?

- Consider using an Ontology when:
  - Schema is large and/or complex and/or used at query time.
    - Can use reasoning to structure and check schema.
    - Inferred answers and/or intentional queries.
  - Not possible/reasonable to assume complete information.
    - E.g., modeling complex structures or activities.
  - Willing to pay potential performance cost.
- Consider using a DB when:
  - Schema is small and/or simple and/or not used at query time.
  - Complete information is available.
    - E.g., booking systems.
  - Need performance guarantees.

Chapter 3:

Medical Ontologies

1. Preface:

The use of ontologies in medicine is mainly focused on the representation and (re-) organization of medical terminologies.

Physicians developed their own specialized languages and lexicons to help them store and communicate general medical knowledge and patient-related information efficiently. Such terminologies, optimized for human processing, are characterized by a significant amount of implicit knowledge.

2. Medicine ontology definition [21]:

An ontology can be viewed as a declarative model of a domain that defines and represents the concepts existing in that domain, their attributes and the relationships between them. It is typically represented as a knowledge base which then becomes available to applications that need to use and/or share the knowledge of a domain. Within health informatics, an ontology is a formal description of a health-related domain.

3. Benefits of medicine ontologies [21]:

- Ontologies can help build more powerful and more interoperable information systems in healthcare.
- Ontologies can support the need of the healthcare process to transmit, re-use and share patient data.
- Ontologies can also provide semantic-based criteria to support different statistical aggregations for different purposes.
- Possibly the most significant benefit that ontologies may bring to healthcare systems is their ability to support the indispensable integration of knowledge and data.

<u>On the negative side:</u>

- Some remain sceptical about the impact that ontologies may have on the design and maintenance of real-world healthcare information systems.

4. Current medicine ontologies:

4.1. Gene Ontology [22]:

Gene **O**ntology (GO) is a controlled biological terminology being created by a consortium of bioinfomaticians. Even though it's relatively new to the world when compared to other ontologies, GO has greater impact on bioinformatics community.

GO started with terminologies from three genomic databases: *Flybase*, *Saccharomyces*

Genome Database and *Mouse Genome Database* and has developed three hierarchies of terms to describe biological processes, cellular components and molecular functions. Gene definitions and comments by the authors are given as annotations in the ontology.

4.2. UMLS [22]:

UMLS acronym for **U**nifies **M**edical **L**anguage **S**ystem was developed by National Library of Medicine. UMLS identifies any entry into three groups:

- *String* – representing a term as terminology.
- *Lexical Group* – strings of same structure can be mapped.
- *Concept* - string of identical meaning.

UMLS is divided into three groups: *Metathesaurus*, *Semantic Network*, and *Special Lexicons*.

- The metathesaurus will hold all the biological terms in the database.
- The *Semantic Network* connects different terms in the metathesaurus semantically.
- The Semantic Network consists of 135 semantic types, organized into a pair of hierarchies with 6,864 relations among them.

The relations show how different semantic types are related to each other. The figure shown below shows the relation between the two layers of UMLS.

The knowledge is represented by actual instances of relations between actual concepts. The figure below illustrates the communication between UMLS layers.

For example, *Disease* is semantic type with around 392 relations (109 semantic relations and 22 other relations). *Pneumonia* categorized under one semantic type *Disease*, but has hundreds of relations.

The figure below shows the two-level structure of UMLS.

The UMLS has probably a greater impact on biomedical ontology work than any other terminology effort because of its long history, its early focus on knowledge representation and its free availability.

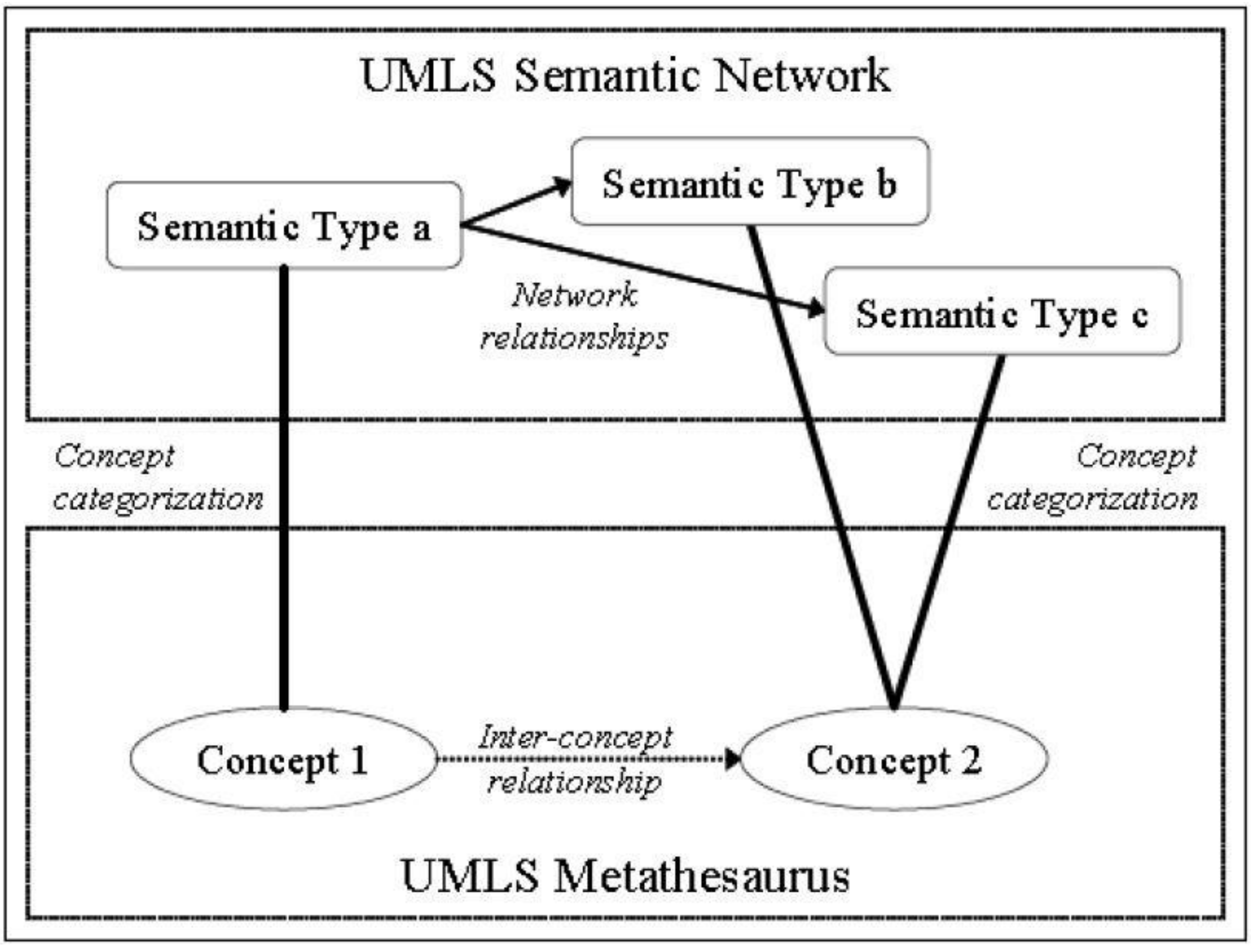


Figure 2: UMLS Metathesaurus and Semantic Network

4.3. OBO Foundry [23] :

The OBO Foundry is a collective of ontology developers that are committed to collaboration and adherence to shared principles. The mission of the OBO Foundry is to develop a family of interoperable ontologies that are both logically well-formed and scientifically accurate.

To achieve this, OBO Foundry participants voluntarily adhere to and contribute to the development of an evolving set of principles including open use, collaborative development, non-overlapping and strictly-scoped content, and common syntax and relations, based on ontology models that work well, such as the Gene Ontology (GO).

OBO Foundry produces over than 20 ontologies, we will cite the two most used.

4.4. Human Disease Ontology (DOID) [24]:

An ontology for describing the classification of human diseases organized by etiology.

The Disease Ontology has been developed as a standardized ontology for human disease with the purpose of providing the biomedical community with consistent, reusable and sustainable descriptions of human disease terms, phenotype characteristics and related medical vocabulary disease concepts through collaborative efforts of researchers at Northwestern University, Center for Genetic Medicine and the University of Maryland School of Medicine, Institute for Genome Sciences.

The Disease Ontology database has been updated to the latest ontology as of 02-03-2017. (Number of terms: 11961)

Figure 3: Diseases hierarchical classes

4.5. Symptom Ontology (SYMP) [25] :

An ontology of disease symptoms, with symptoms encompassing perceived changes in function, sensations or appearance reported by a patient indicative of a disease.

The symptom ontology was designed around the guiding concept of a symptom being: "A perceived change in function, sensation or appearance reported by a patient indicative of a disease". Understanding the close relationship of Signs and Symptoms, where Signs are the objective observation of an illness, the Symptom Ontology will work to broaden its scope to capture and document in a more robust manor these two sets of terms. Understanding that at times, the same term may be both a Sign and a Symptom.

The Symptom Ontology was developed as part of the Gemina project starting in 2005 at TIGR. Work continues on the project at the Institute for Genome Sciences (IGS) at the University of Maryland.

The Symptom Ontology is organized primary by body regions with a branch for general symptoms.

The Symptoms Ontology in July 2008 was submitted for inclusion and review to the OBO Foundry.

This project is open to collaborative development.

Figure 4: Symptom hierarchical classes

5. Clinical Decision Support System [22]

The clinical decision support system (CDSS) can be defined as "interactive computer programs, which are designed to assist physicians and other health professionals with decision making tasks". It can also be defined as "Clinical Decision Support systems link health observations with health knowledge to influence health choices by clinicians for improved health care".

Achieving perfection in providing good health care by physicians is very complex and perplexing. This condition becomes worse when physicians have limited knowledge about the working domain or are working under emergency conditions. With the advances in expert systems, rule engines, case base systems and through combination with advanced methods of information retrieval, information storage, knowledge representation, a decision support system can be designed which will aid novice nurses during their regular practice. For implementing a CDSS we need to have the following components:

1. Knowledge Base – this acts as a brain to your application.

2. Rule Base Engine – this engine will hold rules for your application.

3. Business Model – this model will be responsible for talking to various components and inter-connecting them.

4. UI – possible user interface so that user can interact with the application.

The typical architecture of the CDSS is shown in the figure 16 below. The figure shows how each component communicates with each other in the application.

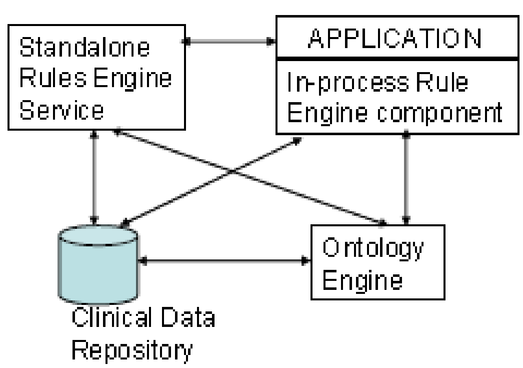


Figure 5: Clinical Decision Support System Architecture

Chapter 4:

EasyDiag System Design & Implementation

1. Preface:

There are currently no ontologies that relate diseases and symptoms and only attempts at their infancy along with some simple proposed models. However, well established ontologies for diseases and for symptoms were already developed in isolation, DOID & SYMP ontologies.

