ONTOLOGICAL PLANT DATA VISUALISATION USING NETWORK GRAPHS

AFRINA ADLYNA BINTI MOHAMAD MATROL

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2019

ONTOLOGICAL PLANT DATA VISUALISATION USING NETWORK GRAPHS

AFRINA ADLYNA BINTI MOHAMAD MATROL

DISSERTATION SUBMITTED IN FULFILMENT OF
THE REQUIREMENT FOR THE DEGREE OF
MASTER OF SCIENCE

INSTITUTE OF BIOLOGICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2019

UNIVERSITY OF MALAYA
ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **AFRINA ADLYNA BINTI MOHAMAD MATROL**

Matric No: **SGR160035**

Name of Degree: **MASTER OF SCIENCE**

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

ONTOLOGICAL PLANT DATA VISUALISATION USING NETWORK GRAPHS

Field of Study: **BIOINFORMATICS**

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

ONTOLOGICAL PLANT DATA VISUALISATION USING NETWORK GRAPHS

ABSTRACT

Plant data is heterogeneous, containing complex pictures and consisting of many terminologies that describe plants typically in textual and image forms. The advancement in information technology has led to the development of online database systems, through which the plant data can be shared; accessed and related information to the users' query be retrieved. However, the retrieved plant data are often presented in lengthy textual and table forms. Consequently, there is inefficiency in elucidating the relationships between plant data in which obtaining new insight from the presented data can be difficult to the users. It should be emphasized that underlining the relationships between data are very important for knowledge enrichment. For the present study, a visual-based representation is proposed to display data to the users in a meaningful way, as it emphasizes the relationships between the data. It involves two main development components; the developments of ontology-driven plant database and plant data visualisation system. Ontological data for plants and samples of trees and shrubs are used as the dataset to be integrated into the proposed plant data visualisation system. The efficiency of the developed visualisation system is measured by performing three types of user evaluation; the usability of heuristic, query and visualisation evaluations, done by the expert and novice users. An ontology-driven plant database named Plant Ontology Universiti Malaya (POUM) and a visualisation system named PlantViz are then developed. POUM consists of plant data and images of 222 plant samples from 43 species of 42 genus for trees and 31 species of 28 genus for shrubs collected around the University of Malaya. PlantViz provides a graphical user interface for users to query the POUM and a graphical viewer to display the results of the query in a form of network graph. The relationships between the data are shown in visualisation form so that users can infer the knowledge and correlate between the data easily. The results from the user evaluation show that the proposed visualisation system is suitable for both users, with or without computer skills. This technique demonstrated the practicability of using computer-assisted tool by providing cognitive analysis in understanding the relationships between data.

Keyword: plant, ontology-driven, visualisation, POUM, PlantViz

VISUALISASI ONTOLOGI DATA TUMBUHAN MENGGUNAKAN GRAF RANGKAIAN

ABSTRAK

Data tumbuhan adalah heterogen, mengandungi gambar kompleks dan banyak istilah digunakan untuk menerangkan tumbuhan yang biasanya dibentangkan dalam bentuk tekstual dan imej. Kemajuan dalam teknologi maklumat telah membawa kepada pembangunan sistem pangkalan data dalam talian. Melalui pangkalan data ini, data tumbuhan boleh dikongsi; diakses dan maklumat yang berkaitan dengan pertanyaan pengguna boleh diperolehi. Walau bagaimanapun, data tumbuhan yang diperolehi sering kali dibentangkan dalam format teks yang panjang dan bentuk jadual. Akibatnya, terdapat ketidakcekapan dalam menjelaskan hubungan antara data tumbuhan di mana mendapatkan pandangan baru dari data yang dibentangkan mungkin sukar bagi pengguna. Ia harus ditekankan, menggariskan hubungan antara data sangat penting untuk pengayaan pengetahuan. Untuk kajian ini, perwakilan berasaskan visual dicadangkan untuk memaparkan data kepada pengguna dengan cara yang bermakna, di mana ia menekankan pada hubungan antara data. Ia melibatkan dua komponen pembangunan utama iaitu pembangunan pangkalan data berasaskan ontologi dan pembangunan sistem visualisasi data tumbuhan. Ontologikal data tumbuh-tumbuhan dan sampel yang terdiri daripada pokok dan pokok renek telah digunakan sebagai dataset untuk di integrasikan ke dalam sistem visualisasi tumbuhan yang dicadangkan. Kecekapan sistem visualisasi yang dibangunkan diukur dengan melaksanakan tiga jenis penilaian pengguna yang merupakan penilaian heuristik kegunaan, penilaian pertanyaan dan penilaian visualisasi. Penilaian telah dilakukan oleh para pakar dan pengguna baru. Pangkalan data tumbuhan berasaskan ontology yang dinamakan Plant Ontology Universiti Malaya (POUM) dan sistem visualisasi yang dinamakan PlantViz kemudiannya dibangunkan. POUM terdiri daripada data tumbuhan dan imej 222 sampel tumbuhan daripada 43 spesies 42 genus untuk pokok dan 31 spesies 28 genus untuk pokok renek yang dikumpulkan di Universiti Malaya. PlantViz menyediakan grafik antara muka pengguna yang membolehkan pengguna menanyakan di POUM dan penampil grafik untuk memaparkan hasil pertanyaan dalam bentuk graf rangkaian. Hubungan antara data yang ditunjukkan dalam bentuk visualisasi membolehkan pengguna untuk mengetahui pengetahuan dan korelasi antara data dengan mudah. Hasil daripada penilaian pengguna menunjukkan bahawa sistem visualisasi yang dicadangkan sesuai untuk pengguna pakar dan pengguna baru, dengan atau tanpa kemahiran komputer. Teknik ini menunjukkan kebolehgunaan menggunakan alat bantuan komputer dengan menyediakan analisis kognitif dalam memahami hubungan antara data.

Kata kunci: tumbuhan, berasaskan ontologi, visualisasi, POUM, PlantViz

ACKNOWLEDGEMENTS

I would like to thank Allah for giving me the chance to do this research and blesses me with strength and guidance in times of difficulty throughout this project.

Infinite gratitude is extended to my parents, my family, and my friends for their never-ending support and belief in me. I would never accomplish this if not for their motivational talks and support.

I would like to express my gratitude and appreciations to my supervisors, Dr. Arpah bt Abu and Dr. Chang Siow Wee for their trust and guidance for this study.

I am also very grateful to University of Malaya for the financial support under the University of Malaya Research Grant (RP038B-15AET) over this study.

Lastly, I would like to express my appreciation to everyone that has helped me in all kinds of forms during this project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

CAD	: Computer Aided Design
ChEBI	: Chemical Entities of Biological Interest
CSS	: Cascading Style Sheets
DAML	: DARPA Agent Markup Language
DBMS	: Database Management System
DL	: Description Logic
GO	: Gene Ontology
GUI	: Graphical User Interface
HTML	: Hypertext Markup Language
IDE	: Integrated Development Environment
IRI	: Internationalized Resource Identifier
IT	: Information Technology
JSON	: JavaScript Object Notation
OIL	: Ontology Inference Layer
OWL	: Web Ontology Language
OWL2	: Web Ontology Language version 2
PlantViz	: Plant Visualisation
PO	: Plant Ontology
POUM	: Plant Ontology Universiti Malaya
RDF	: Resource Description Framework
RDFS	: RDF Schema
SALO	: Saliva Ontology
SQL	: Structured Query Language
SVG	: Scalable Vector Graphics

TBC	: Top Braid Composer
UM	: University of Malaya
W3C	: World Wide Web Consortium

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 Background

Plants are multicellular, autotrophic organisms playing an important role for all forms of life, including mankind. The number of plant species is increasing every year and currently, it is estimated that there are more than 300,000 plant species in the world as stated in Willis (2017).

