Domain-Specific Ontology of Botany

Fang Gu, Cun-Gen Cao, Yue-Fei Sui, and Wen Tian

Key Lab of Intelligent Information Processing, Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100080, P.R. China
E-mail: {fgu, cgcao, yfsui, wtian}@ict.ac.cn
Received May 30, 2002; revised August 29, 2003.

Abstract Domain-specific ontologies are greatly useful in knowledge acquisition, sharing and analysis. In this paper, botany-specific ontology for acquiring and analyzing botanical knowledge is presented. The ontology is represented in a set of well-defined categories, and each concept is viewed as an instance of certain category. The authors also introduce botany-specific axioms, an integral part of the ontology, for checking and reasoning with the acquired knowledge. Consistency, completeness and redundancy of the axioms are discussed.

Keywords domain-specific ontology, botany, ontological analysis, knowledge-based inference, consistency, completeness, redundancy

1 Introduction

Botany is an important scientific domain, and is rich in botanical knowledge. Botanical knowledge is necessary in many knowledge-intensive applications, e.g., intelligent tutoring, question answering, natural language processing, and botanical analysis. BKB (Botany Knowledge Base) is currently the largest, hand-coded botanical knowledge base in the world. However, it is lacking of an analytical tool for checking the problems in the knowledge base and in that it does not adopt any well-established biological classification schemes (e.g., phylum, class, order and family). This makes it hard to be used, e.g., for intelligent tutoring.

We have been acquiring botanical knowledge from several knowledge sources to build a sharable botany knowledge base. In developing such a knowledge base, we have designed a domain-specific ontology for botany, and been extracting botany knowledge from the sources.

Ontology is fundamental in conceptualizing a domain, and it is also a semantic device for sharing knowledge across different domains and among different software agents and human beings. In our practice with ontological theory and application, we find that the existing techniques developed in the ontological engineering community lack specific and explicit semantic definition, and could not be easily shared by others.

The paper is organized as follows. Section 2 presents a multi-perspective structure of botany-specific ontology. Section 3 gives the overall description of the ontology. The representation of botanical ontology is in a natural frame format. Section 4 introduces axioms of the ontology. These axioms are necessary components of the ontology. They provide a basis for checking the botany knowledge to ensure consistency and accuracy and for reasoning with the botany knowledge base. Section 5 analyzes consistency, completeness and redundancy of the axioms themselves. Section 6 concludes the paper.

2 Multi-Perspective Ontology of Botany

Botany has several schemes of classification. Some schemes are similar, but others contradict with one another. It is biased for an ontology to choose one fixed scheme and ignore the others. Methodologically, we consider all possible classification schemes and common concepts (e.g., phylum, class, order, family) in our ontology.

2.1 Concepts, Attributes and Relations

The botanical knowledge space is composed by a set of concepts. The concepts are connected with relations. In other words, each concept is described with attributes and its relations with other concepts.

Definition 1. A concept C is a set of slot-value pairs. Slots are divided into two groups: attributes and relations.

\[ C = \{ (a_1, v_1), (a_2, v_2), \ldots, (a_n, v_n) \} \cup \{ (r_1, C_1), (r_2, C_2), \ldots, (r_m, C_m) \} \]

where \( a_1, a_2, \ldots, a_n \) are attributes and \( v_1, v_2, \ldots, v_n \) are their values, \( r_1, r_2, \ldots, r_m \) are relations, \( C_1, C_2, \ldots, C_m \) are concepts.

In Definition 1, a concept is defined. A concept can be defined according to any classification systems. For example, a concept can be defined for all plants, including those that survive in general, and its possible relations with other three large categories of botanical knowledge.
Fang Gu et al.: Domain-Specific Ontology of Botany

tributes and conceptual relations. Therefore, we also say a concept is a set of attribute-value pairs and relation-concept pairs. Formally,

\[
C = \{ \{a_1, v_1\}, \{a_2, v_2\}, \ldots, \{a_i, v_i\}\} \\
\cup \{(r_1, C_1), (r_2, C_2), \ldots, (r_j, C_j)\},
\]

where \(a_1, a_2, \ldots, a_i\) are called attributes of \(C\), \(v_1, v_2, \ldots, v_i\) are values of these attributes. Attributes and their values represent properties of the concept \(C\). \(C_1, C_2, \ldots, C_j\) are botanical concepts. \(r_1, r_2, \ldots, r_j\) are relations from \(C\) to \(C_1, C_2, \ldots, C_j\).

In Definition 1, each attribute \(a_i\) or relation \(r_j\) defines a perspective of a concept, and several attributes and relations describe an integrated view of the concept. Given values for each attribute and relation, a botanical concept is specified.

Botanical concepts can also be properly classified according to values of the attributes and relations. For example, \(lifecycle\) is a common attribute for all plants, indicating how long a plant can survive in general. From the perspective of \(lifecycle\) and its possible values, plants can be divided into three large categories: annual, biennial and perennial.

2.2 Multi-Perspective Structure of Botanical Concepts

The Swedish biologist Carolus Linnaeus was the firstfruitful biologist who worked on a classification system of all plants. The system was then elaborated by many other biologists. Now it is a considerably perfect taxonomy system. In this system, plants are divided mainly according to their morphological characteristics.

In our botany-specific ontology, the backbone is the taxa used by the whole biological science, e.g., kingdom, phylum, class, order, family, genus and species. This is one of the factors that compose our ontology domain-specific.

According to attributes and relations, botanical concepts are viewed in different perspectives. In some perspectives, new concepts other than the botany taxa, especially common concepts, are included. These concepts are divided according to attributes or relations that define the perspective. For other concepts in the system that cannot be divided according to the attributes or relations, the classification is according to other attributes or relations, or the classification is as that of the botany taxonomy system. With common concepts, the multi-perspective ontology forms a multi-dimensional hierarchical structure. Fig.1 is part of the ontological structure of Plantae from the perspective of \(lifecycle\).