A lot of researches nowadays tries to link these two ontologies to create a core diseases symptoms ontology.

2. Related Works:

2.1. Building a Diseases Symptoms Ontology for Medical Diagnosis [26]:

"An Integrative Approach"

Project Authors:

Osama Mohammed: Department of Software Engineering, Lakehead University, Ontario, Canada

Rachid Benlamri and Simon Fong: Department of Computer and Information Science, University of Macau, Macau, China

Project abstract:

They proposed an alignment algorithm to align these two ontologies. The proposed alignment method has been repeated for 11 inter-related diseases to produce a core ontology for disease diagnosis—Diseases Symptoms Ontology (DSO).

Indeed their method and process can be repeated for any number of diseases to create a larger version of the DSO. The core DSO ontology has been published on university Flash server (http://flash.lakeheadu.ca/~omohamme/DSO.owl) and can be used for analysis by other semantic web applications.

2.2. DDO: a diabetes mellitus diagnosis ontology [27]:

Project Authors:

Shaker El‑Sappagh: Faculty of Computers and Information, Minia University, El‑Minia, Egypt, Egypt.

Farman Ali: Department of Information and Communication Engineering, Inha University, Incheon, South Korea.

Project abstract:

In this study, they take the first step in this direction, by designing an OWL2 diabetes diagnosis ontology (DDO). Protégé 5 software was used for the construction of the ontology.

DDO is developed within the framework of the basic formal ontology and the ontology for general medical science to represent entities in the domain of diabetes, and it follows the design principles recommended by the Open Biomedical Ontology Foundry.

Currently, DDO contains 6444 concepts, 48 properties, 13,551 annotations, and 27,127 axioms. DDO can serve as a diabetes knowledge base and supports automatic reasoning. It represents a major step toward the development of a new generation of patient-centric decision support tools. DDO is available through BioPortal at:

http://www.bioportal.bioontology.org/ontologies/DDO.

2.3. Ontology-Based Diagnosis and Personalization of Medical Knowledge [5]:

Project Authors:

Cristina Romero Tris, **Director**: David Riano

Departament d'Enginyeria Informàt ica i Matemàt iques Rovira Virgili, Spain

Project abstract:

The main objective of this project is to develop an application to guide physicians in the diagnostic process. The specific objectives of this project are three.

- Support physicians in the diagnostic process
- Validate the patient's condition as a whole
- Personalize the knowledge in the Case Profile Ontology

2.4. Toward an Ontological Treatment of Disease and Diagnosis [28]:

Project Authors:

Richard H. Scheuermann, PhD, Department of Pathology and Division of Biomedical Informatics, University of Texas.

Werner Ceusters, MD: Southwestern Medical Center, Dallas, TX; Department of Psychiatry, Center of Excellence in Bioinformatics, University at Buffalo, Buffalo, NY

Barry Smith, PhD: Southwestern Medical Center, Dallas, TX; Department of Philosophy and Center of Excellence in Bioinformatics, University at Buffalo, Buffalo, NY

Project abstract:

The goal of this communication is to outline a terminological framework that encompasses diseases, their causes and manifestations, and diagnostic acts and other entities pertaining to the ways diseases are recognized and interpreted in the clinic. Inspection reveals that such entities have thus far not been adequately treated in standard vocabulary resources.

2.5. Interpreting Patient Data using Medical Background Knowledge [29]:

<u>Project Authors:</u>
Heiner Oberkampf, Sonja Zillner: Siemens AG, Corporate Technology, Munich, Germany
Bernhard Bauer: Software Methodologies for Distributed Systems, University Augsburg,Gr
Matthias Hammon: University Hospital Erlangen, Germany

<u>Project abstract</u>

1. Problem: Integrate patient-data in clinical decision support systems.

2. Approach: Development of a Disease-Symptom Knowledge Model.

3. Conclusion: Relations between diseases and symptoms make possible to integrate unstructured clinical data in decision support systems.

3. Our Approach:

3.1. Problems:

- In order to diagnose a patient disease, physician pass through some steps, which can be sometimes long, ambiguous and complex; especially in the case of diseases which have common symptoms.
- There is no relation between DOID & SYMP ontologies.
- The history records of patients can provide more signs to help in diagnosis.

3.2. Objectives:

Create an application to guide physician through the diagnosis operation, the decision-making and helps him to:

a. <u>Link disease to its symptoms. (Create a new ontology “diag.owl”).</u>

b. <u>Use the patient history record. (Create an ontology “patients-records.owl” from the Database).</u>

c. <u>Infer possible diseases from patient state description.</u>

d. <u>Interrogate patient.</u>

e. <u>Remind possible symptoms.</u>

f. <u>Infer new signs from patient actual/past examinations.</u>

3.3. The design:

3.3.1. System components:

- Database: MySQL database: (patient, Consultation, Symptom, Diagnosis tables)
- DOID.OWL, SYMP.OWL predefined ontologies.
- DIAG.OWL, RECORD.OWL ontologies created.

3.3.2. Model :

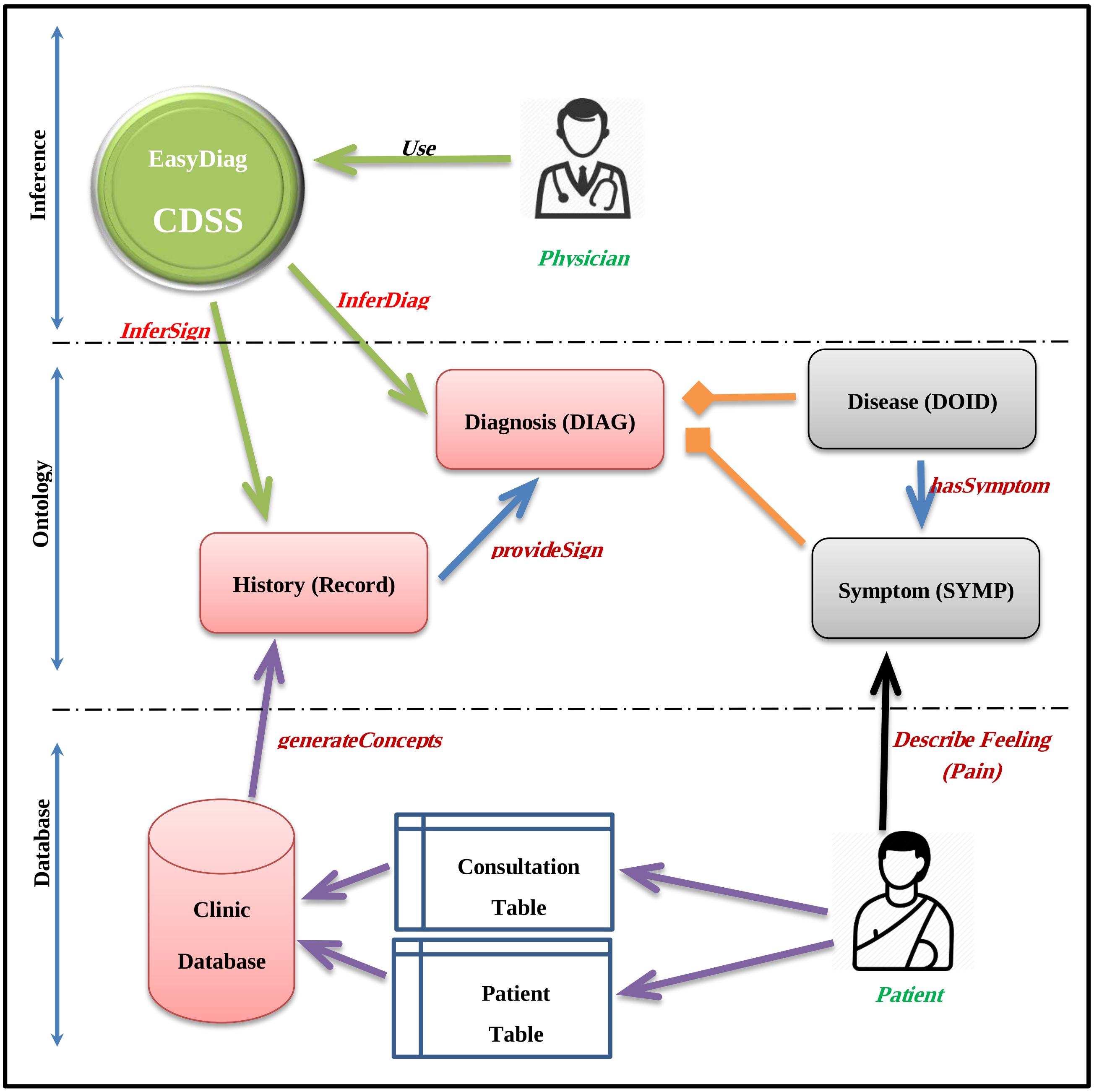


Figure 6: EasyDiag CDSS Model

3.3.3. Explanation of the model:

Database part:

In the first step, the user of EasyDiag Solution should take the patient's general information (first name, last name, birthday…etc.), then create to him the consultation registration, which contains information about the purpose of the visit, the date and some healthy signs like temperature, heartbeat, cardiac frequency…etc.

Ontology part:

In this part the user creates the records of the patients, or update it if it exists, those records are collected in one ontology called "RECORD" under the RDF form (OWL file).

The user can create this ontology by a simple click.

In addition the user handles two predefined ontologies (DOID and SYMP) which are the disease ontology and the symptom ontology, to create another one called "DIAGNOSTIC".

The EasyDiag Solution offers to the user a simple way to link each disease to their symptoms, when he validates the relation the DIAGNOSTIC ontology will be updated automatically.

Inference part:

So in this part we got two main ontologies to help the physician in diagnosis, RECORD & DIAGNOSTIC, the first one provides some signs from the history of patient's consultations using SWRL rules and the second one gives the possible diseases from the feeling description of the patient.

The physician listens to the feeling or the pain description of the patient; he transforms them to their scientific labels using the EasyDiag Solution, which gives a list of all existing symptoms to choose from.

The system infer all the diseases related with those symptoms each one with the percentage of its repetition, the system provides also more possible symptoms extracted from diseases inferred.

The RECORD ontology provides other signs if exists, those signs will be added to the symptoms list, the physician repeat the diagnosis operation to get the final diagnosis.

If the physician is satisfy about the result, he confirms the diagnosis, if not he demands other tests.

3.3.4. Functionalities:

The CDSS application **EasyDiag** is oriented to use by physicians and doctors, and it provides many functionalities, leading to diagnose the patient disease.

To explain this functionalities we used the UML (Unified Modelling Language).