In general, plants species are classified to Angiosperms; which produces flowers, and Gymnosperms; which do not (Campbell et al., 2003). Plants start their life from seeds and then the roots, stems, flowers, fruits and leaves grow (Arteca, 2003). Hence, morphological descriptions such as the root system, the flowering types, and the growth pattern together with taxonomic classification are commonly used to describe plants up to their species. Furthermore, plants are also described through their systematics, behaviour, ecology, diversity and geological distribution. These plant data therefore are commonly in textual and image forms.

It is common to find plant data in paper-based materials or physical documents such as log books, text books (Barrett & Tay, 2016; Hussain et al., 2015; Said et al., 2001), articles in journals (Sreetheran et al., 2011; Webb, 1998), magazines and files in drawers or cabinets. However, the amount of data produced over the years is increasing due to technological advancement and easy access to experimental tools (Marx, 2013). Thus, with the emerging research and information technology (IT), plant data have been stored in online databases such as PlantSearch (Botanic Gardens Conservation International, 2017), TRY Plant Trait Database (Kattge et al., 2011), PLANTS Database (USDA, 2017), Native Plant Database (Evergreen, 2017), maizeGDB (Andorf et al., 2016), MyBIS (MyBIS, 2017) and MyCHM (FRIM, 2017) for data accessibility, sharing and retrieval purposes.

Thus, for the said purpose of online database, many online databases provide the common features such as query and search tool, and present the retrieved results to the user. The data presentation of retrieved results from the database is typically in the textual form, for instance in paragraph, table and list (Bisby et al., 2010; Botanical Missouri Garden, 2018; Evergreen, 2017; FRIM, 2017). Eventhough these are the typical types of data presentation that can be easily viewed by the users to gain information, on the other hand, it may cause difficulties to them in interpreting and analysing the data to gain answer to questions such as 'what is/are the common similarity/ies between the species A1 and A2, since both are the same genus' or 'to see the common difference/s on plant morphology for some species from the same sampling area'. Therefore, an extra task might have to be performed by the users to answer their questions.

The advancement in multidisciplinary field has advanced data visualisation, in a way that the retrieved results from database are presented in graphical form. For instance, plant data are presented in image form (JSTOR, 2018; MyBIS, 2017), statistical graphs (Dash et al., 2012; Lee et al., 2009; Petryszak et al., 2016) and map drawings (Kattge et al., 2011; USDA, 2017), which helps users to deduce a new insight of the plant because the data and their relationships are visualised.

Data visualisation is an interdisciplinary field (Telea, 2014) that conveys the unique properties of the data. It is common in biological field especially for biological-based software (Joachimiak et al., 2006; Junker et al., 2006; Kearse et al., 2012), biological web-based system (Cline et al., 2007; Kasprzyk et al., 2004) and biological-based visual library (Gómez et al., 2013; Schmid et al., 2010).

Concisely, there are three main components in creating data visualisation which are type of visualisation, data visualisation tools, and the most important component, the data modelling for data representation.

The type of data visualisation is important to emphasize the relationships between the data such as the relationships between taxonomical data and samples. Besides that, most of the data visualisation has interactive features, for example zooming (Barrett et al., 2011, Junker et al., 2006), filtering tool (Bingham & Sudarsanam, 2000) and export tool (Ashburner et al., 2000; Avraham et al., 2008), whereby the users can manipulate and explore the graphical viewer, rather than a set of fixed diagrams.

The key to effective data visualisation is the combination of functioning visualisation tools such as rendering maps (Agafonkin, 2017), visualizing complex networks of data (Cline et al., 2007), and generating charts with interactive features (Teller, 2013). A good visual library which consists of a set of programming languages helps in designing any kind of visualisation that needs timeline (Barrett et al., 2011; Secrier et al., 2012), flow chart (Blasi et al., 2011; Brohée et al., 2008), alluvial diagram (Eren et al., 2015; Ruan et al., 2017), and network graph (Cline et al., 2007; Junker et al., 2006; Wang et al., 2013).

Data modelling is very crucial because it controls the logic flow of the data retrieval before the data is visualized. Albeit there are many data modelling concepts such as relational model (Codd, 1970), object-oriented model (Rumbaugh et al., 1991), and graph data model (Kunii, 1990), only the graph data model is the most suitable such as ArangoDB (ArangoDB, 2014), MarkLogic (MarkLogic, 2017), and OrientDB (Tesoriero, 2013) which can be an alternative method in database design through ontology (Martinez-Cruz et al., 2012). Ontology is defined as the specification of conceptualization (Gruber, 1993) which expresses the knowledge in terms of entities in a concrete form. In other words, ontology is referred as the formalization of domain knowledge (Ahmad, 2012). The usage of ontology as the database model in biological field can be seen in many online databases such as Gene Ontology (GO) (Ashburner et al., 2000; GO Consortium, 2017), Plant Ontology (PO) (Avraham et al., 2008; Cooper et al., 2018), IDOMAL (Topalis et al., 2010), Saliva Ontology (Ai et al., 2010), and Textpresso (Müller et al., 2004).

Briefly, plant data is heterogenous, containing complex pictures and many terminologies to describe the plant and typically presented in texts and images. Plant data is as important as other biological data as plant plays a major role in the ecology. Different types of plant data such as morphological description of parts of plant, taxonomical data, and images of plants help to differentiate plant species. Plant data can be obtained from manuscripts and online databases. However, in comparison, there are only a small number of plant databases compared to other biological-based databases (Galperin et al., 2017) and majority of them present the plant data in texts.

This research focuses on the ontological plant data visualisation using network graphs. In this research, ontology-driven plant database, named **P**lant **O**ntology **U**niversiti **M**alaya (POUM) is developed. POUM consists of tree and shrub data collected from University of Malaya (UM), Kuala Lumpur, Malaysia. This ontological data for plants and samples of trees and shrubs becomes the dataset to be integrated into the proposed plant data visualisation system using network graph, named **Plant Vis**ualisation (PlantViz). The purpose of presenting plant data in data visualisation using network graphs is to provide an alternative way of presenting data to the users, which is useful for better data understanding as it improves data interpretation and analysis.

1.2 Problem Statement

As stated previously, there are many biological online databases exist which focus on animal or plant data. Specifically, the plant-based databases consist of many types of plant data such as taxonomic data (Evergreen, 2017; Malaysia Botanical Garden, 2018; The Plant List, 2013; USDA, 2017), sample data (Andorf et al., 2016; County ITS, 2015; USDA, 2017), genomic data (Andorf et al., 2016; Botanic Gardens Conservation International, 2017; FRIM, 2017), sequence data (Andorf et al., 2016; Kattge et al., 2011) in textual and image forms.

However, data retrieved from these databases are commonly presented in the forms of list, table and lengthy textual or paragraph. For instance, Evergreen Native Plant Database (Evergreen, 2017), Lady Bird Johnson Wildflower Center (Lady Bird Johnson Wildflower Center, 2018), Colorado Plant Database (County ITS, 2015) and MyCHM (FRIM, 2017) only provide lengthy textual description which requires more time for the users to digest the information and thus diminishing the value of the information and restrict the users in inferring new perception of the plant species. Whereas, databases such as NParks Flora & Fauna Web (NParks Flora & Fauna Web, 2013), Malaysia Botanical Garden (Malaysia Botanical Garden, 2018), and MyBIS (MyBIS, 2017) only present their data per individual species without highlighting the unique relationships between plant species which devalues the information as there is no knowledge that can be inferred.

