Building an Ontology of Flora of Thailand for Developing Semantic Electronic Dictionary

Jirapong Panawong[*]
Faculty of Management Science, Nakhon Ratchasima Rajabhat University, Thailand

Nattapong Kaewboonma
Faculty of Management Technology, Rajamangala University of Technology Srivijaya, Thailand

Wirapong Chansanam
Faculty of Business Administration, ChaiyaphumRajabhat University, Thailand

Thepchai Supnithi
Language and Semantic Technology Laboratory, National Electronics and Computer Technology Center, Thailand

Marut Buranarach
Language and Semantic Technology Laboratory, National Electronics and Computer Technology Center, Thailand

Abstract

Thailand is one of the tropical countries teeming with diverse flora and vegetation. The diverse vascular plants of Thailand with estimated number of no less than 10,000 species, have been recorded and published in continuation in the Flora of Thailand since 1970. Domain ontologies express conceptualization that are specific for particular domain and greatly useful in knowledge acquisition, sharing and analysis. In this paper, we propose a Thailand's Flora Ontology (TFO) for developing semantic dictionary on the web to discover the flora knowledge for plant biologists across all disciplines of botany. A mixed methods was applied in organizing of the specification of conceptualizations on flora of Thailand using the domain analytic approach in order for developing an ontology. The TFO has been constructed by using HOZO ontology editor. The research methods included 1) Domain analysis for knowledge organization. 2) Ontology development. The results of classification of Thailand's Flora based on concise encyclopedia of plants in Thailand, floral characteristics and area of distribution can be divided into 8 concepts including Plant_Information, Plant_Family, Plant_Genus, Plant_Habitat, Botany_Habit, Uses, Medicinal_Properties, and Floristic_Regions. In the next step, semantic electronic dictionary will be developed by using the TFO from this study.

Keywords: Flora of Thailand; Thailand's flora ontology; Plant ontology; Sematic dictionary.

1. Introduction

Thailand is a country of forests, shrub-studded grasslands, and swampy wetlands dotted with lotuses and water lilies. Since the mid-20th century, the total land area covered by forests has declined from more than half to less than one-third. Forest clearing for agriculture (including for tree plantations), excessive logging, and poor management are the main causes of this decline. Forests consist largely of such hardwoods as teak and timber and resin-producing trees of the Dipterocarpaceae family. As elsewhere in Southeast Asia, bamboo, palms, rattan, and many kinds of ferns are common. Where forests have been logged and not replanted, a secondary growth of grasses and shrubs has sprung up that often limits land use for farming. Lotuses and water lilies dot most ponds and swamps throughout the country (Thailand - Plant and Animal life, 2017).

The study of the Flora of Thailand, with its estimated 10,000 vascular plants species, has been gathering momentum in the last few years and has now reached a well-advanced stage. The Flora of Thailand Project, whose aim is to produce a complete floristic treatment of the entire vascular flora, was initiated in 1963 under Thai-Danish (Thailand - Plant and animal life | history - geography", 2017). However, plants information is mostly described in English with botany technical terms, especially those which are studied taxonomy by botanists. Besides, plants sometimes have more than one botanical name or different names, as well as lack information of botanical characters and area of distribution. These problems make it difficult for ordinary people or even other plant researchers who are not botanists to understand or receive accurate information about plants.

During the previous years, there has been a growing concern on ontology due to its ability to explicitly describe data semantics in a common way, with independent data source characteristics and to provide a schema that allows data interchanging among heterogeneous information systems and users (Saad *et al.*, 2011). An ontology is a classification methodology for formalizing a subject's knowledge in a structured way. In the world of structured information, ontologies, comprising structured controlled vocabularies, play a very important role in facilitating information retrieval (The Plant Ontology Consortium, 2002). Information ontologies which specify the record

structure of databases. Domain ontologies express conceptualization that are specific for particular domain and greatly useful in knowledge acquisition, sharing and analysis (Gu *et al.*, 2004; van Heijst *et al.*, 1997). In this paper, we propose a Thailand's Flora Ontology (TFO) for developing semantic dictionary on the web to discover the flora knowledge for plant biologists across all disciplines of botany.