Uses Case Diagram:

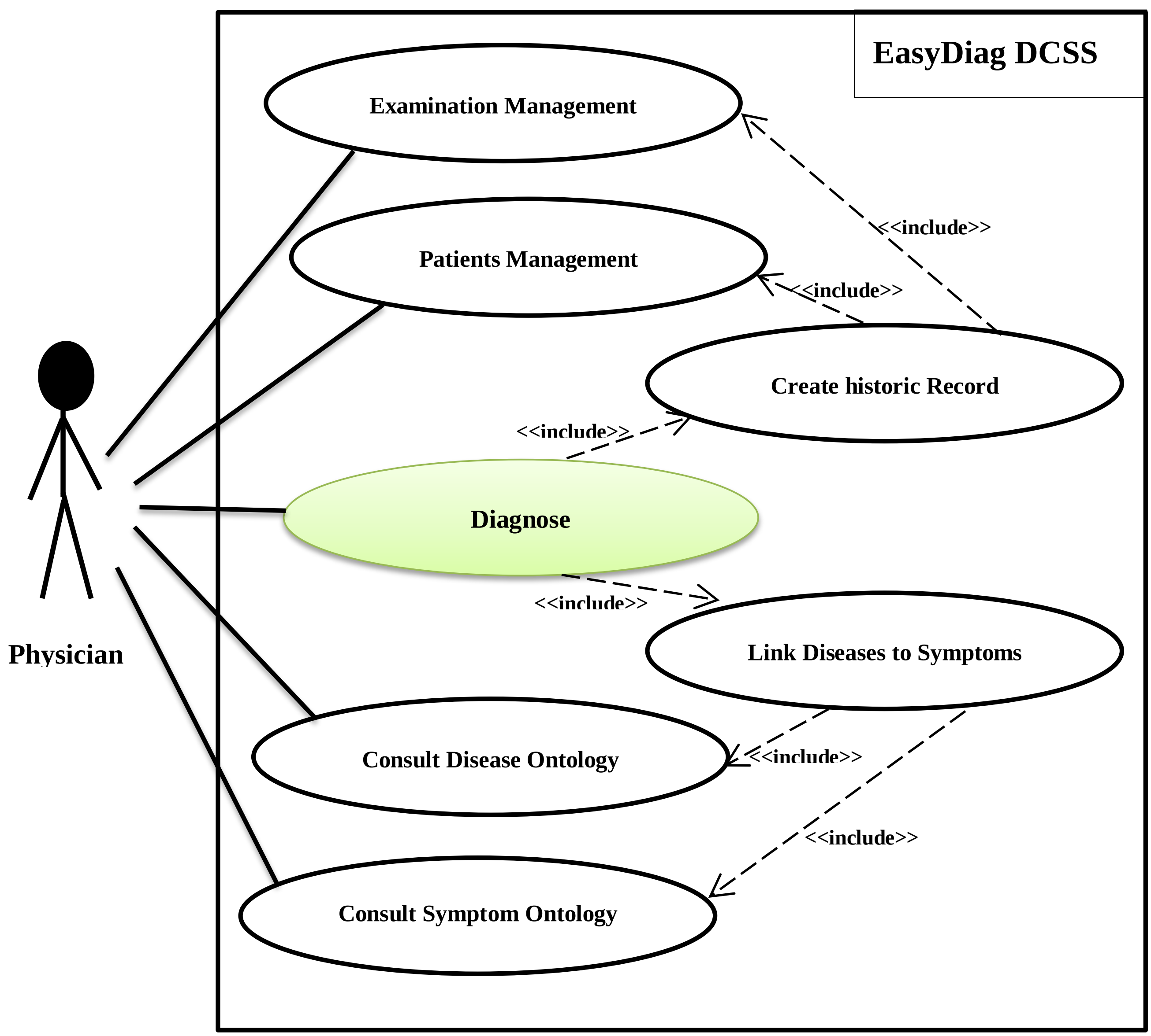


Figure 7: Uses Case Diagram of EasyDiag DCSS

4. Software Used:

4.1. Languages:

4.1.1. Programming language :

We used in the development of **EasyDiag** system, **Java** computer programming language, that is concurrent, class-based, object-oriented, and specifically designed to have as few implementation dependencies as possible. It is intended to let application developers "write once, run anywhere" (**WORA**), meaning that compiled Java code can run on all platforms that support Java without the need for recompilation [30].

4.1.2. Structured Query Language (SQL) :

Is a standardized programming language used for managing relational databases and performing various operations on the data in them. Initially created in the 1970s, SQL is regularly used by database administrators, as well as by developers writing data integration scripts and data analysts looking to set up and run analytical queries [31].

4.1.3. Web Ontology Language (OWL):

The W3C Web Ontology Language (OWL) is a Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relations between things. OWL is a computational logic-based language such that knowledge expressed in OWL can be exploited by computer programs, e.g., to verify the consistency of that knowledge or to make implicit knowledge explicit. OWL documents, known as ontologies, can be published in the World Wide Web and may refer to or be referred from other OWL ontologies. OWL is part of the W3C's Semantic Web technology stack, which includes RDF, RDFS, SPARQL, etc [32].

4.1.4. Simple Protocol and RDF Query Language (SPARQL):

Defines a standard query language and data access protocol for use with the Resource Description Framework (RDF) data model. It works for any data source that can be mapped to RDF. The specification is under development by the RDF Data Access Working Group (DAWG) [33].

4.1.5. Semantic Web Rule Language (SWRL) :

Is a proposed language for the Semantic Web that can be used to express rules as well as logic, combining OWL DL or OWL Lite with a subset of the Rule Markup Language [34].

4.2. Platform:

There are many software platforms supported by Java, we chose the newest one, which is **JavaFX** platform.

JavaFX is a set of graphics and media packages that enables developers to design, create, test, debug, and deploy rich client applications that operate consistently across diverse platforms [35].

And we applied the **MVC** (Model View Controller) design pattern.

It is a design pattern for computer software. It can be considered an approach to distinguish between the data model, processing control and the user interface. It neatly separates the graphical interface displayed to the user from the code that manages the user actions. The objective is to provide a framework which enforces better and more accurate design [36].

4.3. Environment:

4.3.1. NetBeans:

Is a software development platform written in Java. The NetBeans Platform allows applications to be developed from a set of modular software components called modules. Applications based on the NetBeans Platform, including the NetBeans integrated development environment (IDE), can be extended by third party developers [37].

4.3.2. WampServer:

WampServer (Windows, Apache, MySQL, PHP) is a Windows web development environment. It allows you to create web applications with Apache2, PHP and a MySQL database. Alongside, PhpMyAdmin allows you to manage easily your databases [38].

4.4. DBMS:

MySQL, is an open source relational database management system. It is based on the structure query language (SQL), which is used for adding, removing, and modifying information in the database. Standard SQL commands, such as ADD, DROP, INSERT, and UPDATE can be used with MySQL [39].

4.5. Tools:

4.5.1. JavaFX SceneBuilder:

Is a visual layout tool that lets users quickly design JavaFX application user interfaces, without coding. Users can drag and drop UI components to a work area, modify their properties, apply style sheets, and the FXML code for the layout that they are creating is automatically generated in the background. The result is an FXML file that can then be combined with a Java project by binding the UI to the application's logic [40].

4.6. Framework :

4.6.1. Apache Jena :

Is an open source Semantic Web framework for Java. It provides an API to extract data from and write to RDF graphs. The graphs are represented as an abstract "model". A model can be sourced with data from files, databases, URLs or a combination of these. A Model can also be queried through SPARQL [41].

4.6.2. Resource Definition framework (RDF) :

Is a family of World Wide Web Consortium (W3C) specifications originally designed as a metadata data model. It has come to be used as a general method for conceptual description or modeling of information that is implemented in web resources, using a variety of syntax notations and data serialization formats. It is also used in knowledge management applications [42].

4.6.3. Protégé:

Is a free, open source ontology editor and a knowledge management system. Protégé provides a graphic user interface to define ontologies. It also includes deductive classifiers to validate that models are consistent and to infer new information based on the analysis of an ontology.

Like Eclipse, Protégé is a framework for which various other projects suggest plugins. This application is written in Java and heavily uses Swing to create the user interface. Protégé recently has over 300,000 registered users. According to a 2009 book it is "the leading ontological engineering tool".

Protégé is being developed at Stanford University and is made available under the BSD 2-clause license. Earlier versions of the tool were developed in collaboration with the University of Manchester [43].

5. Implementation:

EasyDiag has 10 interfaces each one serves a functionality or some functionalities:

5.1. Principal Interface:

This is the main window executed after the run of **EasyDiag** System.

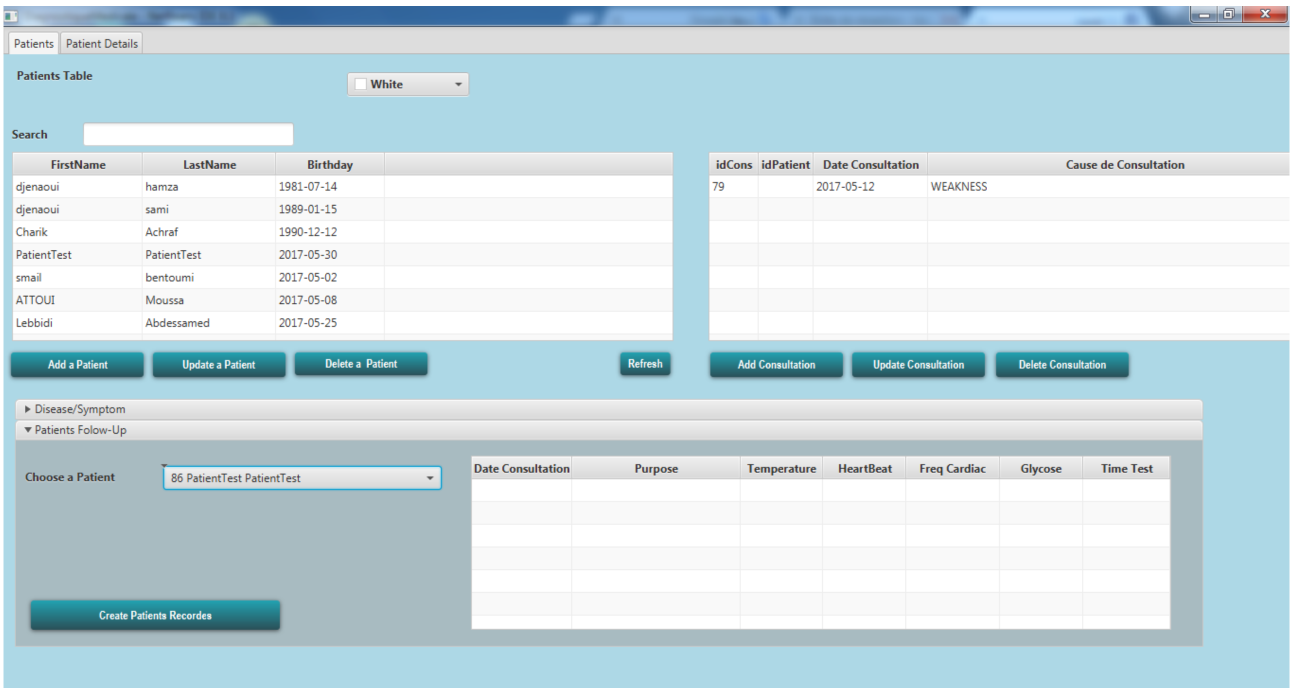


Figure 8: The Main Window

- Contains the patients and the consultations tables.
- Add, Update, and Delete Buttons.
- Consultations shown when patient selected from the patient table.
- Give access to DIAG and RECORD ontologies.
- Contains Patients Details Tab.

