Building an ontology for diagnosing Sidr tree diseases

Zainab M. Jiwar a, Zainab I. Othman b

a Department of Computer Information Systems, College of Computer Science and Information Technology, University of Basrah, Basrah, Iraq.
Email: itpg.zainab.jiwar@uobasrah.edu.iq

b Department of Computer Information Systems, College of Computer Science and Information Technology, University of Basrah, Basrah, Iraq.
Email: zainab.othman@uobasrah.edu.iq

ABSTRACT

Sidr trees are among the important trees in Iraq, especially in the southern regions of the country, as in Basra Governorate. The Sidr crop contributes a good share of the economy in many regions. In addition, it is used in various fields such as medicine. Various pathogens affect the Sidr tree due to many diseases, which cause serious problems that weaken or stop the production of the plant, and may eventually cause the plant to die. Therefore, we find that direct work, whether at the individual level of farmers or institutions specialized in agricultural prevention, for the development of solutions to detect and diagnose diseases quickly, with high accuracy, and recommend treatment for Sidr diseases is necessary and inevitable. In this work, we build an ontology to represent information about Sidr tree diseases. The proposal of this ontology is to support agricultural practices and systems geared towards helping farmers in the early prediction of diseases from their morphological symptoms. The ontology was developed under Protégé 5.5.0 using the Web Ontology Language (OWL) format and defined Competency Questions, DL-Query, and SPARQL queries. It includes 217 classes, 13 object properties, 6 data properties, and 1762 axioms. Experiments conducted through a data set showed the effectiveness of ontology in diagnosing Sidr tree diseases using one or more observations of symptoms provided by farmers. As a contribution to this work, it presents the first ontology to recover knowledge about the diseases of the Sidr tree and the possibility of using this ontology in designing easy-to-use computerized systems (relying on semantic web technologies) that help the farmer in diagnosing diseases and suggesting appropriate treatment quickly, accurately and at a lower cost.

MSC..

https://doi.org/10.29304/jqcm.2023.15.1.1141

1. Introduction

Sidr tree (or Ziziphus Spina – Christi), also known as Nabkh, Ber, Jujube or Christ's Thorn tree is a plant that belongs to the family Rhamnaceae and the genus Zizyphus, which includes several types of deciduous and evergreen trees and shrubs that spread in tropical and subtropical regions of the world [1]. Sidr fruits are primarily seasonal

*Corresponding author

Email addresses:

Communicated by ‘sub etitor’
and available in abundance at certain times of the year. They are mainly consumed fresh and in dried form and have a high nutritional value that boosts energy and strengthens immunity. It has been used in the confectionery industry. The leaves, on the other hand, are rich in calcium, iron, and magnesium. Sidr trees are used as a source of animal fodder, and as ornamental trees in home gardens. All of these benefits and others have made it the focus of researchers’ attention [2, 3, 4, 5, 6]. The fruit has excellent medicinal value; it helps with digestion and blood purification. The plant’s seeds, roots, and stems also have medicinal uses [7].

Sidr trees are among the important trees in Iraq, especially in Basra Governorate, and are planted either overlapping with palm trees or individually and in home gardens. Sidr orchards are also used for honey bee breeding in addition to the production of fruits. Recently, Sidr trees have received great attention from many, and that is for their high ability to withstand drought conditions and wide adaptation to different soils.

Sidr trees are distinguished by their great diversity in the different ecosystems in which they grow, such as dry, semi-arid, saline, and desert environments, and they have widespread in Iraq, especially in the southern region. Statistics from the Department of Agriculture in Basra Governorate indicate that there are 116060 Sidr shrubs distributed in different areas of the province (Planning and Follow-up. 2018). Basra Agriculture Directorate. Basra Governorate. Iraq. The crop contributes a good share of the economy in many localities.

There are a variety of pathogens affecting the Sidr trees and can be considered as a cause of numerous diseases, which impair or stop plant production and may ultimately cause the plant to die. Thus, it is very important to diagnose these diseases and then suggest a proper treatment.

The term plant disease diagnosis can be defined as the process of determining the cause of plant diseases. This process is considered one of the most important priorities in plant protection. The process of diagnosing plant diseases can be carried out through many or few steps, long or short, and that depends on the ability of the person making the diagnosis and the nature of the disease. In general, the diagnosis is made through two methods, the first is the direct method and the second is the indirect method. In the direct method, farmers collect real samples from the affected parts of the plant and take them to the laboratory for the purpose of examination, analysis, verification, and diagnosis of disease. The indirect diagnosis method, in which a set of data is collected from the infected field of plants, and then using a specific mechanism such as the use of an ontology, through which the disease is diagnosed [8]. Within the field of computer and information sciences, ontology consists of the concepts through which the field of knowledge for which this ontology is built is modeled. The basic concepts in any ontology represent classes, attributes, and relationships between members of those classes [9].