2. The Flora of Thailand

At present, the natural forest cover is an estimated 25 % of the total land area. Vast forest areas have been converted into secondary vegetation mainly by the urban developments and the expansion of agricultural lands. Nevertheless, the vegetation of Thailand is varied and can be classified into evergreen and deciduous forest types which are basically based on varying moisture gradients, temperatures and altitudes. The names of the dominant tree species are often used for associations and sub-types of vegetation, more technically, the characteristics of floristic composition is based, such as tropical evergreen rain forest, seasonal evergreen forest (or dry evergreen forest), montane forest, mangrove forest, peat swamp forest, strand vegetation, mixed deciduous forest and deciduous dipterocarp forest.

Botanically, Thailand is included in the Indochinese subdivision of the continental Southeast Asia, and phytogeographically, is situated between two floristic regions, viz. Malesian and Indochinese including Myanmar and South China. Thailand is considered as a collective centre of botanic diversity designated by three floristic regions: Indo-Burmese, Indo-Chinese, and Malesian. As a result, Thailand shares its flora with the neighboring countries. The number of endemic species is, therefore, not high. However, the richness of flora of Thailand comprising of estimated 10,000 vascular plant species, represented by 275 families of spermatophytes and 36 families of pteridophytes. In deciduous forests, plant diversity is rather poor. The main canopy trees of both mixed deciduous and deciduous dipterocarp forests are dominant by the dipterocarp and leguminous tree species.

During the past decades, the much increased population in Thailand combining with economic and infrastructure development has been responsible for forest retreats in every region of the country. Since the establishment of the Royal Forest Department in 1896 and the inauguration of the Department of National Parks, Wildlife and Plant Conservation in 2002, both departments are authorized as the main agencies for the forest and wildlife conservation and sustainable management of the forest resources. At present, there are 147 national parks, 108 forest parks, 57 wildlife sanctuaries, 49 non-hunting areas, 16 botanical gardens and 55 arboreta throughout the country covering over 60 % of the remaining forest areas and containing most of natural resources of ecological importance in the country (Thailand - Plant and animal life | history - geography", 2017).

3. Related Research

The Plant Ontology Consortium (POC) released two major ontologies: (i) Plant Structure Ontology (PSO), which, as one of the top-level parent terms in PO, describes morphological and anatomical structures and includes organs and organ systems, tissues and cell types (ii) whole plant Growth Stages Ontology (GSO), which describes organism growth stages such as 'germination', 'rosette growth', 'flowering' and 'senescence' that cover the vegetative and reproductive lifecycle of an entire plant (Avraham, 2008). Shi and Huang (2009) developed the "Flora-oriented Plant Ontology" for acquiring botanical knowledge form flora. The ontology can reflect the flora domain knowledge hierarchical structure clearly, and can be used for application systems, such as information extraction systems and information retrieval systems. The ontology constructed using a method which is combined with seven-step methods and skeleton methods from the aspect of floristic morphologic descriptions successfully, instead of using the traditional methods of biological classification. The basic relations are used frequently in the Flora-oriented Plant Ontology, such as kind-of, part-of, attribute-of, instance-of, and so on.

More examples of research on flora ontology include the study of Hoehndorf (2016) the Flora Phenotype Ontology (FLOPO), an ontology for describing traits of plant species found in Floras. This study used the Plant Ontology (PO) and the Phenotype And Trait Ontology (PATO) to extract entity-quality relationships from digitized taxon descriptions in Floras, and used a formal ontological approach based on phenotype description patterns and automated reasoning to generate the FLOPO. The resulting ontology consists of 25,407 classes and is based on the PO and PATO. The classified ontology closely follows the structure of Plant Ontology in that the primary axis of classification is the observed plant anatomical structure, and more specific traits are then classified based on parthood and subclass relations between anatomical structures as well as subclass relations between phenotypic qualities. A study by Gu *et al.* (2004) focused on created 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 the ontology domain-specific.