5.2. Add new Patient:

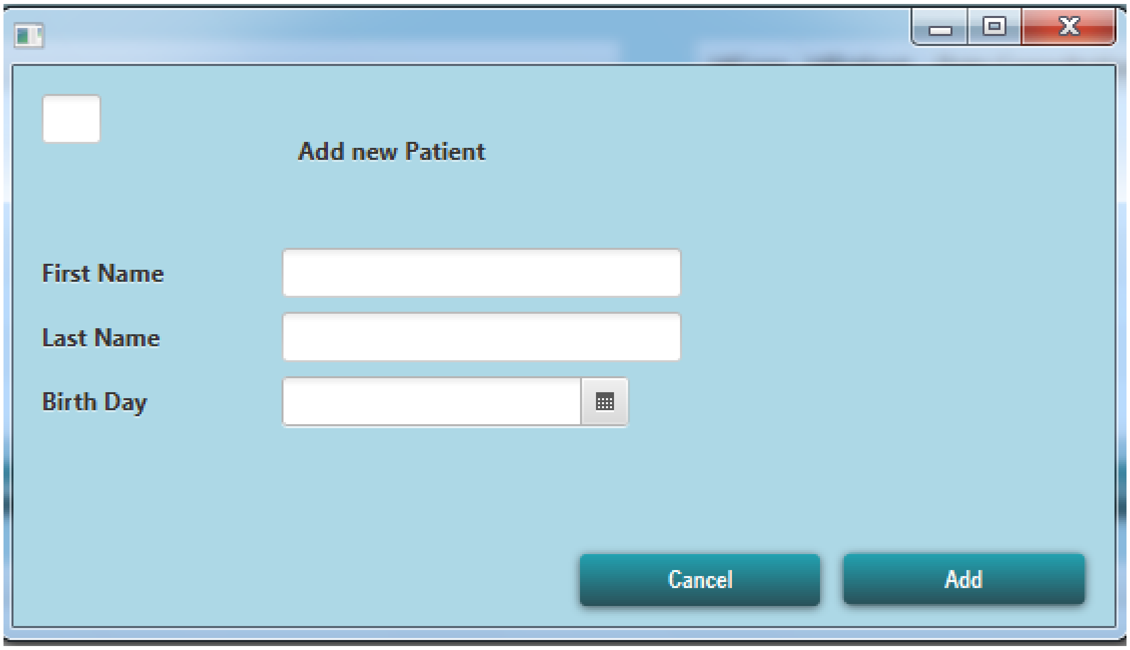


Figure 9: Add new patient window

We can access to this window from the Add button in the principle window, to add new patient, this window contains few information about patient (First, Last Name and Birth Day).

5.3. Add Patient Details:

This is the Patients Details Tab, which serve to complete Patient's information.

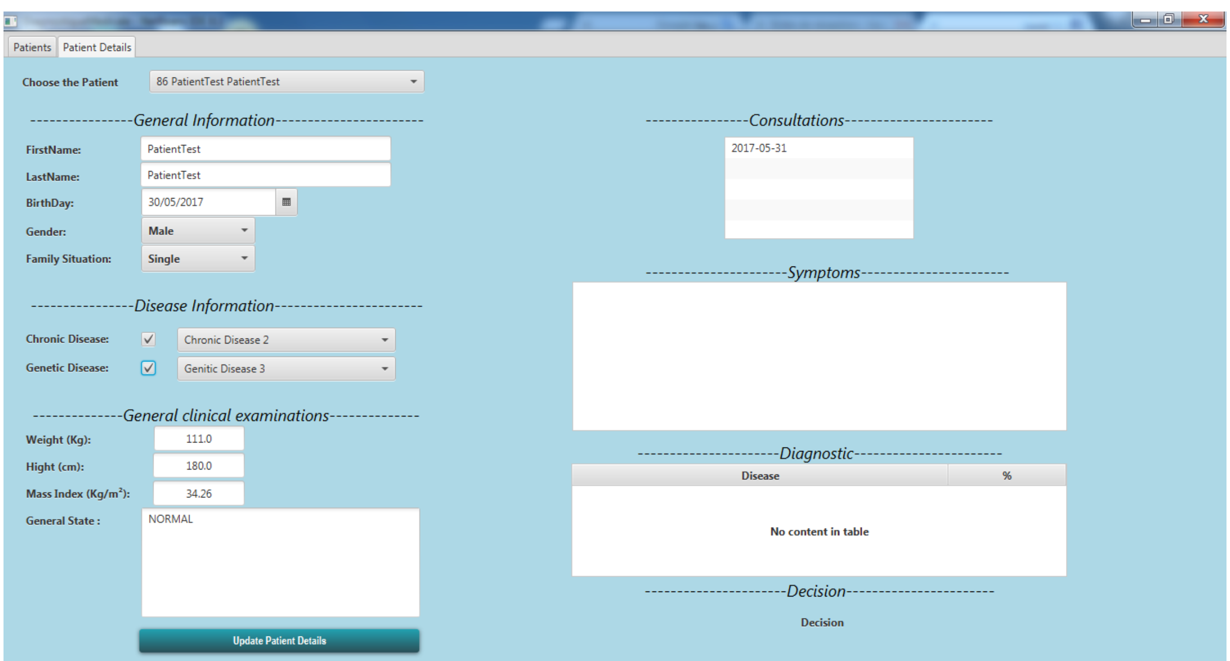


Figure 10: Patient Detail Window

5.4. Modify Patient Info:

We can update patient's information using this window, firstly we should choose the patient from the patient table in the main window and click the update button.

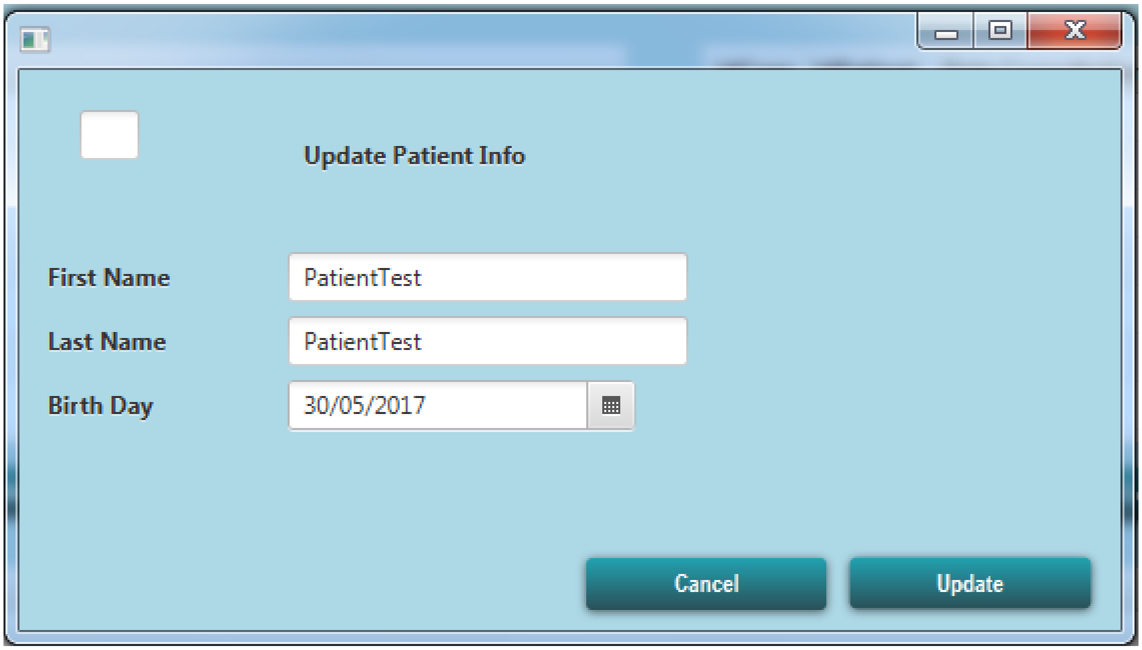


Figure 11: Update Patient Infos

5.5. Delete A Patient:

This is a warning message to alert user that he will delete the patient and all his related information.

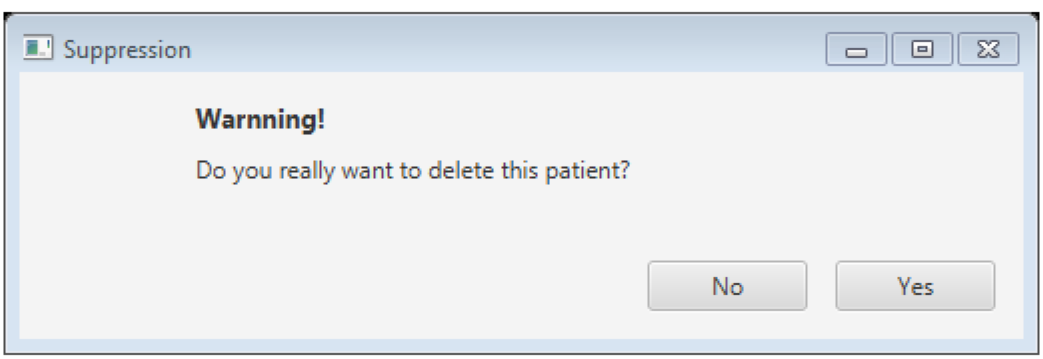


Figure 12: Delete Warning Message

5.6. Relate Disease with Symptoms :

The role of this window is to create the DIAG ontology if not exists, creates disease instances and links them with symptoms with the "**ObjectProperty**": **hasSymptom**.

After choosing the disease and the symptoms related we press the validate button, to generate the OWL code and update our DIAG ontology.