In this research, a special ontology has been built as a knowledge base for diagnosing support systems, relying on semantic web technologies. Then ontology can help farmers diagnose the correct disease and suggest the right treatment, thus avoiding the disease outbreak.

2. Related Works

Many studies are carried out to help farmers for making decisions regarding the control, prevention, and management of plant diseases. Some of these studies are based on the use of the concept of ontology. Here are some of these studies:

Mahmoud A. El-Askary [10] developed an approach consisting of three interrelated components: creating a knowledge base, using a reasoning engine, and writing server-side applications. His approach is used to diagnose date palm diseases and pests and suggest appropriate treatment by identifying unusual disease signs on any part of the date palm.

By using the disease’s symptoms, a group of researchers (Watanee Jearanaiwongkul, Chutiporn Anutariya, and Frederic Andres) [11], show how to identify plant diseases from their existing abnormalities (symptoms). In their first paper, they proposed an ontology-based approach for modeling plant diseases and then show their approach by developing a rice disease ontology. And in their second paper [12], An expert system, called RiceMan, was designed and developed for rice disease identification and control recommendation.
Ontology knowledge-based for durian pests and diseases retrieval system is adopted by Porawat Visutsak [13]. The important come out of his work is a system consisting of the stored knowledge of durian pests and diseases and the diagnosis of durian diseases and suggestions for disease treatments.

Rusul Y. Al-Salhi and Abdulhussein M. Abdullah [14] developed an ontology of Quranic stories depending on the MappingMaster domain-specific language (DSL) technology, through which concepts and individual data are loaded and linked automatically to the ontology from pre-prepared Excel sheets. They used object role modeling (ORM) language to build the conceptual structure. Testing and evaluating their ontology is done through the use of SPARQL query language by asking many competency questions. Then they prove that their ontology answered all those questions well.

In this work, a Sidr tree disease ontology is built to be the first ontology in the field of Sidr tree planting. And then the produced ontology knowledge base is used in a web application for diagnosing and suggesting the proper treatment for the diagnosed diseases.

3. Sidr tree diseases diagnosis ontology

The term "ontology" was first used in philosophy. It focuses on figuring out what sorts of objects genuinely exist and how to characterize them [15]. The most often-used definition of ontology is an “explicit and formal specification of a conceptualization” [16]. Ontology, however, has recently been given a technical definition in computer science that differs significantly from its original meaning. “Ontology is a methodology for describing the domain of knowledge structure in a specific area”. In general, ontology consists of a finite list of terms called classes and relationships (or properties) linking these terms together. In this paper, this last definition will be adopted. In building an ontology some requirements (domain and scope), are required with certain steps that have to be followed as in the next section.

3.1. Ontology requirements

Developing an ontology to facilitate a knowledge base for the Sidr tree disease diagnosis system must only focus on the domain of Sidr tree diseases and cover all abnormalities that may appear on Sidr trees and are observable by human eyes. The required ontology is designed to model the following information and relationships:

• The ontology consists of three main categories: Tree part, abnormality, and disease diagnosis.

• The disease diagnosis categorizes different groups of diseases depending on and according to the input observations.

• A tree part categorizes different tree components, from which trees are made up.

3.2. Competency questions

The competency questions (CQs) that the designed ontology must be able to answer are listed as follows:

1. What are the probable diseases if an abnormality X is observed on the Sidr tree?
2. What is the probable disease if a Sidr tree has abnormality X, Y, ...?
3. What is the probable disease if a Sidr tree has abnormality X, Y, ... on the ‘leaves’ part of the tree?
4. What are all abnormalities that could have been diagnosed on the X Sidr tree part?
5. What are all probable diseases that could have been diagnosed on a Sidr tree?
6. What is the probable Pathogen if the disease diagnosis on the Sidr tree is X?
7. What are the probable diseases that are caused by unfavorable environmental conditions?
The above-mentioned CQs have identified two primary purposes: disease diagnosis and disease retrieval. CQs (1-3) focus on the diagnosis of the disease that occurs based on the symptoms observed. The ontology must be able to answer user query that does not give complete information about the occurring disease. Whereas CQs (4-7) focuses on the ability of the built ontology to retrieve interesting knowledge related to diseases.