4. Methods
4.1. Domain Analysis for Knowledge Organization

A qualitative research method was used to develop the knowledge structure of flora of Thailand. The study was conducted using the following steps: 1) Survey and selection of existing resources on flora of Thailand, plant ontology, plant taxonomy and biological classification. The selected resources included: (1) Flora-oriented Plant Ontology (Shi and Huang, 2009); (2) Domain-Specific Ontology of Botany (Gu *et al.*, 2004); (3) (The Plant Ontology Consortium, 2002); (4) Flora Phenotype Ontology (Hoehndorf, 2016) and the concise Encyclopedia of Plants in Thailand (Thailand - Plant and animal life | history - geography", 2017). 2. Content analysis of the flora from the mentioned selected resources according Existing plant classifications were used as guidelines and as

resources for comparison. 3. Organizing of the flora of Thailand using the domain analytic approach. The concepts of flora of Thailand were term-assigned and categorized. 4. Clarification and modification of the knowledge organization on the flora of Thailand by consulting with domain experts.

4.2. Ontology Development Process

It is to be give emphasis to that developing a new ontology is often dreary and time consuming; it normally requires ontology engineers and developers to have more sufficient knowledge in ontology specifications and familiar with ontology development environment (Chansanam and Tuamsuk, 2016). There is no single correct ontology for any domain (Noy and McGuinness, 2000). In construction our ontologies for the Thailand's Flora Ontology (TFO) on the case study, we consulted with the guidelines suggested in an ontology development 101 (Noy and McGuinness, 2000). We outline our five step iterative ontology development process in Table 1. We provide guideline of the ontology development process on the selected case study in the subsequent sub-sections.

Table-1. Ontology Development Process

Ontology Development Process	Description
Step 1: Synthesize Information Collected	Information gathered and described in the plant and flora documents created through collaboration with biology analysts, ontology engineers, and requirement engineers during the requirements elicitation process are analyzed and synthesized. Terms take out from these documents are applicants for the definition of classes and properties in the ontologies.
Step 2: Look up Existing Ontologies	There are widespread libraries of reusable ontologies available on the Web. As building a new ontology from scratch is a time consuming process, if an existing solution ontology is available and is relevant to the application domain in hand, then it is proposed to refer with the existing ontology to define if we can reuse, refine, or spread out existing classes and properties.
Step 3: State Classes and Properties	A top-down approach is adopted to define the class hierarchy (super-sub-class), i.e. from most general concepts to specialized concepts. A set of potential classes, class hierarchy, and class properties (i.e. attributes, cardinalities, and relationships with other classes) are identified, defined, and specified in the HOZO ontology editor tool. This step occupies the most time.
Step 4: Create Instances	Instances of the classes are created in the Ontology Application Management Framework (OAM) [12]. Constructing class instances can assistance to accurate mistakes and fine-tune the classes and properties in the ontology.
Step 5: Association Ontologies	If one or more relations exist between classes of two ontologies, the ontologies are combined by including the related ontology into the current ontology via HOZO ontology editor tools. Combining related ontologies support ontology engineers better understand the relationships of classes between ontologies, classify conflicts, and make essential changes. Ontology engineers may necessity to revisit one or more previous steps to enhance the ontologies.

Step 1: Synthesize Information Collected

This step is closely correlated to the Concept Development Process (CDP) model proposed in previous work (Buranarach *et al.*, 2012); which we endorse readers to review to get to know more detail about the CDP model. For our case study, the primary target users are scientist that uses the system to browse for flora name, check relation, review information and so forth. The secondary users are the application developers and site administrator. In this

paper, we focus on the primary users. The attributes are potential candidates for the definition of classes in the ontologies of the TFO model.

Step 2: Consult Existing Ontologies

In constructing the ontologies for our TFO model on the case study, there appears one existing ontology that is similar, but not readily available for reusable purpose (Shi and Huang, 2009).

Step 3: State Classes and Properties.

We have consulted some existing ontologies on the approaches in specifying some of the class properties, for examples, Plant_Information class, Plant_Family class, Plant_Genus class, Plant_Habitat class, Botany_Habit class, Uses class, Medicinal_Properties class, and Floristic_Regions class.