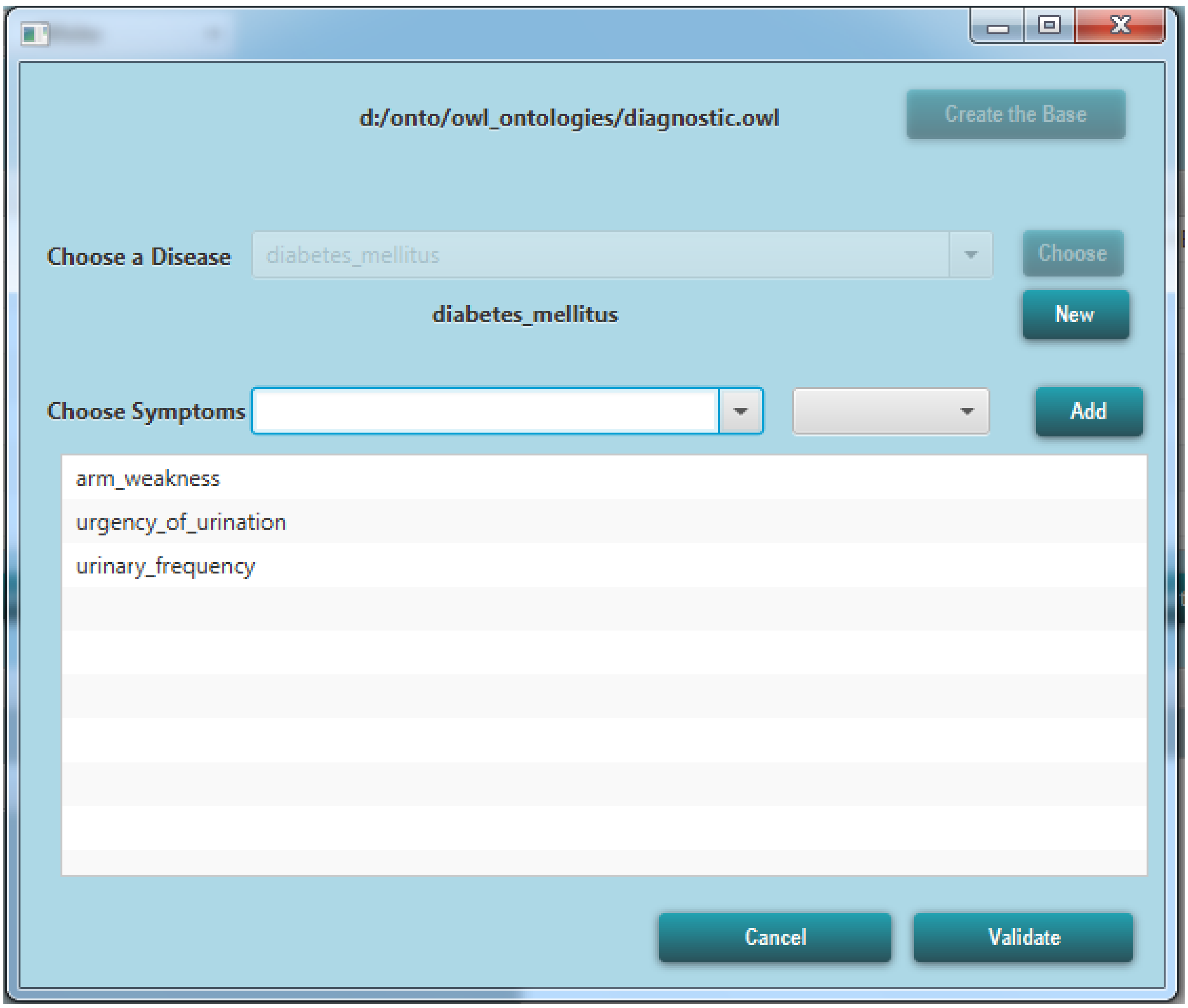


Figure 13: Disease-Symptoms Window

5.7. Generate Patient Records :

In this window, we convert each patient record from our MySQL database to patient instance the RECORD ontology, which is created in advance.

The information retrieved will be organized following some characteristics, to be used later in diagnosis (provide signs).

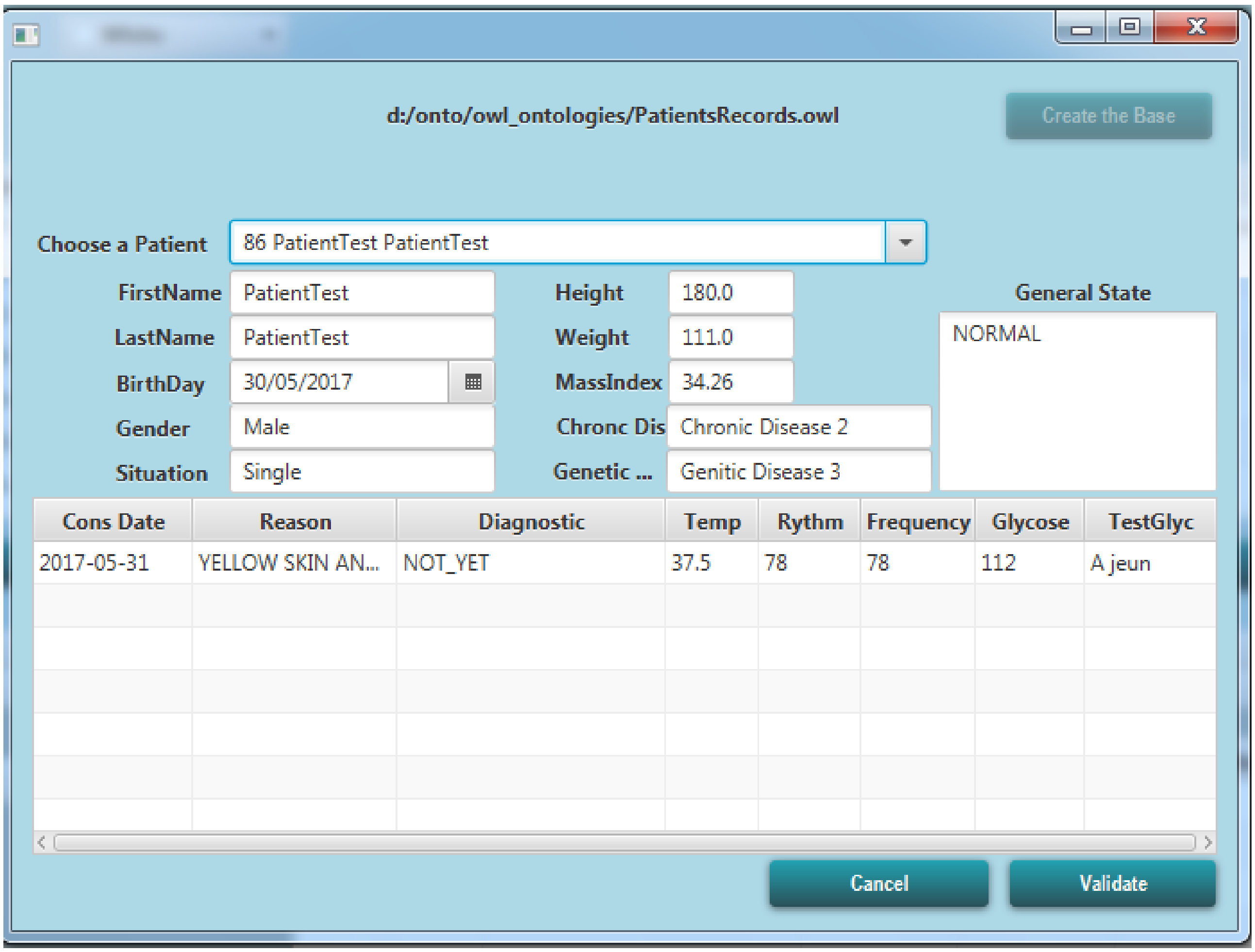


Figure 14: Patient Record Window

5.8. Add new Consultation:

In order to make diagnosis, the physician must pass by this step, the addition of consultation, and fills all fields that concern general exams made.

When the button add consultation clicked, the system insert data in the database and redirect the user to the other tab (Symptoms).

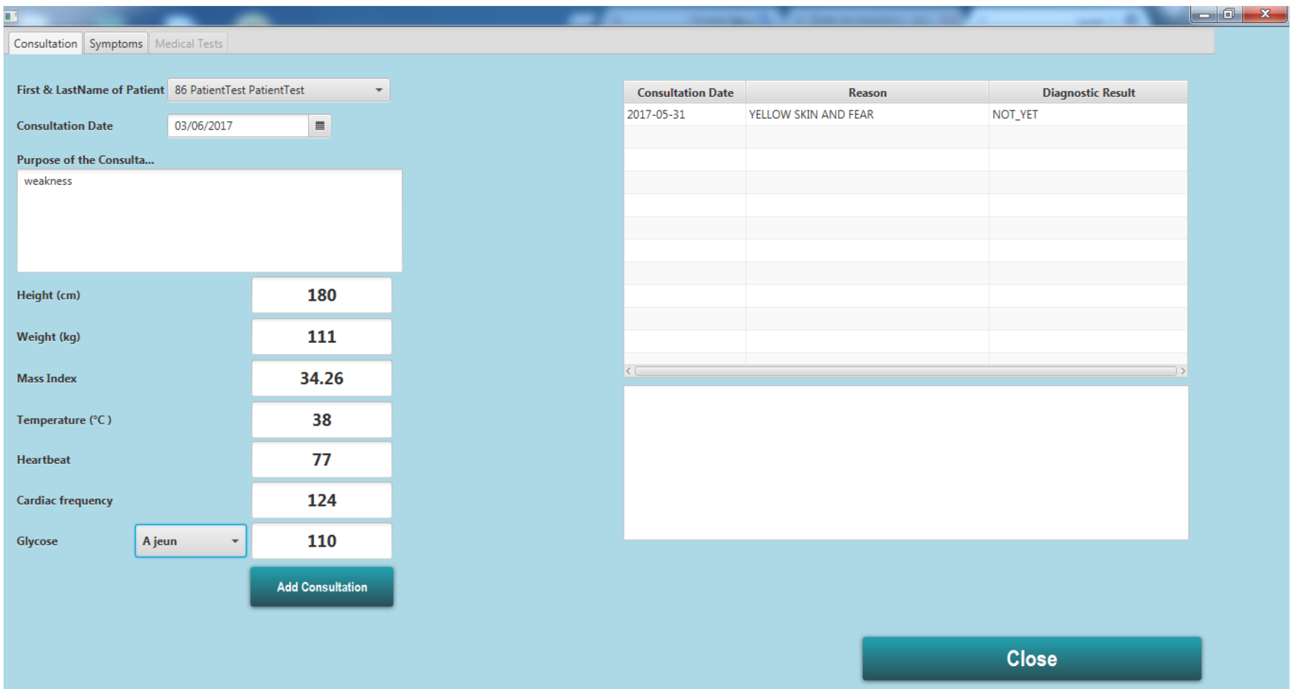


Figure 15: Add New Consultation

5.9. Diagnosis:

Actually this is the most important step in the diagnosis process, that we can in this window use our created ontologies to make decision.

This window provide to physician the standard way to interrogate patient, starting from the description of feelings and pains, passing by interpret this description to symptoms and finally the **EasyDiag** system gives the possible diseases (by clicking Possibilities button).

If the diagnosis proposed isn't appropriate, the system propose alternative solutions:

- <u>Add More Symptoms</u>, by clicking this button, another window will open, and we will explained its role in the next interface (below).
- <u>Use the Patient Records</u>: the physician use here the history of patient, which is stored in RECORD ontology, the **EasyDiag** system retrieve signs or symptoms from it, using some pre-defined rules.

If the final diagnosis is appropriate, the physician can confirm the decision, if not he asks for more tests.