### 3.3. Elements of ontology

The competency questions (CQs) that the designed ontology must be able to answer are listed as follows:

- **Classes**: the class concept is a way of meaningfully grouping things from a domain into categories. Most ontologies center on this. A class (a category) of diseases, for instance, encompasses the whole of the diseases (or all diseases) as a superclass. Subclasses of the superclass can represent concepts that are more specialized than those represented by the superclass. For instance, we can separate the pathogens of tree disease into microorganisms and pests. Then, microorganisms can be further separated into fungi and bacteria, etc. as depicted in Fig 1.

- **Properties**: express the relationships between ontology classes, instances, literals, etc. Properties can be divided into two main categories:
  - **Object properties**: show the relationships between instances. For example, between instances of the two classes: *SidrDisease* class and the *Pathogens* class is the object property *hasFactor*.
  - **Data properties**: these are the linking properties between instances and data values. For example, the class *SidrDisease* has a property called *Name* that relates the *SidrDisease* class to a *string* value.

- **Individuals**: things that have been placed into the class and not classes themselves. They are instances of classes and properties that incorporate particular or actual things from the domain, like tree parts, diseases, etc.

![Fig. 1 - Superclass—Subclass example](image)

### 3.4. Ontology Building

By following Noy and McGuinness’ methodology [17], and using the editor Protégé (version 5.5.0), which is developed by Stanford University [18], the building of an ontology may be done through the following seven steps:

Step 1: Deciding ontology domain and ontology scope.
Step 2: Studying the reusing of existing ontologies.
Step 3: Identifying the important terms in the ontology.
Step 4: Describe the classes and their hierarchy.
Step 5: Clarifying the properties of the classes.
Step 6: Defining the facets of the classes-slots.
Step 7: Initiating instances.

In addition to the above methodology, Sidr tree diseases ontology building is mainly based on collected data from many resources including research papers, reports, books, and articles in the agricultural domain such as the internet and also from national and international agriculture universities. The resulting ontology consists of 217 classes, 13 object properties, and 6 data properties. Some instances are also, added to the ontology for illustration. The resultant Ontology can be described as follows:

− Classes:
  • Abnormality: in Sidr ontology, this class is defined by three possibilities from what a farmer can normally observe on the appearance of the diseased tree. These abnormalities are obtained from the answers to the following questions:
    1. What is the type of abnormality? (The answer for example could be: The infected area becomes slightly raised and rough.),
    2. What is the color of the abnormality? (The answer for example could be: Yellowing of leaves.), and
    3. What is the shape of the abnormality? (The answer for example could be: Partial shedding of leaves occurs.).
    These are the widely noticeable signs or symptoms that may appear on the infected Sidr tree. And according to them a creation of three subclasses for representing the abnormality of the Sidr tree diseases has been done. The three subclasses are Symptom, Color, and Shape.
  • Cases: This class determines the disease case for each disease and its diagnosis.
  • Diagnosis: this class determines the diagnosis of the disease according to observed abnormalities.
  • Pathogens: This class determines different types of tree pathogens such as microorganisms, pests, and physiological pathogens. It has three subclasses which are Microorganism, Pest, and UnfavorableEnvironment. Moreover, the class Microorganism has two subclasses: Bacterium and Fungus, and class Pest has three subclasses Pest_Insect, Pest_Mite, and Pest_Nematode, and the class UnfavorableEnvironment has no subclass.
  • SidrDisease: This class describes a number of Sidr tree diseases. It is classified into three subclasses: Plant_Disease (including two classes: BacterialDiseases and FungalDiseases), Pest_damage (including three classes: Insect, Mite, and Nematode), and PhysiologicalDiseases.
  • SidrPart: This class categorizes different tree components, which trees are made up of, such as leaves, fruits, stems, etc.
  • Treatment: This class describes the proper treatment for the diagnosed diseases.

The hierarchical structure of the ontology classes is depicted in Fig 2.
Fig. 2 - Concepts/classes in the Sidr tree diseases ontology