In Fig.1, it can be seen that Plantae is divided into perennial, biennial and annual plants from the perspective of \(lifecycle\). In a multi-perspective ontology, one concept can be viewed in different perspectives, and divided into different sub-concepts.

In fact, concepts in different layers can be divided from different perspectives in one classification system. It is unreasonable to organize all the botanical concepts in different layers with the same attributes.

3 Botanical Ontology Representation

We have defined a series of categories to represent the botany-specific ontology. The categories are defined in a frame-based formalism, and they form a hierarchical structure according to the relations of concepts.

3.1 Category

First, we design a top category, called \(general\)-category. It contains attributes and relations that are shared by all kinds of botanical concepts. In other words, all of other botanical categories' definitions inherit attributes and relations defined in the \(general\)-category.

There are seven taxa in the botanical taxonomy system, and each taxon includes a set of corresponding botanical concepts. We design seven taxa categories to represent the corresponding concepts as shown in Fig.2. They form a hierarchical structure: lower-level categories inherit attributes and relations defined at higher levels. Fig.3(a) is the definition of \(phylum\)-category. This category is an extension of the \(kingdom\)-category. It inherits all the attributes and relations defined in the \(kingdom\)-category. Because \(kingdom\)-category extends \(general\)-category and inherits all of its attributes and relations, so \(phylum\)-category also in-
herit all of the attributes and relations of general-category.

Table 1. Common Botanical Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>alias</td>
<td>Alternate names of a plant</td>
</tr>
<tr>
<td>producing-area</td>
<td>Area that plant was used to live in</td>
</tr>
<tr>
<td>habitat</td>
<td>Type of environment for plant to live</td>
</tr>
<tr>
<td>lifecycle</td>
<td>Life-span of plant</td>
</tr>
<tr>
<td>pigment</td>
<td>Pigments contained in plant</td>
</tr>
<tr>
<td>ancestor</td>
<td>Ancestors of plant</td>
</tr>
<tr>
<td>number-of-species</td>
<td>Total number of species</td>
</tr>
<tr>
<td>number-of-existing-species</td>
<td>Number of the existing species</td>
</tr>
<tr>
<td>number-of-family</td>
<td>Number of families</td>
</tr>
<tr>
<td>number-of-chromosome</td>
<td>Number of chromosomes in plant</td>
</tr>
<tr>
<td>number-of-cotyledon</td>
<td>Number of cotyledons of plant</td>
</tr>
</tbody>
</table>

Fig. 2. Botanical taxa categories.

def-category phylum-category
    extends kingdom-category
    {attribute: alias : type string
     : value-type multiple-valued
     attribute: lifecycle : type number
     : value-type multiple-valued
     attribute: cell-type : type string
     : value-type single-valued ...
     relation: in-kingdom : type string
     : value-type single-valued ...
    }

def-frame Angiospermae: phylum-category
    {alias: Magnoliophyta
     lifecycle: 1, 2, 3, ..., cell-type: multi-cellular
     number-of-existing-species: 250,000
     number-of-family: 300~500
     number-of-cotyledon: 1 or 2 or 3 or 4
     habitat: forest, grassland, desert, environment of water
     ancestor: primarly angiosperm
     have-organism: flower
     breeding-manner: sexual reproduction, agamogenesis
     in-kingdom: Plant kingdom
     class-division: Monocotyledonae, Dicotyledonae
     not-similar-to: Gymnosperm ...
    }

Fig. 3. Definition of phylum-category and an example.

In Fig. 3(a), keywords in bold face before the colon represent meta-attributes of the category. The meta-attribute attribute defines an entity (e.g., lifecycle) as an attribute for botanical concepts, and the two facets type and value-type indicate the type of an attribute and the type of values the attribute can assume. Table 1 lists a few of the most common attributes of botanical entities. These attributes are also specified in the phylum-category.

Roots, stems, leaves, flowers, fruits and seeds are common components (i.e., organisms) of plants. Each of these organisms may have quite a few special attributes and relations. To represent them, we introduce six categories, i.e., root-category, stem-category, leaf-category, flower-category, fruit-category and seed-category. There are also some other kinds of botanical concepts. For each kind of concepts, we define a category to represent them.

For example, cell is a basic component of any plant. To represent botanical knowledge, we must represent plant cells, and we define a cell-category for this purpose.

To sum up, we have defined fifteen categories in our practice to represent the botanical knowledge. They are general-category, kingdom-category, phylum-category, class-category, order-category, family-category, genus-category, species-category, root-category, stem-category, leaf-category, flower-category, fruit-category, seed-category, and cell-category.

After the botanical categories are defined, we can view a botanical concept as an instance of certain category. For example, Angiospermae is one of the plant phyla, and it can be viewed as an instance of the phylum-category, as shown in Fig. 3(b). The underlined keywords are attributes and relations defined in the phylum-category.
3.2 Structural Representation

Structure of the ontology is represented by relations between botanical concepts. We divide structure of the ontology system into two parts: backbone structure and common classification structure. Each part of the structure is represented by specially defined relations.

3.2.1 Backbone Structure of Botany

Special relations are defined to describe the taxonomy among botanical concepts. They construct the backbone of the botany-specific ontology. These relations are shown in Table 2.

The ‘value’ of a hyponymic relation is a higher-level concept in the botany taxonomy system, and the ‘value’ of a hypernymic relation is lower-level concepts. For example, the order Nymphaeales is divided into 5 families, i.e., Nymphaonaceae, Nymphaeaceae, Barclayaceae, Cabombaceae and Ceratophyllaceae. So the value of family-division of Nymphaeales is the five families, and the value of in-phylum of each of these five families is Nymphaeales. In addition, we use phylum-classification (Nymphaeales, Nelumboaceae, Nympahea-ceae, Barclayaceae, Cabombaceae, Ceratophyllaceae) to indicate that Nelumboaceae, Nymphaeaceae, Barclayaceae, Cabombaceae, Cerato-phyllaceae are the value of phylum-classification of Nymphaeales (Representation of relations will be discussed in Subsection 4.1).