In addition, most plant databases exist independently, in textual form (FRIM, 2017; USM, 2018) or image form (Malaysia Botanical Garden, 2018; MyBIS, 2017), thus the users need to switch between databases frequently before the retrieved data can be gathered into the proper forms for data interpretation and analysis to see the relationships between data.

For Malaysian plants, there is no database providing data visualisation with interactive features to present the data in textual and image forms. This research is concerned with data visualisation of ontological plant data. Data visualisation elucidates the relationships between plant data to obtain new insight from the presented data.

1.3 Objectives

This research aims to produce a prototype of ontological plant data visualisation for the knowledge enrichment. Thus, it is vital that the following objectives are successfully achieved.

i) To examine the vocabularies to describe the plant data and their relationships in both textual and image forms

ii) To develop an ontology-driven plant data management

iii) To present ontological plant data in a visualisation form using network graphs

iv) To evaluate the usability of the developed visualisation system

1.4 Scope of the Study

In this research, plant data consists of plant species and their samples in both textual and image forms, annotated with taxonomical classification, morphological characteristics, ecological attributes and geological distribution which produces ontological plant data, POUM.

Based on the ontological plant data, three relationships between data are emphasized as the followings: -

i) The relationships between one taxon to another taxon

ii) The relationships between taxa and the taxa's sample

iii) The relationships between samples

These relationships are visualised through the graphical viewer in the data visualisation system, PlantViz. Few interactive elements are featured as well to encourage two-way communication between the users and the data whereby it allows data manipulation by the users.

In the testing phase, it involves both expert and novice users to evaluate the usability of the data visualisation system.

1.5 Research Significances

The main purpose of this research is to provide another alternative technique in presenting plant data for the online plant database through data visualisation.

Data visualisation incorporates cognitive references to users (Rowley & Hartley, 2008) which allows the users to analyse the information given and deduce new inference for new knowledge finding. Furthermore, data visualisation is a dynamic data presentation whereby it comes with interactive elements for the users to manipulate the data according to their preferences, thus facilitating data inference.

This research offers a more effective technique in improving data interpretation and analysis for knowledge enrichment. Graphical data presentation allows users to view the data as a whole without any cluttering of texts. Moreover, it is easier for users to study relationships between data as data visualisation emphasizes on correlation between related data. For instance, expert users such as researchers and lecturers or novice users such as layman and students can benefit from this alternative data presentation as the interactive elements in data visualisation gives the users better understanding.

The other significance of this research is the ontological plant data using graph data model, which describes the relationships between the data and thus provides a fortified definition and well described data. The flexibility of ontology allows further extension without changing the whole data structure, useful for future advancement.

1.6 Chapter Organization

This thesis is divided to six chapters. **Chapter 1** provides a summary of topics concerning this research, objectives, scope, and its significances.

Chapter 2 describes the literature review of this research. It presents the finding of plant and plant data, current existing database systems, available tools and related issues regarding data visualisation and data management using ontology.

Chapter 3 presents the analysis from the literature review where details of problems focused in this research are determined. Issues regarding organization of plant data and data presentation approaches are further discussed. This chapter also discusses the outline of the proposed solution and the functional and non-functional requirements.

Chapter 4 presents the methodologies and techniques used in the system development based on the proposed solution. It includes the overall system architecture, description of the development environment and tools involved, and system testing and evaluation that is performed.

Chapter 5 discusses the result and discussion regarding the implemented system. It also describes the strengths and limitation of the implemented system architecture.

Finally, in **Chapter 6**, presents the final conclusion of this research. Future enhancements to improve the system capability that can be done are discussed and presented at the end of this chapter.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

This chapter presents the review done on previous and current literatures obtained from online search via Internet and from reading materials such as e-books, academic journals, and reference books, revolving around plant data, plant databases, data modelling techniques for database development, common data presentation techniques, and data visualisation including its development components.

2.2 Plant Data

Plant is a multicellular, autotrophic organism that is capable of photosynthesis. The total number of plant species worldwide is still unknown as new species are still discovered every year. The species that have the same scientific name and from different locations with different common names complicate this matter. There are approximately more than 300,000 of known species (Willis, 2017). In general, plant species are classified to two groups which are plants producing flowers namely Angiosperms and plants that do not, namely Gymnosperms (Campbell et al., 2003). Cultivated plant is another form of plant species that have been selected for cultivation or hybridized in an environment that is not in its natural habitat (Brittain, 2005) and it is estimated that there are about 35,000 of cultivated plant species (Khoshbakht & Hammer, 2008).

The extensive range of plant species is due to the high differences between each plant species. To the naked eyes, appearances between one plant species to another do not have any dissimilarity. Yet, there is a high distinction of plant species in term of its morphological and physiological forms. Numerous plants grow from seeds (Arteca, 2013). Some plants may have flowers and fruit, and others do not. Roots facilitate water and mineral absorption from the soil. Stems support the plant as well as conducting water and nutrients from the roots to other parts of the plant. Leaves are where most plant's food is made; leaves capture sunlight through photosynthesis and use it to make food (Reece et al., 2013). Flowers and fruits are important to plant growth as flowers contain pollen for pollination and fruit contain seeds of the plant (Abrol, 2011).

Moreover, characteristics of plants are not only limited to their external structural such as leaf, stem, and root, but other information such as genomic, physiological and ecological information are pivotal too. These types of information are useful as it provides better understanding of how plants function. Increasing number of plant-related researches and experiments increases the amount of plant data. In addition, the increasing development of genotyping techniques (Agarwal et al., 2008) has generated a large amount of plant genetic data. These factors demand a comprehensive and structured information resources as well as detailed analysis among different species to provide new insights into specific characteristics of individual plant (Spannagl et al., 2007), and can be fulfilled by organizing information about plant data into a database.