- **Object properties:**
  - `factorOf`: characterizes the relationship between disease `Pathogen` class and `SidrDisease` class.
  - `has_part`: it shows the relationship between SidrPart class and all other tree part classes (Root, Branches, Florals, Leaf, Fruit, Stem, and Seeds classes). This property is the inverse of the `has_part` property.
  - `hasAbnormality`: characterizes the relationship between the Diagnosis class and Abnormality class. Every disease has one or more abnormality/ies.
  - `hasDiagnosis`: characterizes the relationship between the case of the disease (Cases) class and Diagnosis class.
  - `hasFactor`: it shows the relationship between SidrDisease class and the Pathogens class. It is the inverse property of the `factorOf` property.
  - `hasState`: characterizes the relationship between the SidrDisease class and Cases class.
  - `hasSymptom`: characterizes the relationship between SidrPart class and Abnormality class.
  - `hasTreatment`: describes the relationship linking the SidrDisease class and the Treatment class.
  - `location_in`: characterizes the relationship between the Abnormality class and SidrPart class.
  - `part_of`: characterizes the relationship between the tree parts classes (Root, Branches, Florals, Leaf, Fruit, Stem, and Seeds classes) and Sidrpart class.
  - `similarSymptom`: characterizes the relationship between the Abnormality class and itself. Two or more diseases may have the same abnormalities.
  - `StateOf`: it shows the relationship between the case of tree diseases Cases class and SidrDisease class. It is the inverse property of the `hasState` property.
  - `treatmentOf`: it shows the relationship between Treatment class and SidrDisease class. It is the inverse of the `hasTreatment` property.

Table 1, describes more details about the above-defined object properties.
Table 1 - Object properties in Sidr tree diseases ontology

<table>
<thead>
<tr>
<th>No.</th>
<th>Property</th>
<th>Domain</th>
<th>Range</th>
<th>Characteristics</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>factorOf</td>
<td>Pathogens</td>
<td>SidrDisease</td>
<td></td>
<td>Inverse property:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Root, Branches,</td>
<td></td>
<td>hasFactor</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Florals, Leaf,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Fruit, Stem,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>and Seeds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>has_part</td>
<td>SidrPart</td>
<td>Florals, Leaf,</td>
<td></td>
<td>Inverse property:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Root, Branches,</td>
<td></td>
<td>part_of</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Florals, Leaf,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Fruit, Stem,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>and Seeds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>hasAbnormality</td>
<td>Diagnosis</td>
<td>Abnormality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>hasDiagnosis</td>
<td>Cases</td>
<td>Diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>hasFactor</td>
<td>SidrDisease</td>
<td>Pathogens</td>
<td>Inverse functional</td>
<td>Inverse property:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>factorOf</td>
</tr>
<tr>
<td>6</td>
<td>hasSymptom</td>
<td>SidrDisease</td>
<td>Cases</td>
<td>Inverse functional</td>
<td>Inverse property:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>StateOf</td>
</tr>
<tr>
<td>7</td>
<td>hasState</td>
<td>SidrDisease</td>
<td>Treatment</td>
<td></td>
<td>Inverse property:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>treatmentOf</td>
</tr>
<tr>
<td>8</td>
<td>hasTreatment</td>
<td>SidrDisease</td>
<td>Treatment</td>
<td></td>
<td>Inverse property:</td>
</tr>
<tr>
<td>9</td>
<td>location_in</td>
<td>Abnormality</td>
<td>SidrPart</td>
<td>functional</td>
<td>Inverse property:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Root, Branches,</td>
<td></td>
<td>has_part</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Florals, Leaf,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Fruit, Stem,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>and Seeds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>part_of</td>
<td>Abnormality</td>
<td>SidrPart</td>
<td></td>
<td>Inverse property:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Root, Branches,</td>
<td></td>
<td>has_part</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Florals, Leaf,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Fruit, Stem,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>and Seeds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>similarSymptom</td>
<td>Abnormality</td>
<td>Abnormality</td>
<td>Transitive &amp; Symmetric</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>stateOf</td>
<td>Cases</td>
<td>SidrDisease</td>
<td>functional</td>
<td>Inverse property:</td>
</tr>
<tr>
<td>13</td>
<td>treatmentOf</td>
<td>Treatment</td>
<td>SidrDisease</td>
<td></td>
<td>hasState</td>
</tr>
</tbody>
</table>

- **Data properties:**
  - *Description:* assigning one or more sentences or paragraphs of text to detailed descriptions of some concepts of the ontology. The datatype of the *Description* data property is *xsd:Literal*.
  - *Image:* used to assign hyperlinks of images for clarifying some concepts of the ontology, such as *SidrDisease*. The datatype of the *Image* property is *xsd:Literal*.
  - *Name:* used to assign names for some concepts of the ontology. The datatype of the *Name* property is *xsd:Literal*.
  - *Scientific name:* used to assign scientific names for some concepts of the ontology. The datatype of the *Scientific_name* property is *xsd:Literal*.
  - *Value:* used to give values to some concepts of the ontology. The datatype of the *Value* property is *xsd:string*.