3.2.2 Other Classification Structures

Botanical concepts can also be classified according to many other different attributes. For example, according to lifecycle, Plantae is classified into perennial, biennial, and annual, as demonstrated in Fig.1. This kind of classification is as valuable as the taxonomy classification discussed previously, and is represented with two relations: generalization and specialization (see Table 3).

As discussed in Section 2, a concept can be divided into different sets of sub-concepts from different perspectives, and the value of specialization of a concept is different according to different attributes or relations. Therefore we use according-to as a necessary facet of relation specialization. Its value is attributes or relations that the classification is according to. Fig.4 gives the definition of according-to facet, together with an example of its application. Because generalization and specialization are actually applicable to all botanical concepts, we define them in the general-category.

### Table 2. Backbone Relations

<table>
<thead>
<tr>
<th>Type</th>
<th>Relation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyponymic</td>
<td>in-kingdom</td>
<td>Kingdom that the concept belongs to</td>
</tr>
<tr>
<td></td>
<td>in-phylum</td>
<td>Phylum that the concept belongs to</td>
</tr>
<tr>
<td></td>
<td>in-class</td>
<td>Class that the concept belongs to</td>
</tr>
<tr>
<td></td>
<td>in-order</td>
<td>Order that the concept belongs to</td>
</tr>
<tr>
<td></td>
<td>in-family</td>
<td>Family that the concept belongs to</td>
</tr>
<tr>
<td></td>
<td>in-genus</td>
<td>Genus that the concept belongs to</td>
</tr>
<tr>
<td>Hypernymic</td>
<td>phylum-classification</td>
<td>Phyla that belong to the concept</td>
</tr>
<tr>
<td></td>
<td>class-classification</td>
<td>Classes that belong to the concept</td>
</tr>
<tr>
<td></td>
<td>order-classification</td>
<td>Orders that belong to the concept</td>
</tr>
<tr>
<td></td>
<td>family-classification</td>
<td>Families that belong to the concept</td>
</tr>
<tr>
<td></td>
<td>genus-classification</td>
<td>Genera that belong to the concept</td>
</tr>
<tr>
<td></td>
<td>species-classification</td>
<td>Species that belong to the concept</td>
</tr>
</tbody>
</table>

### def-category general-category

(a)

### def-frame Plantae: kingdom-category

(b)

Fig.4. (a) Definition of according-to. (b) An example.
3.3 Part-Whole Relation Representation

Part-whole relation is one of the most important relations in the ontology of botany. For example, petal is a part of flower, flower is a part of rose, and willow has root, stems, and leaves as its parts. To represent the part-whole relation between concepts, we define the following relations (see Table 4). The relations of the part with other concepts (e.g., another part) are represented with relations defined in the category of the part. And we will discuss them in other articles.

<table>
<thead>
<tr>
<th>Table 4. Part-Whole Relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relation</td>
</tr>
<tr>
<td>have-part</td>
</tr>
<tr>
<td>part-of</td>
</tr>
</tbody>
</table>

4 Botany-Specific Axioms

We have been building a very large botany knowledge base from several large knowledge sources[2,4,5]. In our practice, we find that it is extremely important to ensure that the botany knowledge stored in the knowledge base is accurate and consistent. Obviously, manual analysis, as done in MindNet[20–22], is time-consuming, biased and error prone.

We have summarized a list of botany-specific axioms both for identifying inconsistency and inaccuracy in the acquired knowledge and for reasoning with the acquired knowledge. They form a second-order axiomatic system, and are an integral part of our whole ontology of botany. When a piece of botany knowledge is stored into the knowledge base, it is first checked by these axioms. If one of the axioms is violated, relevant information is reported to a knowledge engineer.

To represent botanical axioms, we have designed a formal language. The syntax of the language is given in the Appendix. An axiom has two parts: (Context) and (Body). (Context) is the precondition of the axiom, such as time, place, and discipline in which the axiom makes sense. There exist axioms without (Context) because these axioms are true universally in botany. But in general, (Context) is necessary for two reasons:

- Axioms are botany-specific, and they do not necessarily make sense in other domains.

For example, order-classification is a relation in botany-specific ontology. Axioms about order-classification may be true in a zoological ontology, but make no sense in a physical ontology.

- Some attributes and relations are closely related to time, place and other factors, therefore axioms about such attributes and relations need to take time, place, and other factors into consideration (see the following for examples).

The (Body) of an axiom is subdivided into two parts: (Variable-definition) and (Formula). (Variable-definition) defines types of variables, and the type of variables can be any defined category: integer, string, etc. (Formula) is first-order formula with slight variation. For example, we use \( R(X, Y) \) to mean that there is no relation \( R \) from \( X \) to \( Y \). (Predicate-item) is classified into two groups: (Normal-predication-item) and (Relation-predication-item). (Relation-predication-item) is the predication of relations that have been defined in botanical categories. The form of (Relation-predication-item) is \( R(C_1, C_2) \), meaning \( C_1 \) is in relation \( R \) with \( C_2 \), and \( R \) is defined in the category of \( C_1 \). Predications of (Normal-predication-item) are specially defined. Table 5 gives some of the predications. Parameters of all predications in botany axioms may be constants, defined variables, mathematical functions, predications, attributes of concepts, relations of concepts, etc. It should be mentioned that parameters, attributes of concepts and relations of concepts are represented in a dot notation: \( C \cdot A \) and \( C \cdot R \). \( C \cdot A \) means the attribute \( A \) of concept \( C \), and it may be used as a variable or value of the variable. \( C \cdot R \) means \( C \) is related in \( R \) with other concepts, and may also be used as a variable or value of the variable.