2.3 Plant Data Sources

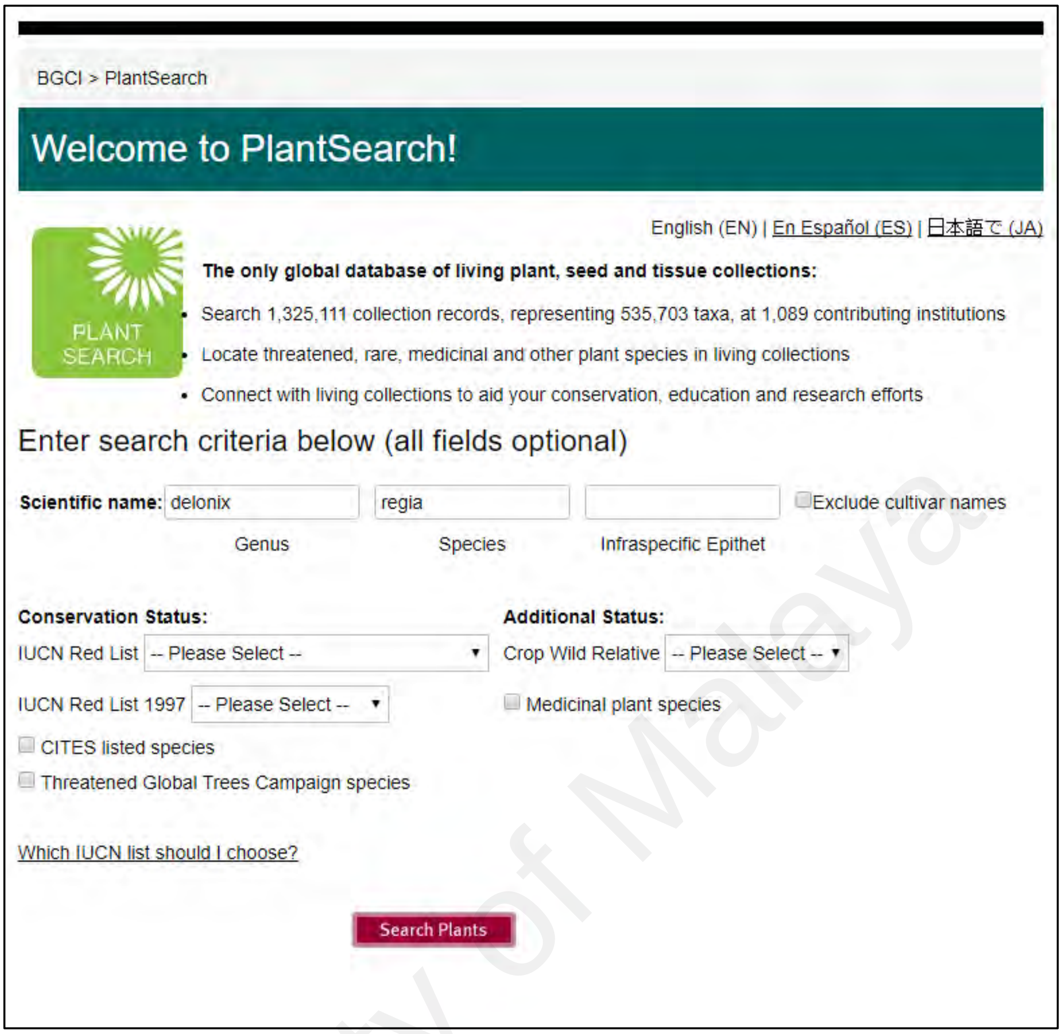
Plant data is commonly found in the paper-based materials and physical documents such as log books, text books (Barrett & Tay, 2016; Hussain et al., 2015; Said et al., 2001), articles in journals (Sreetheran et al., 2011; Webb, 1998), magazines and files in drawers and cabinets. However, the amount of data produced over the years is increasing due to the advancement in technology and easy access to experimental tools (Marx, 2013). Thus, the plant data have been stored into the online databases for data accessibility, sharing and retrieval purposes.

2.3.1 Plant Database Systems

There are a number of public plant databases. Some databases only contain textual information or images separately, and some contain image with annotated textual information. The following are examples of plant database systems that had been reviewed in this research.

i) PlantSearch (https://www.bgci.org/plant_search.php)

PlantSearch (Botanic Gardens Conservation International, 2017) is an online database that contains taxon-level information of plant, seed and tissue. It is the only global database for botanic gardens plant species and tracks threatened species of botanical plant as it also includes plant's conservation status. In addition, it also connects data from external plant database website to facilitate users for extra information. Currently, it contains more than 1.3 million records from more than 1000 contributing institutions. Figure 2.1 shows the search tool where the users can perform textual based search to search for the scientific name of plant species and filter the results by the conservation status and the relative of the plant species. Per Figure 2.2, data retrieved from the database are displayed in a simple list, divided to taxonomical classification of the plant species, conservation status, and links of the plant species in other external databases. A significant feature of this database is it provides links to other external databases and the total number of *ex situ* botanical collections of the plant species available worldwide.



BGCI > PlantSearch

Welcome to PlantSearch!

English (EN) | En Español (ES) | 日本語で (JA)

The only global database of living plant, seed and tissue collections:

- Search 1,325,111 collection records, representing 535,703 taxa, at 1,089 contributing institutions
- Locate threatened, rare, medicinal and other plant species in living collections
- Connect with living collections to aid your conservation, education and research efforts

Enter search criteria below (all fields optional)

Scientific name: delonix (Genus) regia (Species) (Infraspecific Epithet) ☐ Exclude cultivar names

Conservation Status:

IUCN Red List: -- Please Select --

IUCN Red List 1997: -- Please Select --

☐ CITES listed species

☐ Threatened Global Trees Campaign species

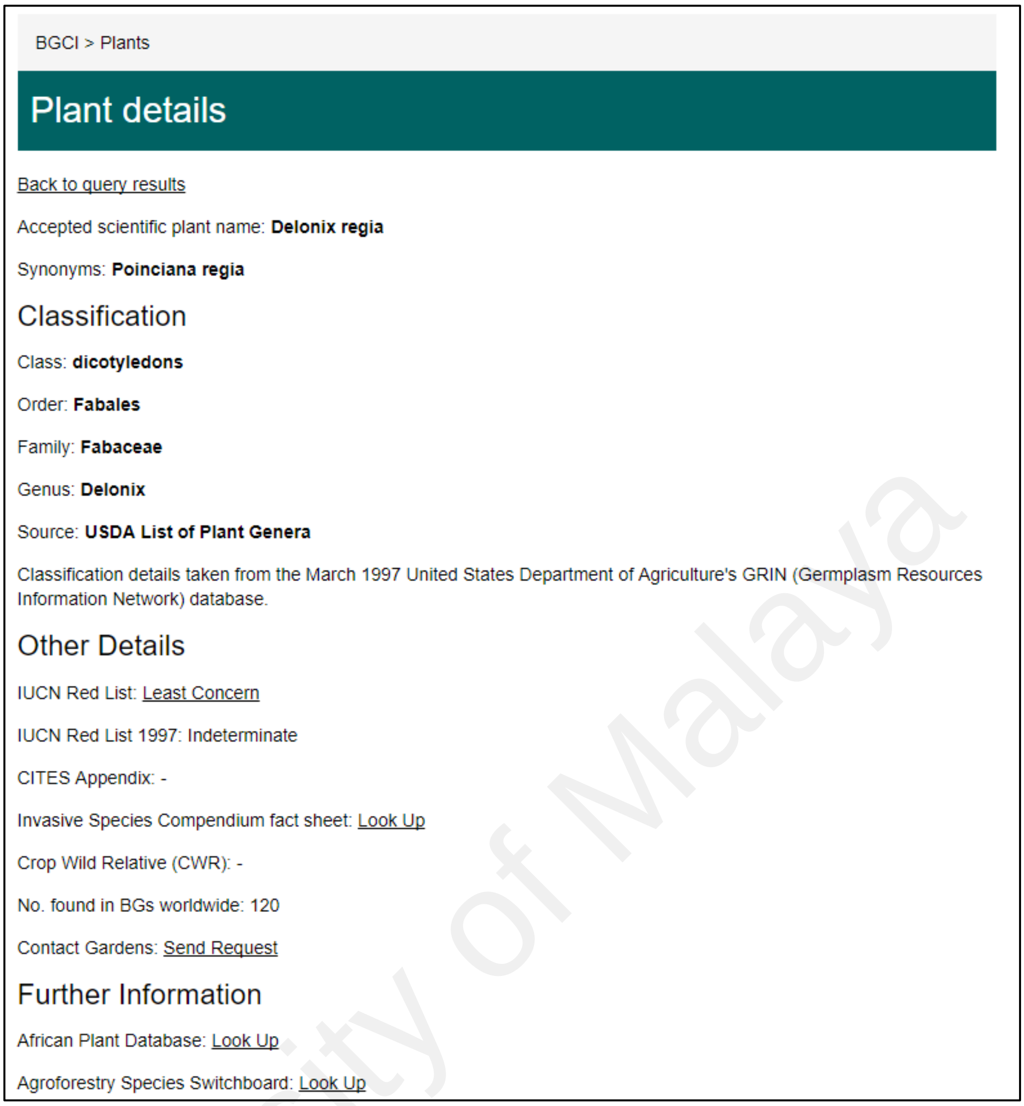
Which IUCN list should I choose?