Step 4: Create Instances

After the classes are defined and specified in Hozo ontology editor tools, instances of the classes are created in OAM tool. A representative name for an instance of a class is chosen and displayed in the "Instance Browser". The instance name may be selected from any one or combination of the values of the properties (or slots in Hozo). Values for the properties are filled in in the "Instance Editor" of the Hozo tool. One or more instances may be created for a class. In our case study, for example, we created an instance for the Plant_Information class and selected the value of the attribute Common_name as the representative instance name displayed in the "Instance Browser". An alternative for the instance name could be the Plant_ID attribute. Due to space limitation, we show in Fig. 1 the instance created for the Plant_Information concept of the TFO in the Hozo. The instances created in Hozo for the TFO Ontology can be found in details.

Step 5: Association Ontologies

The final step in our ontology development process is to combine related ontologies to support to better understand the relationships of classes between ontologies and make necessary corrections to refine the ontologies. If there exist one or more relationships between classes of two ontologies, then the ontologies are combined by including the related ontology into the current ontology. The included classes and properties are displayed in Hozo as browsing pane to distinguish from the classes in the current ontology. An included ontology may also be added with the current ontology to form a single ontology via the "add Ontology in the Project" and all classes and properties of the merged ontologies are then displayed as browsing pane.

Figure-1. Thailand's Flora Ontology Class
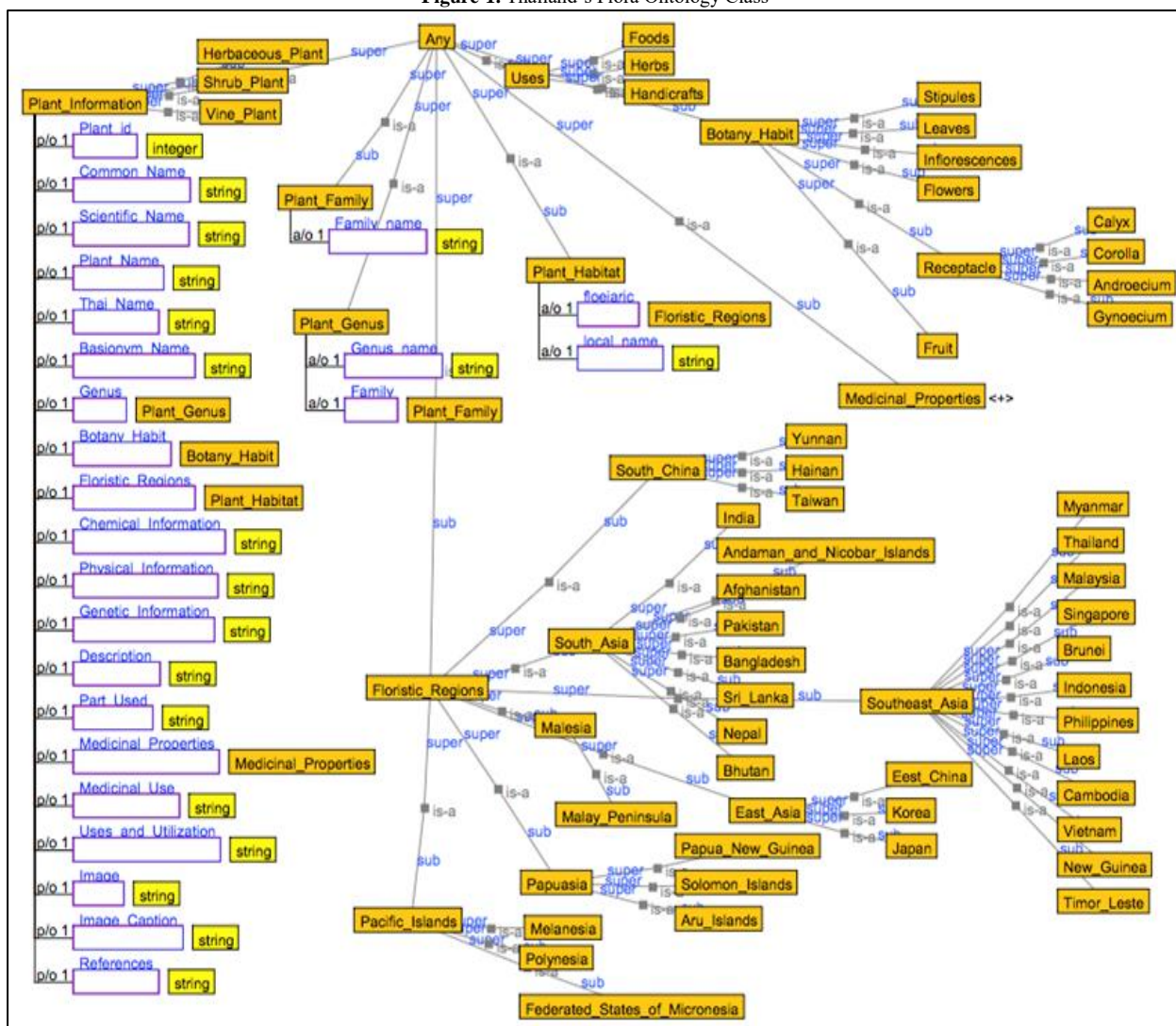
5. Results