Because the backbone structure (the taxonomical structure) of botany is unique, we classify the botany-specific axioms into two classes: axioms about the backbone structure of botanical ontology, and axioms about non-structural botanical knowledge. All the axioms are used to check the acquired botany knowledge in order to identify bugs (i.e., inconsistency and incompleteness) in it. They are also used in botany knowledge inference, which is more complex than analysis. Inference is conducted with axioms about known facts in the botany knowledge base, and draws conclusions that are not explicitly specified.

4.1 Axioms

4.1.1 Relevant

The backbone structure of botany is formed by taxa (i.e., kind, family, genus, and species). These concepts are the root categories of these taxonomic classes, and it is easy to verify that they be represented in botanical categories. For example, if a class, then it be represented in botanical categories.

Definition 4.1.1.4-tuple: BS = (\( S \), \( V \), \( E \), \( \alpha \))

1) \( S = \{ \text{class, in-phylum} \}
2) \( V = \{ \text{class, phylum} \}
3) \( E = \{ \text{is-a} \}
4) \( \alpha \) is a mapping
\( \alpha(p) = \{ v \} \)

Definition

Notions and relations structured in the backbone category, such as: cat, us-category, nextt1, t1-category, phylum, category, phylogenetic category, next(t, tcategory), (genus category, or (family category), (class category), (phylum category, or (kingdom category).
Table 5. Examples of Predications

<table>
<thead>
<tr>
<th>Predicate</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>equal(X,Y)</td>
<td>X is equal to Y</td>
</tr>
<tr>
<td>not-equal(X,Y)</td>
<td>X is not equal to Y</td>
</tr>
<tr>
<td>greater-than(X,Y)</td>
<td>X is greater than Y</td>
</tr>
<tr>
<td>greater-or-equal(X,Y)</td>
<td>X is greater than or equal to Y</td>
</tr>
<tr>
<td>less-than(X,Y)</td>
<td>X is less than Y</td>
</tr>
<tr>
<td>less-or-equal(X,Y)</td>
<td>X is less than or equal to Y</td>
</tr>
<tr>
<td>contain(X,Y)</td>
<td>X and Y are sets, X contains Y</td>
</tr>
<tr>
<td>not-contain(X,Y)</td>
<td>X and Y are sets, X does not contain Y</td>
</tr>
<tr>
<td>element-of(X,Y)</td>
<td>Y is a set, X is an element of Y</td>
</tr>
<tr>
<td>not-element-of(X,Y)</td>
<td>Y is a set, X is not an element of Y</td>
</tr>
<tr>
<td>intersection(X,Y)</td>
<td>The intersection of set X and set Y</td>
</tr>
<tr>
<td>union(X,Y)</td>
<td>The union of set X and set Y</td>
</tr>
<tr>
<td>cardinality(X)</td>
<td>The cardinality of set X</td>
</tr>
<tr>
<td>abs(X)</td>
<td>The absolute value of number X</td>
</tr>
<tr>
<td>sqrt(X)</td>
<td>The square root of X</td>
</tr>
<tr>
<td>percent(X,Y)</td>
<td>The percentage of X in Y</td>
</tr>
</tbody>
</table>

4.1 Axioms of Backbone Structure (BS)

4.1.1 Relevant Definitions of BS

The backbone structure of botanical ontology is formed by concepts belonging to botany taxa (i.e., kingdom, phylum, class, order, family, genus, and species). Relations among these concepts are presented in Table 2. From definitions of these relations and botany knowledge, it is easy to verify that hyponymic relations can be represented by corresponding hyponymic relations. For example, if C1 is a phylum, and C2 is a class, then class-classification(C1, C2) implies inphyllum(C2, C1), and vice versa. In the following, therefore, we will only present and discuss axioms about the hyponymic relations.

Definition 2. We define botany structure as a 4-tuple: BS = (SR, V, E, α), where

1) SR = \{ in-genus, in-family, in-order, in-class, in-phyllum, in-kingdom \};
2) V = \{ C | C is a species, genus, family, order, class, phylum or Plantae \};
3) E = \{ (v1, v2) | v1 ∈ V, v2 ∈ V \};
4) α is a mapping: SR → E. For all p ∈ SR, α(p) = \{ (v1, v2) | p(v1, v2) is true in botany \}.

Definition 3. We define the following useful notions and functions:

structural-category =_def_ \{ species-category, genus-category, family-category, order-category, class-category, phylum-category, kingdom-category \};

next(t1, t2) =_def_ \{ (species-category, genus-category), (genus-category, family-category), (family-category, order-category), (order-category, class-category), (class-category, phylum-category), (phyllum-category, kingdom-category) \}.  

\[
d(C_1, C_2) = \begin{cases} 
0, & \text{if } C_1 = C_2; \\
1, & \text{if } C_1 \text{ and } C_2 \text{ belong to structural categories } t_1 \text{ and } t_2 \text{ respectively, next } (t_1, t_2), \text{ and there exists a } p \in SR \text{ such that } p(C_1, C_2) \text{ is true in botany}; \\
n, & \text{if there exists } C, d(C_1, C) = 1, \\
\cdots & , \\
\infty, & \text{otherwise}. 
\end{cases}
\]

We can draw several conclusions from the definitions above.

Corollary 1. The type of C is a structural-category iff C is in V.

Corollary 2. If there exists some C in V such that d(C1, C) ≠ ∞ and d(C2, C) ≠ ∞, then d(C1, C2) = d(C1, C) + d(C, C2).

4.1.2 Axioms about BS

In the following, we introduce the axioms about BS, and we use BSA to indicate the set of the axioms about BS.

Axiom 1 (Fact Axiom). ∀X: structural-category, ∀P: SR, in(P(X,Y), BS) ∧ equal(d(X,Y), 1) → in(P(X,Y), BSA).

Fact axioms are basic facts about relations between concepts in V. They cannot be deduced by any other axioms, and they are parts of the axiomatic system about BS.

We introduce two tree axioms below.