Additional Status:

Crop Wild Relative: -- Please Select --

☐ Medicinal plant species

Figure 2.1: Plant search tool



BGCI > Plants

Plant details

Back to query results

Accepted scientific plant name: **Delonix regia**

Synonyms: **Poinciana regia**

Classification

Class: **dicotyledons**

Order: **Fabales**

Family: **Fabaceae**

Genus: **Delonix**

Source: **USDA List of Plant Genera**

Classification details taken from the March 1997 United States Department of Agriculture's GRIN (Germplasm Resources Information Network) database.

Other Details

IUCN Red List: Least Concern

IUCN Red List 1997: Indeterminate

CITES Appendix: -

Invasive Species Compendium fact sheet: Look Up

Crop Wild Relative (CWR): -

No. found in BGs worldwide: 120

Contact Gardens: Send Request

Further Information

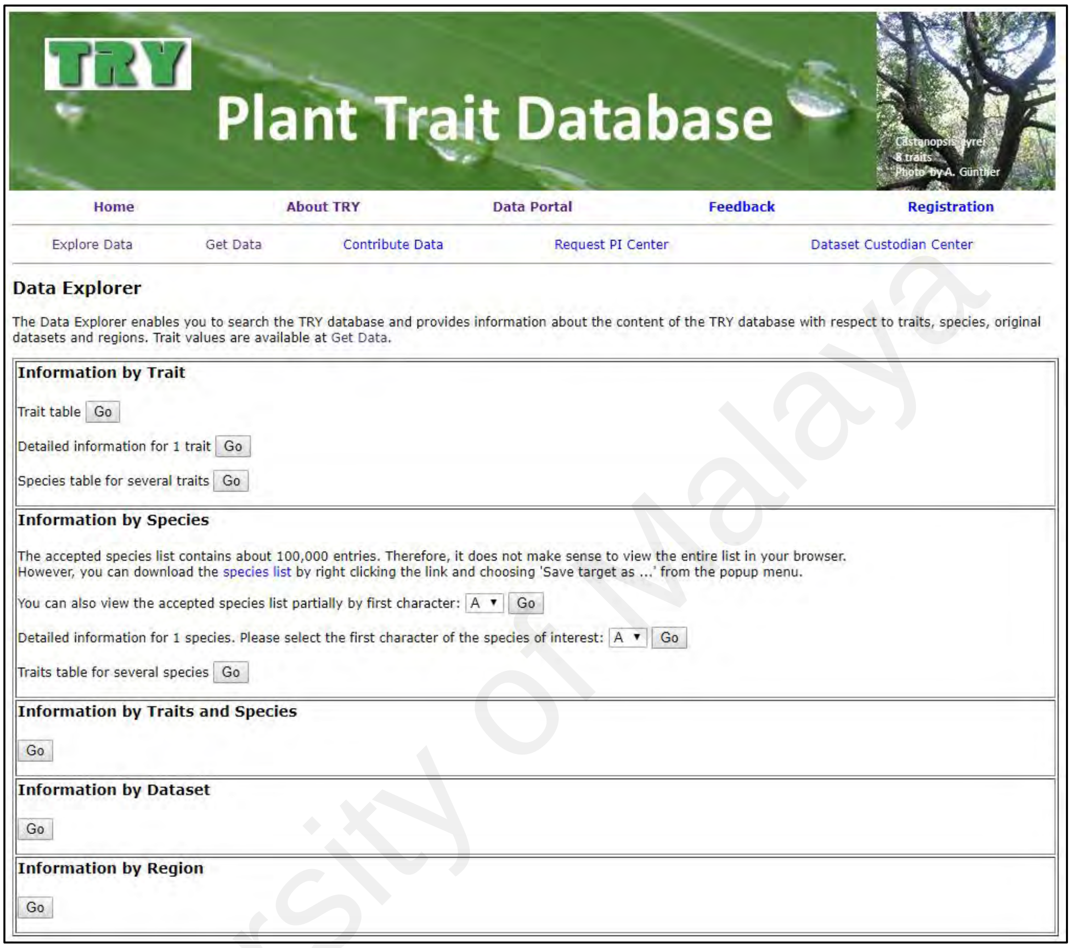
African Plant Database: Look Up

Agroforestry Species Switchboard: Look Up

Figure 2.2: PlantSearch query result

ii) TRY Plant Trait Database (https://www.try-db.org/TryWeb/Home.php)

TRY Plant Trait Database (Kattge et al., 2011) is an open access database consists of curated plant trait data with more than 6 million trait records for 148,000 plant taxa currently. The purpose of the development of this database is to improve the observed plant data and bring together different plant trait database at one centre. As shown in Figure 2.3, this database provides a textual based search where the users can search for data on traits, species, or dataset submitted by institutions or region. Results of the query are displayed in tables and map image as shown in Figure 2.4 and Figure 2.5 respectively. Users can download the results in tab delimited text format.



Data Explorer

The Data Explorer enables you to search the TRY database and provides information about the content of the TRY database with respect to traits, species, original datasets and regions. Trait values are available at Get Data.

Information by Trait

Trait table Go

Detailed information for 1 trait Go

Species table for several traits Go

Information by Species

The accepted species list contains about 100,000 entries. Therefore, it does not make sense to view the entire list in your browser.
However, you can download the species list by right clicking the link and choosing 'Save target as ...' from the popup menu.

You can also view the accepted species list partially by first character: A ▾ Go

Detailed information for 1 species. Please select the first character of the species of interest: A ▾ Go

Traits table for several species Go

Information by Traits and Species

Go

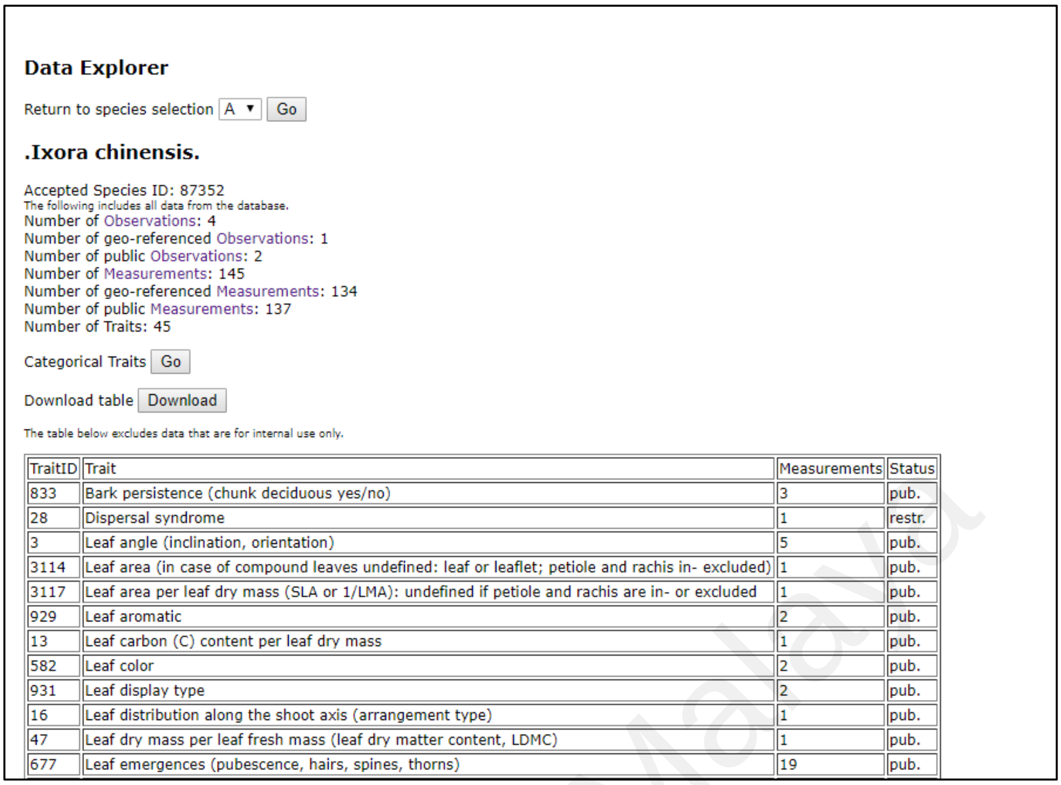
Information by Dataset

Go

Information by Region

Go

Figure 2.3: TRY Plant Trait Database result shown in the table form



Data Explorer

Return to species selection A ▾ Go

.Ixora chinensis.