The Knowledge structure. The knowledge structure has constructed by mixed method are using 5 steps: 1) Survey and selection of existing resources, 2) content analysis, 3) organizing using the domain analytic approach, and 4) clarification and modification of the knowledge organization. The knowledge structure comprised of 8 main classes including 1) Plant_Information, 2) Plant_Family, 3) Plant_Genus, 4) Plant_Habitat, 5) Botany_Habit, 6) Uses, 7) Medicinal_Properties, and 8) Floristic_Regions. Each main class is subdivided into sub-classes and the relationships of the topics were identified.

The ontology of Thailand's Flora. The knowledge structure from created was analyzed and used for ontology development by using Hozo ontology editor. Class is defined base on relation, is-a relation, part-of relation, and attribute-of relation. The ontology comprised of 8 concepts, and 67 of classes and sub-classes is shown in Figure 1.

Thailand's Flora Ontology. In this research focused on main class is Plant_Information as shown in Figure 2., class Plant_Information can be defined some properties (Table II) Plant_Id, Common_Name, Scientific_Name, Plant_Name, Thai_Name, Basionym_Name, Genus, Botany_Habit, Floristic_Regions, Chemical_Information, Physical_Information, Genetic_Information, Description, Part_Used, Medicinal_Properties, Medicinal_Use, Uses_and_Utilization, Image, Image_Caption and References. The class Plant_Information can be defined 3 sub-classes 1) class Herbaceous_Plant, 2) class Shrub_Plant, and 3) class Vine_plant. Class Plant_Information can be relation to other classes.