Axiom 2. ∀X: structural-category, ∀Y1, Y2, ..., Yn: structural-category, ∀p1, p2, ..., pn: SR, in(p1(Y1, Y2), BS) ∧ ... ∧ in(pn(Yn-1, Yn), BS) ∧ equal(d(Yn-1, Yn), 1) → equal(d(Yn, Yn+1), 1).

Axiom 3. ∀X: structural-category, ∀Y1, Y2, ..., Yn: structural-category, ∀p1, p2, ..., pn: SR, in(p1(Y1, Y2), BS) ∧ ... ∧ in(pn(Yn-1, Yn), BS) ∧ equal(d(Yn-1, Yn), 1) → equal(d(Yn, Yn+1), 1).

In botany, the basic relations between taxon concepts (relations defined in Axiom 1) and the taxon concepts form a directed tree. Concepts of
botany species are leaves, and Plantae is the only root. The tree goes upwards from leaves to the root, and its depth is 7. Axiom 2 and Axiom 3 are called tree axioms because they describe the tree structure of the taxon concepts system. Axiom 2 indicates that there exists a directional path from each concept to the root Plantae. Axiom 3 indicates that there exists only one directional path (length of each step is 1) from each concept to the root.

Before presenting the triangle axiom, we need to define the path between two structural categories.

**Definition 4 (Path).** \( \text{path}(X, Y) \overset{\text{def}}{=} (X, Z_1, Z_2, \ldots, Z_m, Y) \) for \( X, Z_1, Z_2, \ldots, Z_m, Y : \) structural-category, iff there exist \( p_0, p_1, p_2, \ldots, p_m : \) SR, \( p_0(X_1) \land \cdots \land p_{i-1}(Z_{i-1}, Z_i) \land \cdots \land p_m(Z_m, Y) \land \text{d}(X, Z_1) = 1 \land \cdots \land \text{d}(Z_m, Y) = 1 \).

**Axiom 4 (Triangle Axiom).** \( \forall X, Y, Z : \) structural-category, \( \forall p_1 : \) SR, \( \forall p_2 : \) SR, \( p_1(Y, Z) \land p_2(Z, Y) \rightarrow \text{p}_2(Y, Z) \).

Axiom 4 is depicted graphically in Fig. 5. In the figure, nodes indicate botanical concepts \( X, Y \) and \( Z \) of some structural categories, and edges represent relations among these concepts. As an example, let \( p_1 \) be in-class, \( p_2 \) be in-phyllum. Then Axiom 4 states that: if \( X \) belongs to class \( Y \), and \( Y \) belongs to phylum \( X \), then \( Z \) belongs to phylum \( Y \).

![Fig. 5](image)

From Axioms 1 to 4, a few useful conclusions can be drawn.

**Proposition 1.** \( \forall X, Y : \) structural-category, if not-equal(X, Y), then:
1) there exists \( p : \) SR and \( p(X, Y) \) implies that there exists one path, \( \text{path}(X, Y) \);
2) if \( \text{path}(X, Y) = (X, Z_1, Z_2, \ldots, Z_m, Y) \), then \( \text{d}(X, Y) = m + 1 \).

Proof. First, we prove the first part.

If \( \text{d}(X, Y) = 1 \), then \( \text{path}(X, Y) = (X, Y) \).

If \( \text{d}(X, Y) \neq 1 \), then according to Axiom 4, there must exist \( S_1 \) (whose category is structural-category) and \( p_1 \) such that \( p_1(X, S_1) \land \text{path}(S_1, Y) \). If \( \text{d}(X, S_1) \neq 1 \), repeat the above argument. At last, there must exist \( Z_1, Z_2, \ldots, Z_m : \) structural-category, and \( q_0, q_1, \ldots, q_m : \) SR, such that \( q_0(X, Z_1) \land \cdots \land q_{i-1}(Z_{i-1}, Z_i) \land \cdots \land q_m(Z_m, Y) \land \text{d}(X, Z_1) = 1 \land \cdots \land \text{d}(Z_m, Y) = 1 \).

We prove the uniqueness of the paths between \( X \) and \( Y \) by contradiction.

Suppose there exist two paths from \( X \) to \( Y \): \( \text{path}_1(X, Y) = (X, Z_1, Z_2, \ldots, Z_m, Y) \) and \( \text{path}_2(X, Y) = (X, W_1, W_2, \ldots, W_m, Y) \), and \( \text{path}_1(X, Y) \neq \text{path}_2(X, Y) \). According to Axiom 2, there exists a path: \( \text{path}(Y, \text{Plantae}) \). Suppose \( \text{path}(Y, \text{Plantae}) = (Y, M_1, M_2, \ldots, M_n, \text{Plantae}) \), then \( (X, Z_1, Z_2, \ldots, Z_m, Y, M_1, M_2, \ldots, M_n, \text{Plantae}) \) is a path from \( X \) to \( \text{Plantae} \), and we use \( \text{path}_1(X, \text{Plantae}) \) to denote it. Similarly, there exists a path: \( \text{path}_2(X, \text{Plantae}) = (X, W_1, W_2, \ldots, W_m, Y, M_1, M_2, \ldots, M_n, \text{Plantae}) \). Since \( \text{path}_1(X, Y) \neq \text{path}_2(X, Y) \), then \( \text{path}_1(X, Y) \neq \text{path}_2(X, \text{Plantae}) \). This is inconsistent with Axiom 3. So there exists only one path from \( X \) to \( Y \).

The second part of this proposition is direct from Definition 3.

**Proposition 2.** \( \forall X, Y, Z : \) structural-category, \( \forall p_1 : \) SR, \( \forall p_2 : \) SR, \( p_1(X, Y) \land p_2(Y, Z) \rightarrow \text{d}(X, Z) = \text{d}(X, Y) + \text{d}(Y, Z) \).

Proof. Since \( p_1(X, Y) \), suppose \( \text{path}(X, Y) = (X, Z_1, Z_2, \ldots, Z_m, Y) \), then \( \text{d}(X, Y) = m + 1 \).