Accepted Species ID: 87352
The following includes all data from the database.
Number of Observations: 4
Number of geo-referenced Observations: 1
Number of public Observations: 2
Number of Measurements: 145
Number of geo-referenced Measurements: 134
Number of public Measurements: 137
Number of Traits: 45

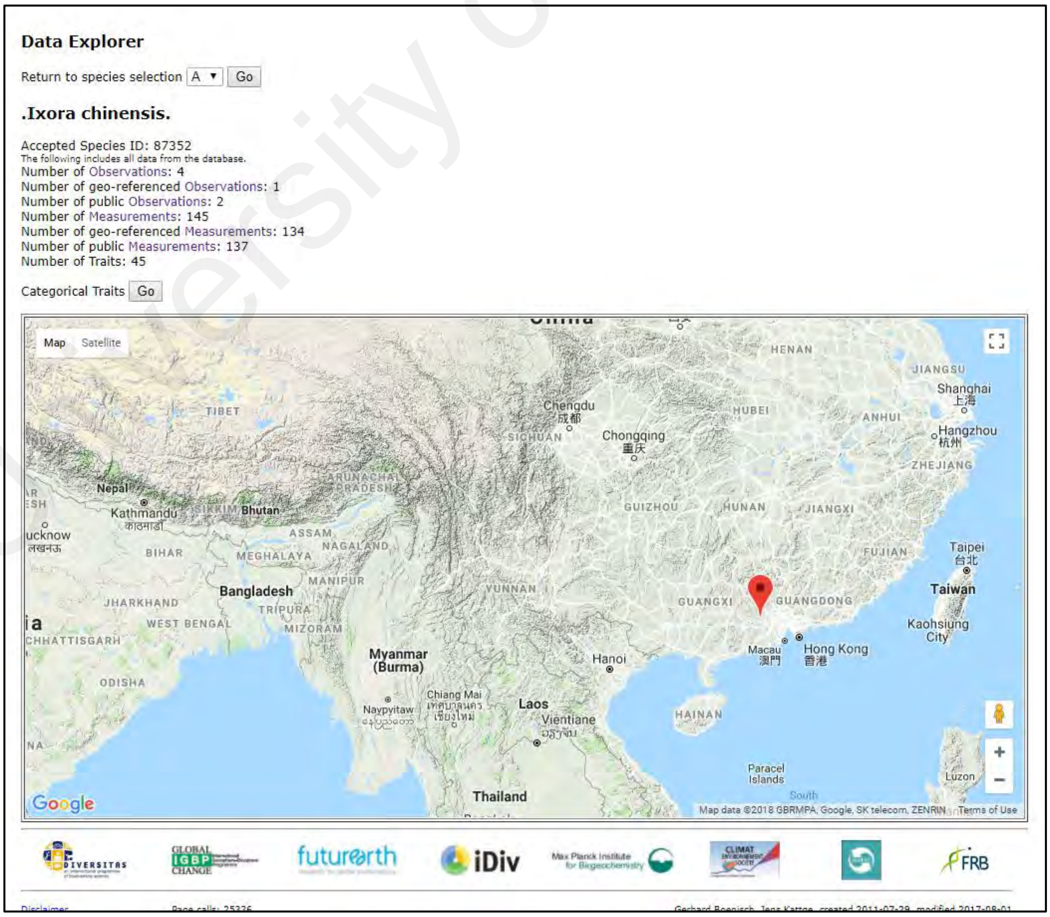
Categorical Traits Go

Download table Download

The table below excludes data that are for internal use only.

TraitID	Trait	Measurements	Status
833	Bark persistence (chunk deciduous yes/no)	3	pub.
28	Dispersal syndrome	1	restr.
3	Leaf angle (inclination, orientation)	5	pub.
3114	Leaf area (in case of compound leaves undefined: leaf or leaflet; petiole and rachis in- excluded)	1	pub.
3117	Leaf area per leaf dry mass (SLA or 1/LMA): undefined if petiole and rachis are in- or excluded	1	pub.
929	Leaf aromatic	2	pub.
13	Leaf carbon (C) content per leaf dry mass	1	pub.
582	Leaf color	2	pub.
931	Leaf display type	2	pub.
16	Leaf distribution along the shoot axis (arrangement type)	1	pub.
47	Leaf dry mass per leaf fresh mass (leaf dry matter content, LDMC)	1	pub.
677	Leaf emergences (pubescence, hairs, spines, thorns)	19	pub.

Figure 2.4: TRY Plant Trait Database search tool



Data Explorer

Return to species selection A ▾ Go

.Ixora chinensis.

Accepted Species ID: 87352
The following includes all data from the database.
Number of Observations: 4
Number of geo-referenced Observations: 1
Number of public Observations: 2
Number of Measurements: 145
Number of geo-referenced Measurements: 134
Number of public Measurements: 137
Number of Traits: 45