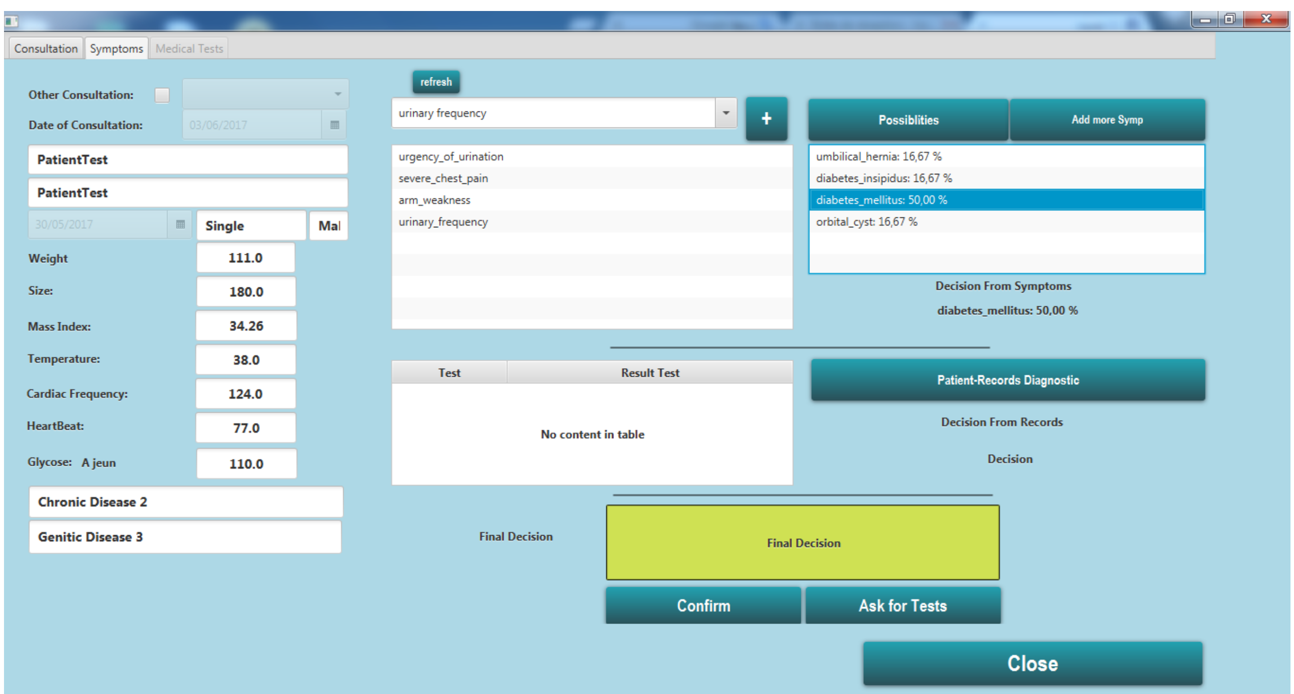


Figure 16: Diagnosis Window

5.10. Add more Symptoms:

This window was created to help physician to remember the symptoms of each expected disease, and adds the accepted ones to the symptoms list, to repeat the diagnosis operation.

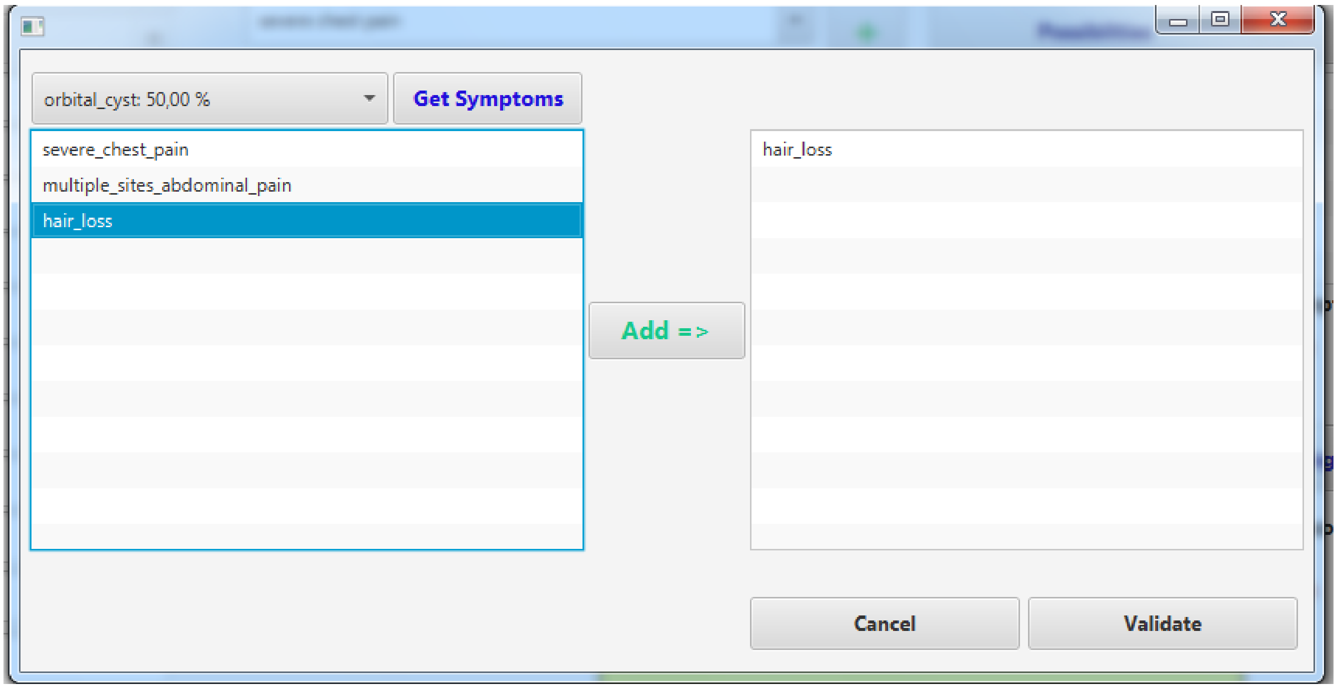


Figure 17: Add More Symptoms Window

Conclusion:

Medical diagnosis is a major concern in modern science, especially artificial intelligence and semantic web, where there are organizations for development and research in this important area.

But despite all this attention, medical diagnosis still needs a lot of efforts, especially the part about creating its own ontology, because what exists now are sporadic attempts that need to be intensified and unifying efforts.

Our work in this research is about creating and using an ontology linking diseases and their symptoms, as well as using patient data to arrive at a correct medical diagnosis, to help the physician to make the right decision, especially in similar and complex situations.

We have relatively reached the main objective of this research, through:

- Establish a management system to medical clinic and link it to a database of patients.
- Giving the program user the ability to link diseases with their symptoms, including the establishment of a diagnostic ontology,
- Transferring patient data to an ontology for the history of each patient's information, examinations and tests.
- Use of two ontologies created in diagnosis.
- Follow the approved method of diagnosis.

Despite all of that but we think that it is not sufficient, because the creation of diagnosis ontology needs a group of experts on medicine, and we have some suggestions to complete the research in the future.

Suggestions:

On the way to get a successful work, we suggest a set of proposals:

- Create a group of medicine experts to set the rules, to link Disease-Symptom.
- Create a system to generate rules and extract signs from patient's data.
- The use of fuzzy-logic can give more efficient results.

Finally we hope that this modest work gave an addition to this important domain "Medical Diagnosis"

Bibliography

[1] Berners-Lee, The Semantic Web, 2001.

[2] "dictionary.com," [Online]. Available: http://www.dictionary.com/browse/diagnosis. [Accessed 14 05 2017].

[3] oxford dictionaries, "oxforddictionaries," [Online]. Available: https://en.oxforddictionaries.com/definition/diagnosis. [Accessed 14 05 2017].

[4] cambridge dictionary, "dictionary.cambridge.org," [Online]. Available: http://dictionary.cambridge.org/dictionary/english/diagnosis. [Accessed 14 05 2017].

[5] C. R. Tris, Ontology-Based Diagnosis and Personalization of Medical Knowledge., Universitat Rovira i Virgili, 2009, p. 8.

[6] M. Critchley, Medical Dictionary, London: Butterworths, 1986.

[7] medical-dictionary, "The free dictionary," [Online]. Available: http://medical-dictionary.thefreedictionary.com/diagnosis. [Accessed 14 05 2017].

[8] World Health Organization (WHO), "Good diagnostic practice," in *Good clinical diagnostic practice*, Cairo, Regional Office for the Eastern Mediterranean, 2005, p. 15.

[9] C. Nordqvist, "What Are Signs And Symptoms And Why Do They Matter?," *Medical News Today ,* 7 6 2016.

[10] C. Nordqvist, "What Are Signs And Symptoms And Why Do They Matter?," *Medical News Today*, 7 6 2016.

[11] A. Gupta and A. Gupta, Differential Diagnosis, New Delhi: JAYPEE BROTHERS, MEDICAL PUBLISHERS (P) LTD, 2005.

[12] S. Bleisch, Databases, Geographic Information Technology Training Alliance (GITTA), 2007.

[13] Hans-Petter Halvorsen, Introduction to Database, Southeast Norway: University College of Southeast Norway, 2016.

[14] TutorialsPoint, "DataBase Management System (DBMS)," 2015. [Online]. Available: https://www.tutorialspoint.com/dbms/dbms_tutorial.pdf. [Accessed 16 5 2017].

[15] I. Horrocks, Ontologies and the Semantic Web, Manchester: University of Manchester.

[16] F. P. M. A. K. a. O. C. Catherine Roussey, An Introduction to Ontologies and Ontology, Hardcover: Ontologier in urban development projects, 2011.

[17] M. M. Taye, "Web-Based Ontology Languages and its Based Description Logics," The Research Bulletin of Jordan ACM, Jordan.

[18] T. R. Gruber, "TOWARDS PRINCIPLES FOR THE DESIGN OF ONTOLOGIES USED FOR KNOWLEDGE SHARING," in *nternational Workshop on Formal Ontology*, 1993.

[19] E. M. Gontier, "Web Semantic and Ontology," Scientific Research Publishing Inc., Paris,, 2015.

[20] I. Horrocks, Ontologies and Databases, Oxford University Computing Laboratory.

[21] Open Clinical, "Knoledge management for health care," Open Clinical, [Online]. Available: http://www.openclinical.org/ontologies.html. [Accessed 18 05 2017].

[22] R. K. Saripalle, Current status of ontologies in Biomedical and Clinical Informatics, Storrs: University of Connecticut.

[23] OBO Foundry, "The OBO Foundry," OBO Foundry, [Online]. Available: http://obofoundry.org/. [Accessed 18 05 2017].

[24] OBO Foundry, "Diseases Ontology," OBO Foundry, [Online]. Available: http://www.disease-ontology.org/. [Accessed 18 05 2017].

[25] OBO Foundry, "Symptom Ontology," OBO Foundry, [Online]. Available: http://symptomontologywiki.igs.umaryland.edu/mediawiki/index.php/Main_Page. [Accessed 18 05 2017].

[26] O. Mohammed, R. Benlamri and S. Fong, "Building a Diseases Symptoms Ontology for Medical Diagnosis: An Integrative Approach," in *Conference: IEEE International Conference on Future Generation Commnication Technology (FGCT 2012)*, 2012.

[27] S. El-Sappagh1 and F. Ali, "DDO: a diabetes mellitus diagnosis ontology," in *El-Sappagh and Ali Appl Inform*, 2016, 2016.

[28] W. C. a. B. S. Richard H. Scheuermann, "Toward an Ontological Treatment of Disease and Diagnosis," in *AMIA Summit on Translational Bioinformatics*, San Francisco, 2009.