Since \( p_2(Y, Z) \), suppose \( \text{path}(Y, Z) = (Y, W_1, W_2, \ldots, W_n, Z) \), then \( \text{d}(Y, Z) = n + 1 \).

Because \( (X, Z_1, Z_2, \ldots, Z_m, Y, W_1, W_2, \ldots, W_n, Z) \) is a path from \( X \) to \( Y \), \( \text{path}(X, Z) = (X, Z_1, Z_2, \ldots, Z_m, Y, W_1, W_2, \ldots, W_n, Z) \), and \( \text{d}(X, Z) = m + n + 2 \).

So \( \text{d}(X, Z) = \text{d}(X, Y) + \text{d}(Y, Z) \).

**Proposition 3.** \( \forall X, Y, Z : \) structural-category, \( \forall p_1, p_2 : \) SR, \( p_1(X, Z) \land p_2(Y, Z) \land \text{d}(Y, \text{Plantae}) < \text{d}(Y, \text{Plantae}) \rightarrow p_1(Y, Z) \).

The proof of Proposition 3 is straightforward. Nevertheless, we would like to illustrate its meaning graphically in Fig. 5: if \( Z \) belongs to \( X \), \( Z \) belongs to \( Y \), and it is further from \( Y \) to Plantae than from \( X \) to Plantae, then \( Y \) belongs to \( X \). For example, if \( Z \) is an order, \( Z \) belongs to phylum \( X \), and \( Z \) belongs to class \( Y \), then \( Y \) belongs to phylum \( Z \). This is logical in botany, but it is not necessarily logical in other disciplines.

With the preparation above, we now formally define the BSA as follows.
Definition 5. \( BSA = \{Axiom \ 1, \ Axiom \ 2, \ Axiom \ 3, \ Axiom \ 4\} \).

4.2 Axioms about Non-Structural Knowledge (NSK)

4.2.1 Definitions of NSK

Axioms about NSK are used for non-structural knowledge analysis and inference. NSK includes all the botany knowledge except for knowledge about the botanical taxonomy structure that we have discussed above, e.g., knowledge about evolution, distribution, and composition of plants. Knowledge about common classification of botanical concepts is also included in this group. Definition 6 defines the non-structural knowledge about botany. \( BC \) is the set of all botanical concepts, and \( CR \) is relations among these concepts defined in botanical categories.

Definition 6. \( NSK = \{BC, CR, E, \alpha\} \), where

1. \( BC = \{C | C \text{ is a botanical concept}\} \);
2. \( CR = \{R | R \text{ is a common relation defined in botanical categories}\} \);
3. \( E = \{(v_1, v_2) | v_1 \in BC, v_2 \in BC\} \);
4. \( \alpha \) is a mapping: \( CR \rightarrow E \). For all \( r \in CR \), \( \alpha(r) = \{(v_1, v_2) | r(v_1, v_2) \text{ is in botany}\} \).

4.2.2 Axioms of NSK

For the purpose of discussion, we classify all the axioms about NSK into several groups according to the content they talk about. Because axioms about NSK are much more than axioms about BS, we just demonstrate several axioms for each group. We use botany-category to indicate any category we have defined such as phylum-category, stem-category and cell-category.

Fig.6 gives some examples of the axioms about non-structural knowledge. Axioms 5–7 are about common classification of botanical concepts, Axioms 8–10 are about evolution of plants, Axioms 11–13 are about number of plants, Axioms 14–16 are about distribution of plants, Axioms 17–19 are about living style of plants, Axioms 20–22 are about components of plants, Axioms 23–25 are about parts of plants, and Axioms 26–28 are about use of plants.

\begin{align*}
\text{Axiom 5:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \text{generalization}(Y, X) \leftrightarrow \text{specialization}(X, Y). \\
\text{Axiom 6:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \text{generalization}(Y, X) \land \text{generalization}(X, Y) \leftrightarrow \text{equal}(X, Y). \\
\text{Axiom 7:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \text{generalization}(Y, X) \land \text{generalization}(Z, X) \leftrightarrow \text{generalization}(Z, Y). \\
\text{Axiom 8:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \forall (X-\text{ancestor}, Y) \rightarrow \text{advanced-than}(X, Y). \\
\text{Axiom 9:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \forall (X-\text{ancestor}, Y) \land \text{equal}(Y-\text{ancestor}, Z) \leftrightarrow \text{equal}(X, Z). \\
\text{Axiom 10:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \text{less-than}(X-\text{time-of-origin}, X-\text{time-of-extinction}). \\
\text{Axiom 11:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \text{less-or-equal}(X-\text{number-of-known-genus}, X-\text{number-of-known-species}). \\
\text{Axiom 12:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \text{less-or-equal}(X-\text{number-of-familiar-class}, X-\text{number-of-existing-class}). \\
\text{Axiom 13:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \forall (X, Y): \text{class-category, } \text{class-classification}(X, Y, Y_1, Y_2, \ldots, Y_n) \leftrightarrow \text{equal}(X-\text{number-of-known-species}, Y-\text{number-of-known-species}). \\
\text{Axiom 14:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \text{contain}(X-\text{producing-area}, X-\text{primary-producing-area}). \\
\text{Axiom 15:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \text{contain}(X-\text{producing-area}, X-\text{primary-producing-area}). \\
\text{Axiom 16:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \forall (X, Y): \text{specialization}(Y, X_1, X_2, \ldots, X_n) \leftrightarrow \text{equal}(X-\text{producing-area}, Y_1, Y_2, \ldots, Y_n-\text{producing-area}). \\
\text{Axiom 17:} & \quad \forall X: \text{flower-category, } \forall Y: \text{flower-category, } \text{contain}(X, Y), X \text{ begin-of-florescence}, Y \text{ end-of-florescence}. \\
\text{Axiom 18:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \text{grow-into}(X, Y) \rightarrow \text{grow-out-of}(Y, X). \\
\text{Axiom 19:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \text{accrete-with}(Y, X) \rightarrow \text{accrete-with}(X, Y). \\
\text{Axiom 20:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \exists (X-\text{special-material}, X-\text{material}). \\
\text{Axiom 21:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \text{contain}(X-\text{photosynthesis-pigment}, X-\text{pigment}). \\
\text{Axiom 22:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \forall Z: \text{exist}(X-\text{cardinal-number-of-chromosomes}). \\
\text{Axiom 23:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \text{part-of}(X, Y) \rightarrow \text{have-part}(X, Y). \\
\text{Axiom 24:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \forall (X, Y): \text{contain}(X-\text{primary-use}, Y-\text{use}). \\
\text{Axiom 25:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \forall Z: \text{exist}(X-\text{primary-use}, X-\text{use}). \\
\text{Axiom 26:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \forall Z: \text{exist}(X-\text{life-cycle}, Y-\text{life-cycle}). \\
\text{Axiom 27:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \forall Z: \text{exist}(X-\text{primary-use}, Y-\text{use}). \\
\text{Axiom 28:} & \quad \forall X: \text{botany-category, } \forall Y: \text{botany-category, } \forall Z: \text{exist}(X-\text{special-use}, Y-\text{use}).
\end{align*}
Axiom 11 gives a 'botanical constraint' on the attributes number-of-known-genus and number-of-known-species. It says that given any botanical phylum, its number of genera must be less than that of species. In other words, a violation of this axiom is an inconsistency bug in our botanical knowledge base. Axiom 22 deals with the number of chromosomes. It states that the number of chromosomes of plants equals $N$ times the cardinal number of chromosomes of plants.