We develop a web base application to test the proposed the ontology of Thailand's Flora. We use the semantic search systems as backend and the RAP - RDF API for PHP web technology as fronted in additional using a multi representative framework (Westphal and Bizer) to interconnect between the representatives. We combined with a PHP library for spell checking and another library for one and the same (Princeton University). In our semantic search testing, we have collected 240 queries from different sources such as domain specialists, ontology professional through surveys and various information floras' websites. We categorize the 240 queries based on the terms related to the Thailand's flora domain. Then we mark up them manually to find how many terms associated to them. After that, we test the queries in the semantic search system to measure the perception. Table III shows the enlargement of the queries on the groups and the performance of our system query understanding.
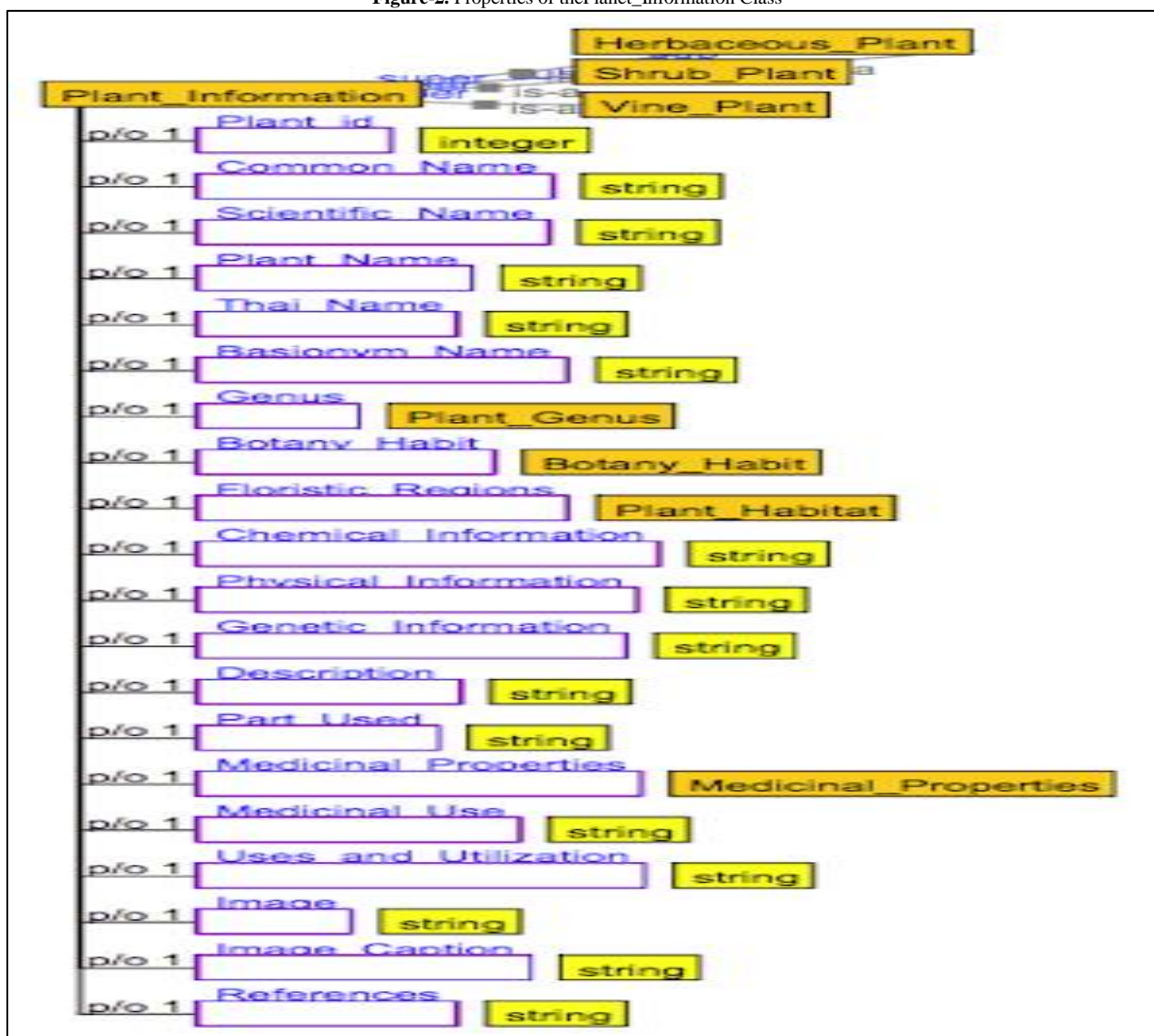
Figure-2. Properties of thePlanet_Information Class

Table-2. Plant_Information

Properties	Description
Plant_ID	Plant code
Common_name	Common name of the plant
Scientific_Name	Scientific name of the plant
Thai_Name	Name of plant in Thai Language
Basionym_Name	Original name of the plant
Genus	Relation to class Plant_Genus for description about genus of the plant
Botany_Habit	Relation to class Botany_Habit for description about botany habit of the plant
Floristic_Regions	Relation to class Plant_Habitat for description about address of the plant
Chemical_Information	Chemicals found in the plant
Physical_Information	Description physical characteristics of the plant
Genetic_Information	Description genetic of the plant
Description	Description about of the plant
Part_Used	Relation to class Uses for description about take advantage of the plant
Medicinal_Properties	Relation to class Medicinal_Properties for description about medicinal properties of the plant
Medicinal_Use	Description about medicinal use of the plant
Uses_and_Utilization	Description about use and utilization of the plant
Image	Image of the plant
Image_Caption	Image caption of the plant
References	Reference of the plant
Sub class	Herbaceous_Plant
Sub class	Shrub_Plant
Sub class	Vine_Plant