- **Individuals:**
  After completing the creation of the classes and defining the relationships between them based on the collected knowledge of resources in the agricultural field from the Internet, as well as from local and international agricultural universities, a group of individuals was inserted for demonstration, including 79 cases of *Abnormality* class, 21 diagnoses of Diagnosis class, 21 types of Pathogens class, 21 kinds of SidrDisease class, 14 parts of *SidrPart* class, 21 cases of *Cases* class, and 21 treatments recommendation of *Treatment* class. Fig 3, illustrates the individuals of class *SidrPart* visualized by Protégé.
3.5. **Ontology Implementation and Evaluation**

Implementation of the built Sidr ontology can be accomplished by the following services:

A) Sidr tree disease diagnosis, and

B) Sidr tree disease-related information retrieval.

**A. Disease Diagnosis**

The disease diagnosis service is done by inputting from 1 to 3 observed abnormalities on specific tree parts, after that, the system shows a list of disease names sorted descending upon the recurrences of each disease name in the list. The disease name at the head of the list can be selected as the nearest disease that can have the given abnormalities. The following scenarios are used as examples:

1. Suppose ‘shape9’ (shape9 which means Plants eventually die.) is observed on the Sidr tree, what are the diagnoses of the probable diseases?

2. Suppose ‘color2’ (color2 which means The leaves show yellowish and brownish discolorations on the upper surface and drop prematurely.), ‘shape4’ (shape4 which means Sooty tuft-like circular to irregular black spots on the underside of the leaves.), and ‘symptom2’ (symptom2 which means Sooty appearance covers the entire lower surface.) are observed on the Sidr tree, what is the probable disease?

3. Suppose ‘color16’ (color16 which means Yellowing of leaves.), ‘shape28’ (shape28 which means drying of leaves.), and ‘symptom9’ (symptom9 which means a white mealy secretion that covers the body with side secretions that vary in number from one type to another.) are observed on the ‘part1’ (part1 which means leaves), what are the probable diseases?

These three scenarios are compatible with the competency questions (cf. CQs 1-3) mentioned in section 3.2. Scenario 1 represents the situation where only one abnormality has been found on a Sidr tree whereas scenarios 2 and 3 represent a situation where more than one abnormality has been found on a Sidr tree.

In order to verify and validate the disease diagnosis service, we implement the Description Logic Query (DL-Query) [19], a standard Protégé plugin built on the Manchester OWL syntax. The following querying example shows the first scenario.

Example:

- The scenario: Suppose ‘shape9’ is observed on the tree, what are the diagnoses of the probable diseases?
- DL-Query: “Diagnosis and hasAbnormality value shape9”. The result of this DL-Query is depicted in Fig 4.
B. Disease-related Knowledge Retrieval

The ontology can answer all competency questions and retrieve knowledge related to Sidr tree diseases in various dimensions. The following querying example represents CQ 4, which refers to the retrieval of all abnormalities that can occur in a particular Sidr tree part by using the SPARQL query [20].

Example:

- The competency question (CQ 4): What are all abnormalities that could have been diagnosed on "A" Sidr tree part?
- SPARQL-Query and its results are depicted in Fig 5, as an example of implementing the service of disease-related knowledge retrieval.

4. Conclusion

Many knowledge management and decision support systems are based on ontologies. Ontology is considered an effective technique for representing the knowledge domain. This work presents the ontology for diagnosing...
diseases of the Sidr tree. It can serve as the knowledge base of a web-based application for supporting farmers’ decisions in diagnosing the disease quickly and accurately and implementing the recommended treatment for the diagnosed disease in order to protect Sidr trees. The ontology permits the inference of new knowledge and it also takes into consideration the semantics of concepts in different terminology or even in several languages. Also, it lets easy combination and reuse of knowledge from multiple sources over the Web. The modeled ontology has been built mainly on collected data from different resources including research papers, reports, books, and articles in the agricultural domain from the internet and also from local and global agriculture institutes, universities and organizations. The obtained ontology uses the abnormalities of Sidr tree diseases in terms of symptoms, colors, shapes, and infected tree parts. The relationships between Sidr tree diseases and abnormalities are defined by subclass relations. The resulting ontology consists of 217 classes, 13 object properties, and 6 data properties. In order to identify Sidr tree disease, DL queries and reasoners are used to find subclasses of farmers’ observations or abnormalities. The future goals are to use this ontology as a knowledge base in a web application to retrieve information about Sidr trees and speed up the process of diagnosing diseases and suggesting appropriate treatment for them.

Acknowledgments

The authors extend their thanks to Professor Muhammad Fayyad, a specialized professor from the University of Basrah / College of Agriculture, for his evaluation and verification of the validity and accuracy of the collected data and the possibility of using them in designing the contents of the computer programs for the diagnosis of plant diseases.

References