[29] S. Z. B. B. M. H. Heiner Oberkampf, "Interpreting Patient Data using Medical Background Knowledge," in *SIEMENS AG, Corporate Technology*, Munich.

[30] Wikipedia, "Java Programming Language," Wikipedia the free Encyclopedia, [Online]. Available: https://en.wikipedia.org/wiki/Java_(programming_language). [Accessed 19 05 2017].

[31] M. Rouse, "SQL (Structured Query Language)," [Online]. Available: http://searchsqlserver.techtarget.com/definition/SQL. [Accessed 19 05 2017].

[32] W3C Semantic Web, "Web Ontology Language (OWL)," W3C, 12 11 2012. [Online]. Available: https://www.w3.org/OWL/. [Accessed 19 05 2017].

[33] V. Beal, "SPARQL," webopedia, [Online]. Available: http://www.webopedia.com/TERM/S/SPARQL.html. [Accessed 19 05 2017].

[34] Wikipedia, "Semantic_Web_Rule_Language," Wikipedia the free Encyclopedia, [Online]. Available: https://en.wikipedia.org/wiki/Semantic_Web_Rule_Language. [Accessed 19 05 2017].

[35] ORACLE, "What is JavaFx?," ORACLE, [Online]. Available: http://docs.oracle.com/javafx/2/overview/jfxpub-overview.htm. [Accessed 19 05 2017].

[36] techopedia, "model-view-controller-mvc," techopedia, [Online]. Available: https://www.techopedia.com/definition/3842/model-view-controller-mvc. [Accessed 19 05 2017].

[37] Wikipedia, "NetBeans," Wikipedia the free Encyclopedia, [Online]. Available: https://en.wikipedia.org/wiki/NetBeans. [Accessed 19 05 2017].

[38] WampServer, "WampServer," WampServer, [Online]. Available: http://www.wampserver.com/en/. [Accessed 19 05 2017].

[39] Tach Terms, "MySql," Tach Terms, [Online]. Available: https://techterms.com/definition/mysql. [Accessed 19 05 2017].

[40] ORACLE, "JavaFX Scene Builder," ORACLE, [Online]. Available: http://www.oracle.com/technetwork/java/javase/downloads/javafxscenebuilder-info-2157684.html. [Accessed 19 05 2017].

[41] Wikioedia, "Jena framework," Wikipedia the free Encyclopedia, [Online]. Available: https://en.wikipedia.org/wiki/Jena_(framework). [Accessed 19 05 2017].

[42] Wikipedia, "Resource Description Framework," Wikipedia the free Encyclopedia, [Online]. Available: https://en.wikipedia.org/wiki/Resource_Description_Framework. [Accessed 19 05 2017].

[43] Wikipedia, "Protégé," Wikipedia the free Encyclopedia, [Online]. Available: https://en.wikipedia.org/wiki/Prot%C3%A9g%C3%A9_(software). [Accessed 19 05 2017].

[44] businessdictionary, "businessdictionary," [Online]. Available: http://www.businessdictionary.com/definition/diagnosis.html. [Accessed 14 05 2017].

[45] R. Elmasri, "FUNDAMENTALS OF DATABASE SYSTEMS," The University of Texas at Arlington, Texas , 2011,.

Annex N°1: Diabetes anemia symptoms and diagnosis

Source: (https://www.thediabetescouncil.com/diabetes-and-anemia-are-they-related/)

The symptoms of diabetes and anemia can be very similar, so it may be difficult to know if a person has both. With mild anemia, there may be no signs of the disease. Once the disease progresses, symptoms may include:

- Tiredness
- Dizziness
- Depression
- Shortness of breath
- Weakness
- Irritability
- Brittle nails
- Pale skin
- Cold hands and feet
- Numbness in the hands and feet
- Chest pain
- Irregular heartbeat
- Hard time concentrating
- Craving ice

Diagnosis

A doctor will initially assess your symptoms and then do blood work to see if you are anemic. Normal hemoglobin for men is between 13.8 and 17.2 while for women it is between 12.1 and 15.1. Hematocrit ranges for men are 40.7 to 50.3 and 36.1 to 44.3 for women. For children, the range changes according to their age. Values with each laboratory test will vary to some extent, so do not be alarmed if they are different.

After establishing low counts, the next step is to figure out what is causing the problem. The doctor may ask questions about your family history, your diet, and the medications that you are taking. They will also do an assessment to see if you have any physical signs that could point to the cause.

Other tests that can be done include:

- Blood work to check ferritin and iron levels or vitamin levels
- A colonoscopy or esophagogastroduodenoscopy to check for internal bleeding
- Occult stool tests to see if there is blood in your stool
- Blood work to check the function of the kidneys

If your doctor still can’t figure what the cause is, you may be referred to a physician that specializes in blood. They are called hematologists and may run more advanced tests to find the cause.

Annex N°2: Link Anemia Symptoms to SYMP ontology

N°	Anemia Symptom	Class ID in SYMP	Literal (Label)
01	Tiredness	SYMP_0000187	tiredness
02	Dizziness	SYMP_0000610	dizziness
03	Depression	SYMP:0000022	depression
04	Shortness of breath	SYMP:0000868	breathing problems
05	Weakness	SYMP:0000177	weakness
06	Irritability	SYMP:0000654	irritability
07	Pale skin	SYMP_0000510	pallor
08	Numbness in the hands and feet	SYMP_0000435	paresthesia
09	Chest pain	SYMP_0000576	chest pain
10	Irregular heartbeat	SYMP_0000287	arrhythmia
11	Hard time concentrating	SYMP_0000817	inability to concentrate
12	Diaphoresis	SYMP_0019152	diaphoresis
13	Rapid heart beat	SYMP:0000668	abnormal heart rhythms

Anemia in Disease ontology DOID: **Anemia**

ID = DOID_2355

Annex N°3: Diabetes Mellitus Signs and Symptoms

There are three main types of diabetes:

Type 1 Diabetes: About 5 to 10 percent of those with diabetes have type 1 diabetes. It's an autoimmune disease, meaning the body's own immune system mistakenly attacks and destroys the insulin-producing cells in the pancreas. Patients with type 1 diabetes have very little or no insulin, and must take insulin everyday. Although the condition can appear at any age, typically it's diagnosed in children and young adults, which is why it was previously called juvenile diabetes.

Type 2 Diabetes: Accounting for 90 to 95 percent of those with diabetes, type 2 is the most common form. Usually, it's diagnosed in adults over age 40 and 80 percent of those with type 2 diabetes are overweight. Because of the increase in obesity, type 2 diabetes is being diagnosed at younger ages, including in children. Initially in type 2 diabetes, insulin is produced, but the insulin doesn't function properly, leading to a condition called insulin resistance. Eventually, most people with type 2 diabetes suffer from decreased insulin production.

Gestational Diabetes: Gestational diabetes develops during pregnancy. It occurs more often in African Americans, Native Americans, Latinos and people with a family history of diabetes. Typically, it disappears after delivery, although the condition is associated with an increased risk of developing diabetes later in life.

If you think that you have diabetes, visit your doctor immediately for a definite diagnosis. Common symptoms include the following:

- Frequent urination
- Excessive thirst
- Unexplained weight loss
- Extreme hunger
- Sudden vision changes
- Tingling or numbness in the hands or feet
- Feeling very tired much of the time
- Very dry skin
- Sores that are slow to heal
- More infections than usual

Some people may experience only a few symptoms that are listed above. About 50 percent of people with type 2 diabetes don't experience any symptoms and don't know they have the disease.

Annex N°4: Link Diabetes mellitus Type 2 Symptoms to SYMP ontology

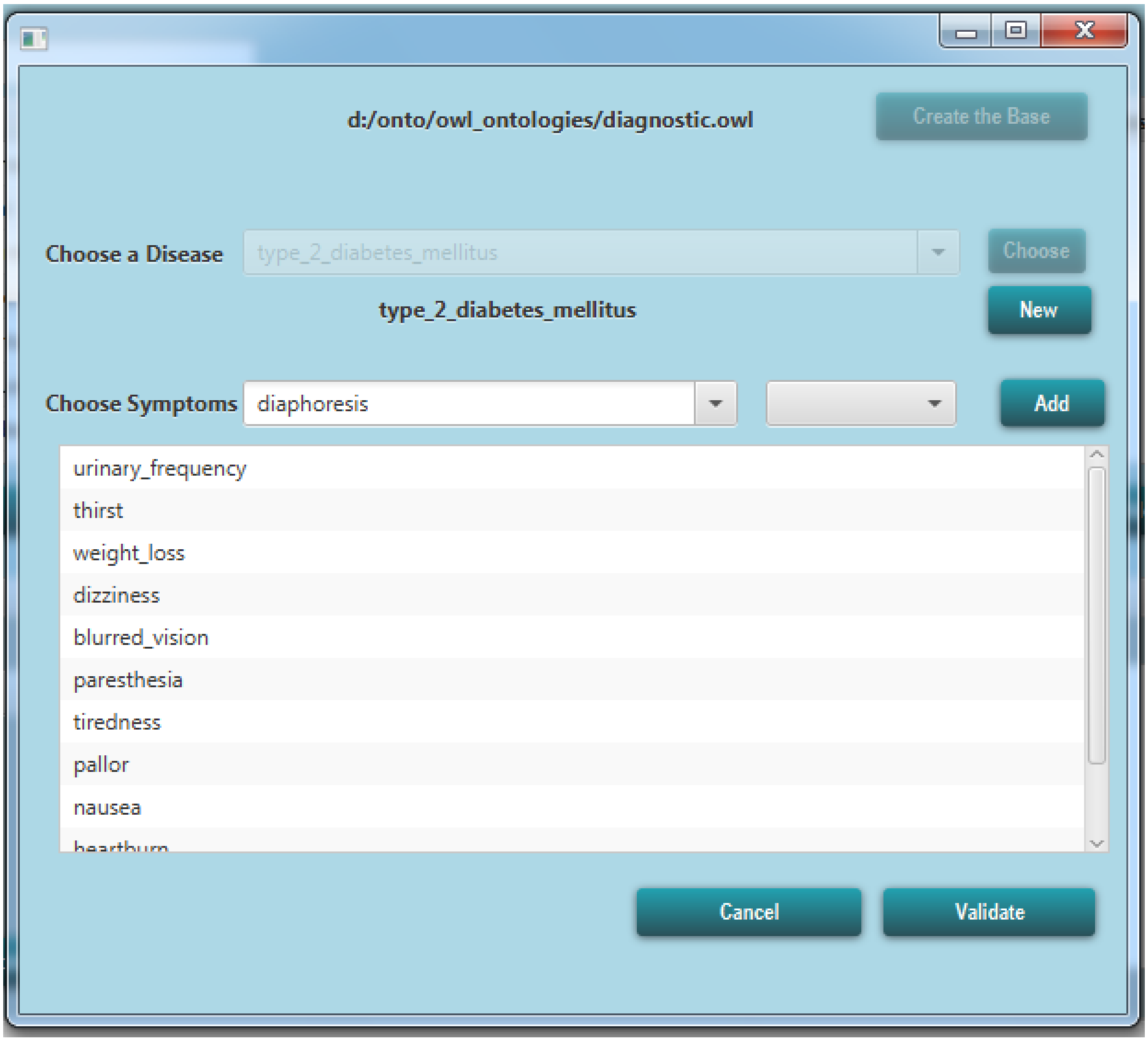
N°	Diabetes Symptom	Class ID in SYMP	Literal (Label)
01	Frequent urination	SYMP_0000563	urinary frequency
02	Excessive thirst	SYMP_0000156	thirst
03	Unexplained weight loss	SYMP_0000178	weight loss
04	Dizziness	SYMP_0000610	dizziness
05	Sudden vision changes	SYMP_0000012	blurred vision
06	Tingling or numbness in the hands or feet	SYMP_0000435	paresthesia
07	Feeling very tired much of the time	SYMP_0000187	tiredness
08	Very dry skin	SYMP_0000510	pallor
09	Nausea	SYMP_0000458	nausea
10	Heartburn	SYMP_0000439	heartburn
11	Diaphoresis	SYMP_0019152	diaphoresis