Table-3. Performance Categories and Distribution of theKnowledgeRetrieval

Query Category	Total			Precision	Recall	F-Measure
	Terms Manually	Found Terms	Method Terms			
Plant_information	45	43	43	1	0.96	0.98
Plant_Family	9	9	1	1	0.9	0.95
Plant_Genus	15	12	12	1	0.8	0.89
Plant_Habitat	10	6	5	0.83	0.5	0.63
Botany_Habit	10	8	8	1	0.8	0.89
Uses	20	18	18	1	0.9	0.95
Medicinal_Properties	25	24	24	1	0.96	0.98
Floristic_Regions	20	18	18	1	0.9	0.95
Total	155	138	137			
Average				0.98	0.84	0.9

We evaluate the system performance by calculating (1) the precision value which is the total matched terms divided by the total found terms, (2) the recall value which is the total matched terms divided by the total terms found manually and (3) the F-measure using the equation (Al-Nazer *et al.*, 2014):

The results of the knowledge retrieval showed that the semantic search application was effective regarding values of precision, recall, and F-measure. The application demonstrates that Thailand's Flora Ontology has been developed persistently with the concise encyclopedia of plants in Thailand. It provided functions of concept-based search system which improve efficiency of the information query by excluding non-relevant information items during query answering process.

In this part, we presents the semantic electronic dictionary of Thailand's Flora has search mode are 3 modes 1) normal search show in Figure 3., user can be input part of flora name and search result to display that flora information are flora name, Science name, Thai name, other name, botanical, details of flower, native land and references. 2) advance search show in Figure 4., this mode implementation of semantic search on Plant_Information concept and the result is processed with selected conditions. However the result can be link to related data follow the

concept of semantic technology, and 3) family search shown in Figure 5., this mode list of Flora Family, user can be select it for drill down, until found the name of flora do you want. Our system defined English language and Thai language.

Figure-3. Thailand's Flora Normal Search Testing

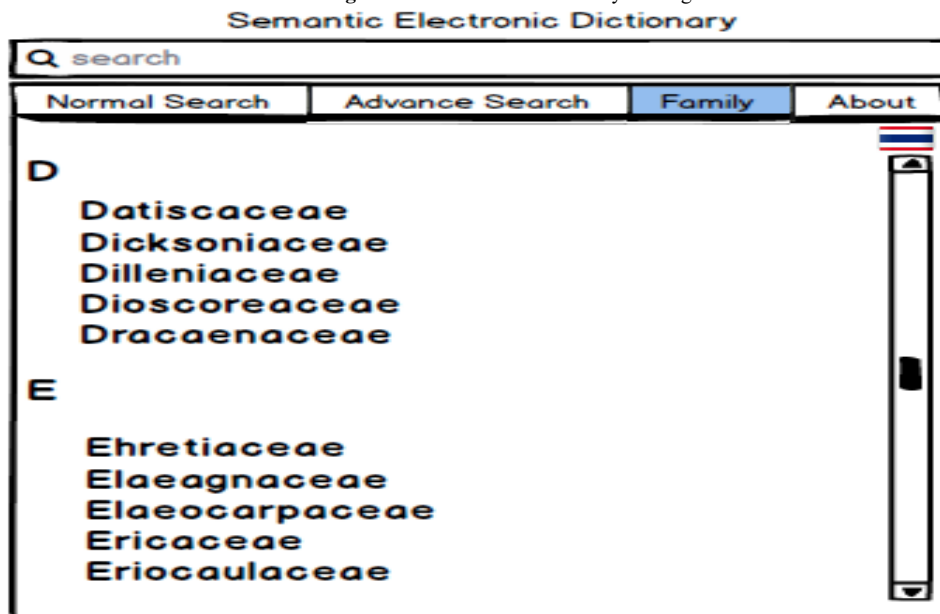**Figure-4.** Thailand's Advance Search Testing