All these axioms demonstrate that botany-specific ontologies are very different from others. It also means that general-purpose ontologies are sometimes hard to be tailored to a specific domain.

Formally, we use NSKA to indicate the axiom system of NSK. It is defined in Definition 7. Obviously, Axioms 5-28 belong to NSKA.

Definition 7. NSKA = $\{A | A$ is an axiom about common botany knowledge $\}$.

5 Axiomatic Analysis

In this section, we analyze the axioms from the perspectives of completeness, consistency and redundancy. Accuracy of the axioms is also very important, but it cannot be analyzed just with the axioms themselves. In fact, it should be checked with botany knowledge in the knowledge base.

Definition 8. We define the botany knowledge system $BK = (BC, BR)$, where

1) $BC = \{C | C$ is a botanical concept $\}$;
2) $BR = \{R | R$ is a relation between botanical concepts $\}$.

To analyze the completeness of axioms about the backbone structure, we give out the definition of completeness first. We use BSA-Completeness to indicate the completeness of axioms about the backbone structure.

Definition 9 (BSA-Completeness). For any $p$ in $SR$, for any $X, Y$ in $V$, if $BK \vdash p(X, Y)$ implies $BSA \vdash p(X, Y)$, then $BSA$ is complete.

Theorem 1. BSA is complete.

Proof. For any $p$ in $SR$, for any $X, Y$ in $V$, if $BK \vdash p(X, Y)$, then $d(X, Y) \neq \infty$. Let $d(X, Y) = N$, we prove the conclusion by induction.

When $N = 1$: $BSA \vdash p(X, Y)$ is true.

Suppose that if $BK \vdash p(X, Y)$, then $BSA \vdash p(X, Y)$ when $N = n - 1$. We consider the case when $N = n$. If $BK \vdash p(X, Y)$, then according to the botanical taxonomy system (the structure of the botanical taxonomy system is a rigid tree; every species of botany directly belongs to certain genus, then family, order, class, phylum, and Plantae at last; so do genus, family, order, class, and phylum), there exist $Z$ in $V$ and $p_1$ in $SR$ such that $BK \vdash p_1(X, Z)$, $BK \vdash p(Z, Y)$, $d(X, Z) = 1$ and $d(Z, Y) = n - 1$; then $BSA \vdash p_1(X, Z)$ and $BSA \vdash p(Z, Y)$, then $BSA \vdash p(X, Y)$.

In conclusion, if $BK \vdash p(X, Y)$, then $BSA \vdash p(X, Y)$. That is to say, BSA is complete. □

Definition 10 (BSA-Consistency). For any $p$ in $SR$, for any $X, Y$ in $V$, if $BK \vdash !p(X, Y)$ implies $BSA \vdash !p(X, Y)$, then $BSA$ is consistent.

Theorem 2. BSA is consistent.

Proof. Suppose $BSA \vdash p(X, Y)$.

Since $BK \vdash !p(X, Y)$, there are two kinds of cases that can be found:

There exists only one $p_1(\neq p)$ in $SR$ such that $BK \vdash p_1(X, Y)$. Or there exists no $p_1$ in $SR$ such that $BK \vdash p_1(X, Y)$. In other words, $d(X, Y) = \infty$.

Under the first condition, because $BK \vdash p_1(X, Y)$, then $BSA \vdash p_1(X, Y)$, which is proved in Theorem 1. So there exists a path Path1 from $X$ to $Y$ whose last relation is $p_1$. On the other hand, because $BSA \vdash p(X, Y)$, there exist a path Path2 from $X$ to $Y$ whose last relation is $p$. Since $p \neq p_1$, Path1 $\neq$ Path2, which contradicts Proposition 1.

Under the second condition, because $BSA \vdash p(X, Y)$, then there exists a path Path1 from $X$ to $Y$ whose last relation is $p$, then $d(X, Y) < \infty$, which contradicts $d(X, Y) = \infty$.

In general, the supposition that $BSA \vdash p(X, Y)$ is wrong, and if $BK \vdash !p(X, Y)$, then $BSA \vdash !p(X, Y)$.

So BSA is consistent. □

Definition 11 (BSA-Redundancy). If there exists an $A$ in $KS$, for any $p$ in $SR$, and for any $X, Y$ in $V$, if $BK \vdash p(X, Y)$ and $BSA \vdash \{A\} \vdash p(X, Y)$, then $BSA$ is redundant.