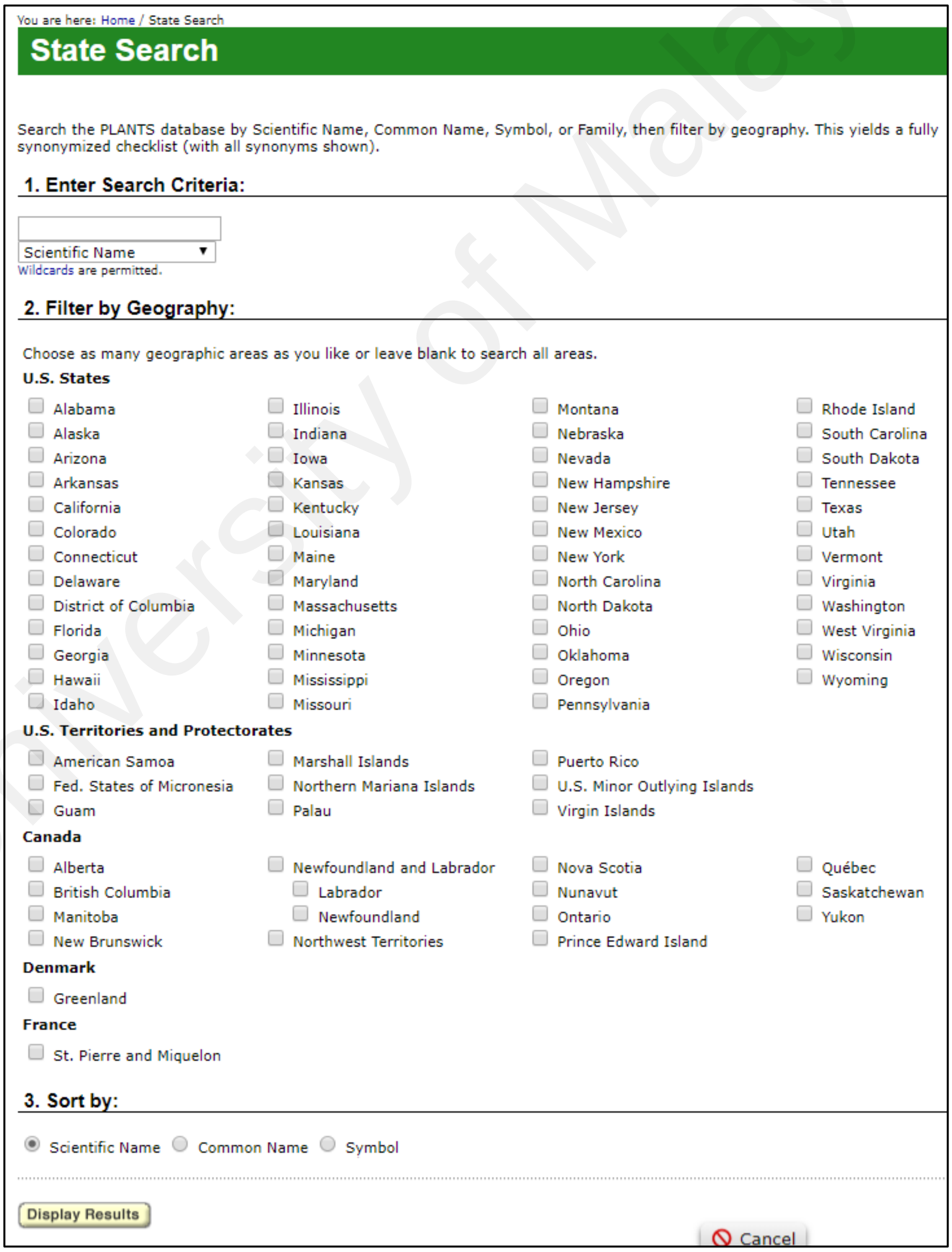
Categorical Traits Go

Disclaimer Page calls: 25336 Gerhard Boenisch, Jens Kattge, created 2011-07-29, modified 2017-08-01

Figure 2.5: TRY Plant Trait Database result shown in map image

iii) PLANTS Database (https://plants.usda.gov/java/)

PLANTS Database (USDA, 2017) is an open access database with textual and images information of vascular plants, mosses, liverworts, hornworts and lichens of the United States and its territories. Figure 2.6 shows the search tool where there are four parameters for users to search for which are 'Scientific Name', 'Common Name', 'Symbol' and 'Family' and users can filter the query result based on the geographic areas or sort by 'Scientific Name', 'Common Name', or 'Symbol'.



You are here: Home / State Search

State Search

Search the PLANTS database by Scientific Name, Common Name, Symbol, or Family, then filter by geography. This yields a fully synonymized checklist (with all synonyms shown).

1. Enter Search Criteria:

Scientific Name ▼

Wildcards are permitted.

2. Filter by Geography:

Choose as many geographic areas as you like or leave blank to search all areas.

U.S. States

□ Alabama
□ Alaska
□ Arizona
□ Arkansas
□ California
□ Colorado
□ Connecticut
□ Delaware
□ District of Columbia
□ Florida
□ Georgia
□ Hawaii
□ Idaho
□ Illinois
□ Indiana
□ Iowa
□ Kansas
□ Kentucky
□ Louisiana
□ Maine
□ Maryland
□ Massachusetts
□ Michigan
□ Minnesota
□ Mississippi
□ Missouri
□ Montana
□ Nebraska
□ Nevada
□ New Hampshire
□ New Jersey
□ New Mexico
□ New York
□ North Carolina
□ North Dakota
□ Ohio
□ Oklahoma
□ Oregon
□ Pennsylvania
□ Rhode Island
□ South Carolina
□ South Dakota
□ Tennessee
□ Texas
□ Utah
□ Vermont
□ Virginia
□ Washington
□ West Virginia
□ Wisconsin
□ Wyoming

U.S. Territories and Protectorates

□ American Samoa
□ Fed. States of Micronesia
□ Guam
□ Marshall Islands
□ Northern Mariana Islands
□ Palau
□ Puerto Rico
□ U.S. Minor Outlying Islands
□ Virgin Islands

Canada

□ Alberta
□ British Columbia
□ Manitoba
□ New Brunswick
□ Newfoundland and Labrador
□ Labrador
□ Newfoundland
□ Northwest Territories
□ Nova Scotia
□ Nunavut
□ Ontario
□ Prince Edward Island
□ Québec
□ Saskatchewan
□ Yukon

Denmark

□ Greenland

France

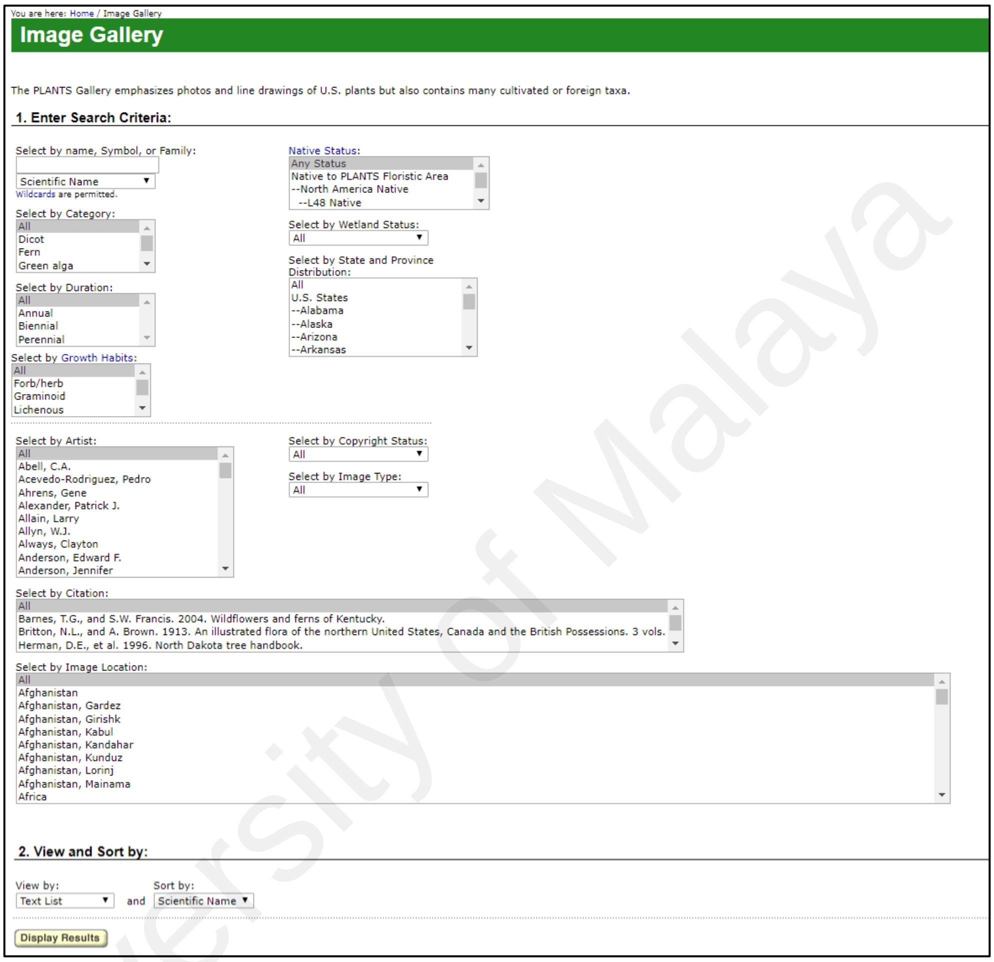
□ St. Pierre and Miquelon

3. Sort by:

● Scientific Name ○ Common Name ○ Symbol

Figure 2.6: PLANTS search tool

Besides that, Image Gallery tool as shown in Figure 2.7 enables users to search for images available in the database. It also shows that there are a number of parameters to filter the result of the images such as native status of the plant species, growth habits, location of the images and by artist.