Diabetes mellitus type 2 in Disease ontology DOID: **type 2 diabetes mellitus**

ID = DOID_9352

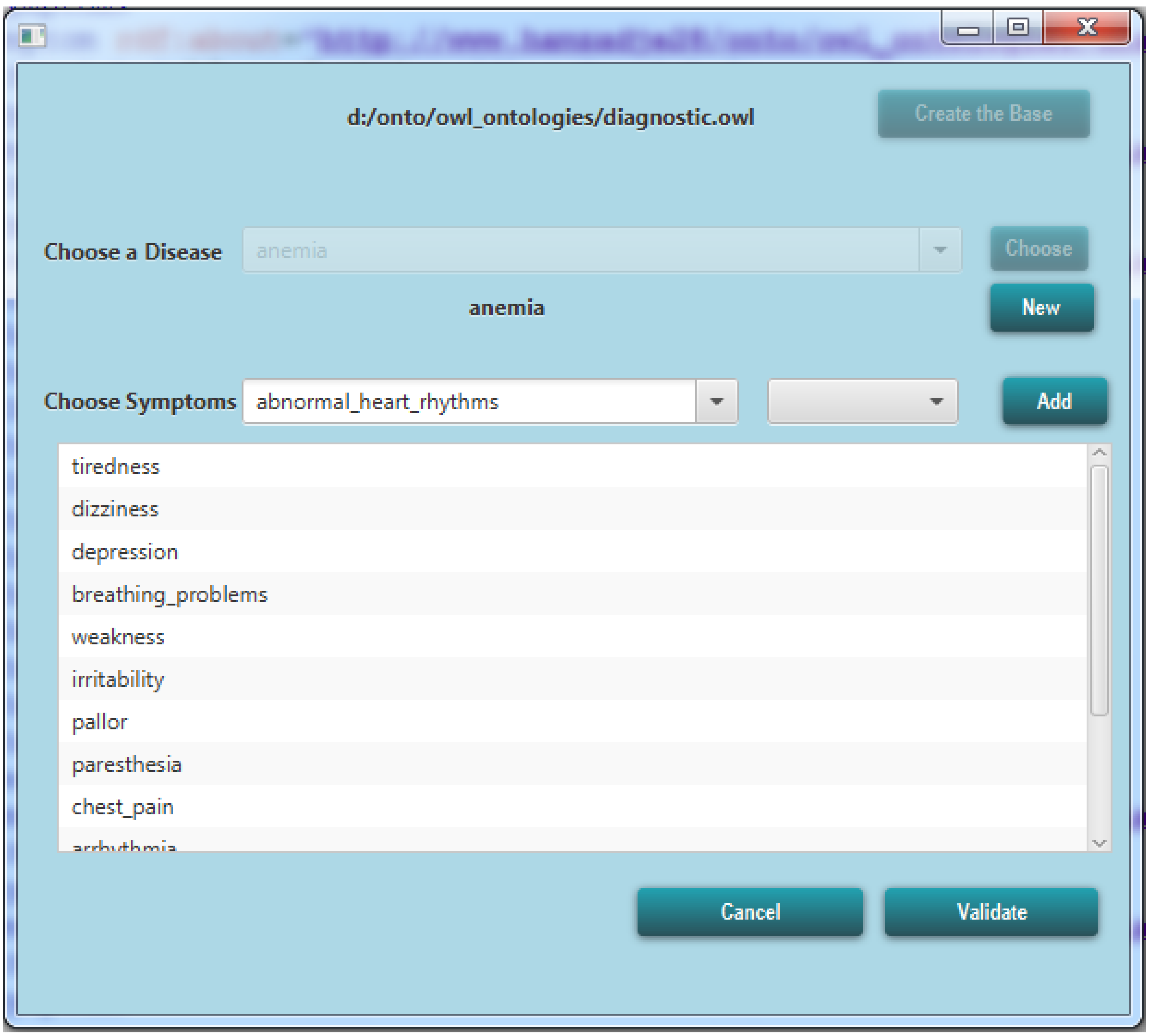
Annex N°5: Generate Disease-Symptoms using EasyDiag

Diabetes mellitus type 2

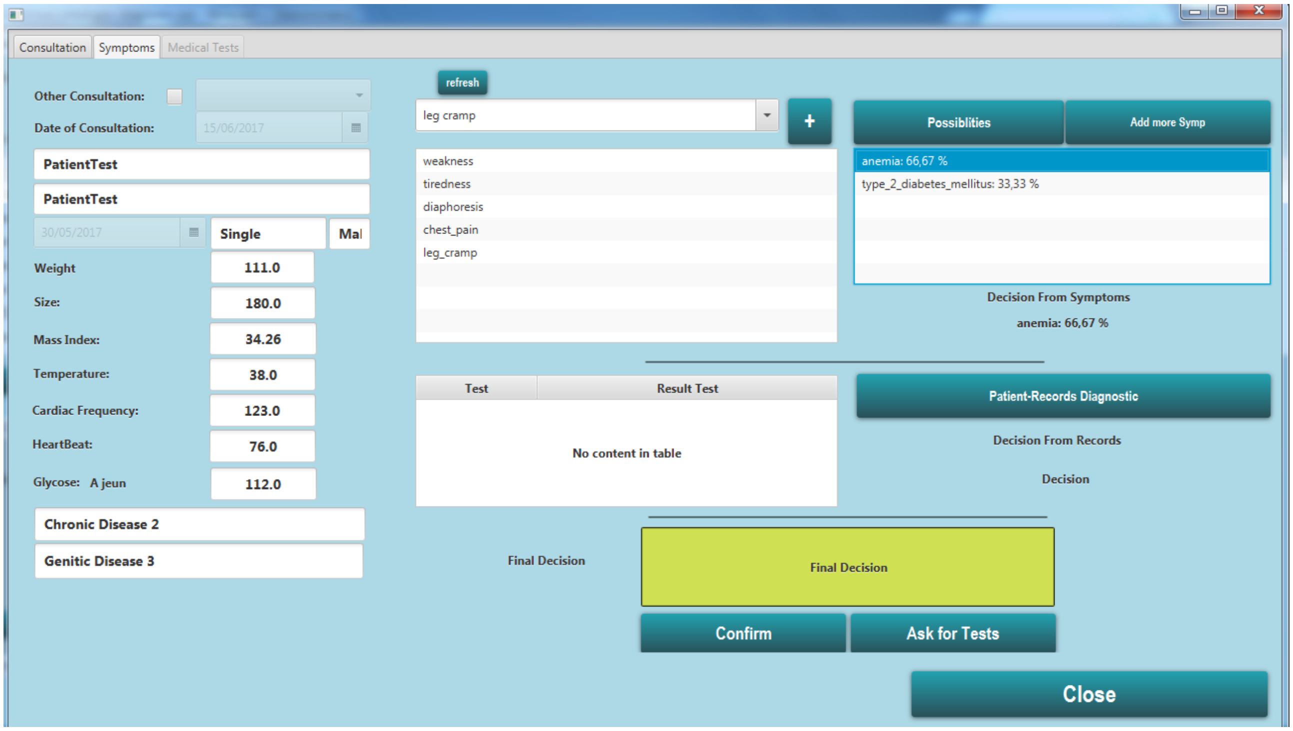


Annex N°6: Generate Disease-Symptoms using EasyDiag

Anemia



Annex N°7: Diagnose a disease using EasyDiag



Patient symptoms:

1. Weakness
2. Tiredness
3. Diaphoresis
4. Chest pain
5. Leg cramp

Diagnosis:

Anemia: 67.67 %

Diabetes mellitus type 2: 33.33 %

ملخص:

هذا العمل هو عبارة عن نظام دعم القرار الطبي، و يهدف إلى مساعدة الأطباء في عملية التشخيص الطبي واتخاذ القرار، وذلك باستخدام أنتولوجيات طبية SYMP (أنتولوجيا الأعراض) و DOID (أنتولوجيا الأمراض)، و كذا أنتولوجيا سجلات المريض المنشأة انطلاقا من تاريخ المريض المخزن في قاعدة البيانات.

و مقاربتنا تعتمد على:

- ربط الأمراض بأعراضها و توليد أنتولوجيا جديدة للتشخيص (DIAGNOSTIC).
- إنشاء أنتولوجيا جديدة من قاعدة بيانات المريض والتي تحتوي كل سجلاته (RECORD).
- استنتاج علامات من الأنتولوجيا RECORD باستخدام قواعد SWRL.
- اتباع إجراءات التشخيص الطبي ومساعدة الطبيب في تشخيص المرض.

الكلمات المفتاحية: عملية التشخيص، التشخيص الطبي، أنتولوجيا، أمراض، أعراض، DOID, SYMP, SWRL.

Abstract:

This work is a clinical decision support system (CDSS) consists to help physicians in medical diagnosis process and decision-making, using medical ontologies SYMP (symptoms) and DOID (diseases), and patient's record ontology which is created from the patient history stored in a database.

Our approach depends on:

- Linking diseases to its symptoms and generate a new ontology (DIAGNOSTIC).
- Creating a new ontology from the patient's database which contain their records (RECORD).
- Infer more signs from RECORD ontology using SWRL (Rules).
- Follow medical diagnostic procedures, and help physician to diagnose the disease.

Keywords: Diagnostic, Medical Diagnosis, Ontology, Disease, Symptom, DOID, SYMP, SWRL.

Résumé :

Ce travail est un système de support de décision médicale, qui consiste à aider les médecins au processus de diagnostic médical et la prise de décision, en utilisant les ontologies médicales SYMP (symptômes) and DOID (maladies), et l'ontologie des recordes des patients, qui a été créé à partir de l'historique des patients enregistrée dans une base de données.

Notre approche consiste à :

- Relier les maladies aves leurs symptômes et la génération d'une ontologie (DIAGNOSTIC).
- Créer une ontologie de la BDD des patients qui contient leurs enregistrements (RECORD).
- Déduire des signes depuis l'ontologie RECORD en utilisant les règles (SWRL).
- Suivre la procédure de diagnostic médical et aider le médecin à découvrir la maladie.

Mots Clés : Diagnostic, Médical, Ontologie, Maladie, Symptôme, DOID, SYMP, SWRL.