Theorem 3. $BSA$ is not redundant.

Proof. Axiom 1 indicates that the basic facts of biology taxonomy are included in the axiomatic system, they cannot be deduced by another axiom, but can be derived from the knowledge base.

Axioms 2 and 3 indicate the tree structure of the backbone, and Axiom 4 indicates relations among the relations in SR. Obviously, they indicate different characters of the backbone, and cannot be deduced from each other, even companied with Axiom 1.

Therefore, BSA is not redundant. □

To end this section, we discuss about the completeness, consistency and redundancy of NSKA.

6 Conclusion

A botanical ontology includes two categories accounting for ontologies.

Or we can say that the specification of a botanical ontology characterizes the true relationships among botanical relationships. A knowledge base is considered as relations among the constructs of botany and construct and complete the theory of taxonomic.

In this paper, our idea is of constructing a knowledge base with ontologies. We constructed a model of botany, a structure of botany would be used in the concepts more in the future of users. Concepts of botany are integrated to one.

Our botanical knowledge base is the view of botany knowledge, complete, consistent and redundant. Any violation of completeness, consistence and redundancy analytical, then there is also ontology inconsistency. Bases are those pieces of knowledge of the other knowledge base.

Acknowledgments

The authors, Xiang-Yun Zhu of Inst. of Botany, CAS, thank for their contributions. Authors are also appreciative to the students for their support.
Fang Gu et al.: Domain-Specific Ontology of Botany

The general analysis of completeness of NSKA is difficult. In fact, there is no algorithm for this task. Many knowledge systems, such as BKB[2], CYC[16], WordNet[18] and MindNet[20], do not focus on the completeness issue. The consistency of NAKA that we have discovered can be guaranteed. But since more axioms to be discovered or designed, a general proof of the consistency is intractable, and we need some sort of heuristic algorithm for consistency checking. This remains in our research agenda. Finally, the redundancy of NSKA can be guaranteed just as BKA. But an efficient algorithm is needed since the NSKA may be eventually huge.

6 Conclusions

A botanical ontology is a specific system of categories accounting for a certain vision of the world. Or we can say that botanical ontology is an explicit specification of a conceptualization of botany. The ontology characterizes the essence of things and the true relationship among things, especially categorical relationships. In the context of AI, the knowledge base is constructed with a set of concepts and relations among these concepts. It is natural to construct and organize the knowledge base with theory of taxonomy.

In this paper, we discussed our current work of constructing and organizing a botany knowledge base with ontological engineering and theory. We constructed a multi-perspective ontological structure of botany in order to organize the botanical concepts more naturally and meet different needs of users. Concepts of plants and categorical terms of botany are instances of corresponding categories.

Our botanical ontology is not merely a semantic view of botany itself. It plays a key role in checking the consistency of the acquired botany knowledge. Any violation of ontological axioms is a piece of evidence of inconsistency in the knowledge base. Redundancy analysis for the botany knowledge base is also ontology-driven. There is no need to keep those pieces of knowledge that can be derived from other knowledge and ontological axioms.

Acknowledgement We appreciate Dr. Xiang-Yun Zhu and Dr. Yao-Dong Qi of the Institute of Botany, the Chinese Academy of Sciences, for their contribution to this project. We would also appreciate our colleagues and other graduate students for their contributions.

References

Appendix. Representation of Axioms

Axiom ::= [(Context) ; (Body)]
(Context) ::= (Context-item ; (Context
-item-value) ; (Context-item-value))
(Context-item) ::= (discipline, time, position, range, basis, ...)
(Context-item-value) ::= (x | x is a discipline) ∪ (x | x is a time) ∪ (x | x is a position) ∪ ...
(Body) ::= (Variable-definition ; (Formula))
(Variable-definition) ::= (∀ ∃) (Variable-list ; (Variable-type))
(Variable-list) ::= (Variable , (Variable))
(Variable) ::= (a, b, ..., z) ∪ {A, B, ..., Z} ∪ {X, X ∈ (a, b, ..., z) ∪ {A, B, ..., Z}, i ∈ {0, 1, 2, ...}}
(Variable-type) ::= (general-category, kingdom-category, phylum-category, class-category, order-category, family-category, cell-category, ...) ∪ {integer, real-number, boolean, character, string, ...}
(Formula) ::= (And-item ; (And-item ; (And-item)) ; (And-item ; (And-item)) ; (And-item ; (And-item)))
(And-item) ::= (Predicate ; (Predicate ; (And-item)))
(Predicate) ::= (Ordinary-predicate ; (Relational-predicate))

Note that (Relational-predicate) is defined in botanical categories.

Rule Extraction from Neural Networks?

Zhi-Hua Zhou
National Laboratory for Novel Software Technology
E-mail: zhouzh@nlsnt.org.cn
Received July 21, 2001.

Abstract In this paper, the behavior of an ANN is identified the fidelity and accuracy of rule extraction for neural networks is evaluated from the rule quality.

Keywords

1 Introduction

An inherent drawback of the learned knowledge from ANN is the lack of connections, which makes it difficult to derive the explicit or implicit knowledge and to transfer the ANN knowledge in the real world. Although the ANN reveals that clinical knowledge may be represented by the ANN, it is not clear whether the ANN can be used for the purpose of knowledge representation and discovery. Therefore, rule extraction from neural networks is an attractive and promising approach to solve this problem.

This paper addresses the problem of rule extraction from neural networks. The main focus is to evaluate the quality and accuracy of the extracted rules. The evaluation criteria are based on the fidelity and accuracy of the extracted rules. The fidelity of the extracted rules is evaluated by comparing the extracted rules with the ground truth. The accuracy of the extracted rules is evaluated by comparing the extracted rules with the rules generated by expert systems.

The work was supported by the National Natural Science Foundation of China.