Figure-5. Thailand's Flora Family Testing

6. Conclusion

This paper presented the ontologies of the Flora of Thailand. It will be further developed into ontology-based semantic retrieval system. The research findings can be concluded as follows: 1) The Thailand's Flora Ontology (TFO) of this research was conducted by the domain experts and team of ontology engineers based on the concise encyclopedia of plants in Thailand, 2) The TFO can be divided into 8 concepts including the Plant_Information, Plant_Family, Plant_Genus, Plant_Habitat, Botany_Habit, Uses, Medicinal_Properties, and Floristic_Regions. Existing flora classifications were used as guidelines for developing the ontologies, and 3) In the next step, semantic electronic dictionary on the mobile application will be developed by using the TFO from this study.

References

Al-Nazer, A., Helmy, T. and Al-Mulhem, M. (2014). User's profile ontology-based semantic framework for personalized food and nutrition recommendation. *Procedia Computer Science,* 32: 101-08.

Avraham, S. (2008). The Plant Ontology Database, A community resource for plant structure and developmental stages con-trolled vocabulary and annotations. *Nucleic Acids Res.,* 36(Database issue): D449–D54.

Buranarach, M., Myat, Y. and Supnithi, T., 2012. "A community-driven approach to development of an Ontology-based appli-cation management framework." In *Proceedings of the Second Joint International Conference (JIST2012), Nara, Japan, De-cember 2012.* pp. 306-12.

Chansanam, W. and Tuamsuk, K., 2016. "Development of imaginary beings Ontology, In Digital Libraries, Knowledge, Infor-mation, And data in an open access society." In *18th International Conference on Asia-Pacific Digital Libraries, ICADL 2016, Tsukuba, Japan, December 7–9, 2016, Proceedings, A. Mor-ishima, A. Rauber, and C. L. Liew, Eds. Cham: Springer Inter-national Publishing, 2016, pp. 231–242.*

Gu, F., Cao, C. G., Sui, Y. F. and Tian, W. (2004). Domain-specific ontology of botany. *Journal of Computer Science and Technology,* 19(2): 238–48.

Hoehndorf, R. (2016). The flora phenotype ontology FLOPO, Tool for integrating morphological traits and phenotypes of vascular plants. *J. Biomed. Semantics,* 7:

Noy, N. and McGuinness, D. (2000). *Ontology development 101, A guide to creating your first Ontology.* Stanford University.

Princeton University Word Net a lexical database for English. Available: http://wordnet.princeton.edu/

Saad, S., Salim, N., Zainal, H. and Muda, Z., 2011. "A process for building domain ontology, An experience in developing Solat ontology." In *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics, 2011.* pp. 1–5.

Shi, J. and Huang, L., 2009. "Development of flora-oriented plant Ontology." In *Second International Symposium on Knowledge Acquisition and Modeling, 2009.* pp. 7-10.

Thailand - Plant and Animal life (2017). History - geography, Ency-clopedia Britannica. Available: https://www.britannica.com/place/Thailand/Plant-and-animal-life

Thailand - Plant and animal life | history - geography", E. B. (2017). *Online. Available M. P. Brown and K. Austin*

The New Physique Publisher Name, Publisher City, 2005. 25–30. https://www.britannica.com/place/Thailand/Plant-and-animal-life

The Plant Ontology Consortium (2002). The Plant Ontology™ Con-sortium and Plant Ontologies. *Comp. Funct. Genomics,* 3(2): 137–42.

van Heijst, G., Schreiber, A. T. and Wielinga, B. J. (1997). Using ex-plicit ontologies in KBS development. *International Journal of Human-Computer Studies,* 46(2): 183–292.

Westphal, D. and Bizer, C. Introduction to RAP.: Available: http://wifo5-03.informatik.uni-mannheim.de/bizer/rdfapi/tutorial/introductionToRAP.htm