You are here: Home / Image Gallery

Image Gallery

The PLANTS Gallery emphasizes photos and line drawings of U.S. plants but also contains many cultivated or foreign taxa.

1. Enter Search Criteria:

Select by name, Symbol, or Family:
Scientific Name ▼
Wildcards are permitted.

Select by Category:
All
Dicot
Fern
Green alga

Select by Duration:
All
Annual
Biennial
Perennial

Select by Growth Habits:
All
Forb/herb
Graminoid
Lichenous

Native Status:
Any Status
Native to PLANTS Floristic Area
--North America Native
--L48 Native

Select by Wetland Status:
All ▼

Select by State and Province Distribution:
All
U.S. States
--Alabama
--Alaska
--Arizona
--Arkansas

Select by Artist:
All
Abell, C.A.
Acevedo-Rodriguez, Pedro
Ahrens, Gene
Alexander, Patrick J.
Allain, Larry
Allyn, W.J.
Always, Clayton
Anderson, Edward F.
Anderson, Jennifer

Select by Copyright Status:
All ▼

Select by Image Type:
All ▼

Select by Citation:
All
Barnes, T.G., and S.W. Francis. 2004. Wildflowers and ferns of Kentucky.
Britton, N.L., and A. Brown. 1913. An illustrated flora of the northern United States, Canada and the British Possessions. 3 vols.
Herman, D.E., et al. 1996. North Dakota tree handbook.

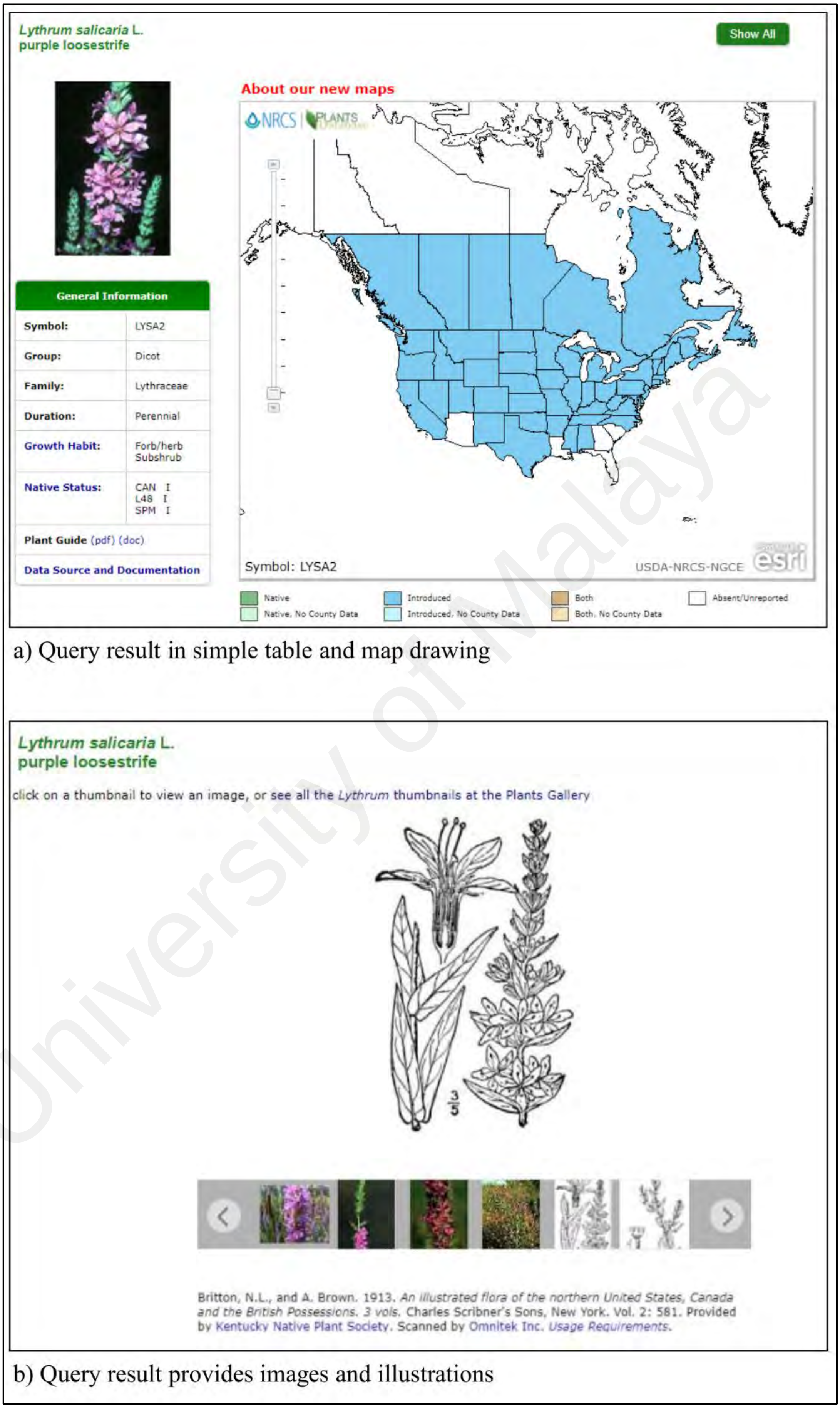
Select by Image Location:
All
Afghanistan
Afghanistan, Gardez
Afghanistan, Girishk
Afghanistan, Kabul
Afghanistan, Kandahar
Afghanistan, Kunduz
Afghanistan, Lorinj
Afghanistan, Mainama
Africa

2. View and Sort by:

View by: Text List ▼ and Sort by: Scientific Name ▼

Figure 2.7: PLANTS Image Gallery tool

Query results are presented in textual and images forms. For example, Figure 2.8 shows the result for query 'Lythrum salicaria'. Figure 2.8(a) shows the general information of *Lythrum salicaria* in table form and locations of the plant species highlighted in map drawing while Figure 2.8(b) displays a number of plant species' images including its illustrations. Furthermore, it also provides links to other external websites where users can find more information of the plant species.



Lythrum salicaria L.
purple loosestrife

Show All

General Information	
Symbol:	LYSA2
Group:	Dicot
Family:	Lythraceae
Duration:	Perennial
Growth Habit:	Forb/herb Subshrub
Native Status:	CAN I L48 I SPM I
Plant Guide (pdf) (doc)	
Data Source and Documentation	

About our new maps

🟩 Native	🟦 Introduced	🟫 Both	□ Absent/Unreported
🟩 Native, No County Data	🟦 Introduced, No County Data	🟫 Both, No County Data	

a) Query result in simple table and map drawing

Lythrum salicaria L.
purple loosestrife

click on a thumbnail to view an image, or see all the *Lythrum* thumbnails at the Plants Gallery

Britton, N.L., and A. Brown. 1913. *An illustrated flora of the northern United States, Canada and the British Possessions. 3 vols.* Charles Scribner's Sons, New York. Vol. 2: 581. Provided by Kentucky Native Plant Society. Scanned by Omnitek Inc. *Usage Requirements*.

b) Query result provides images and illustrations

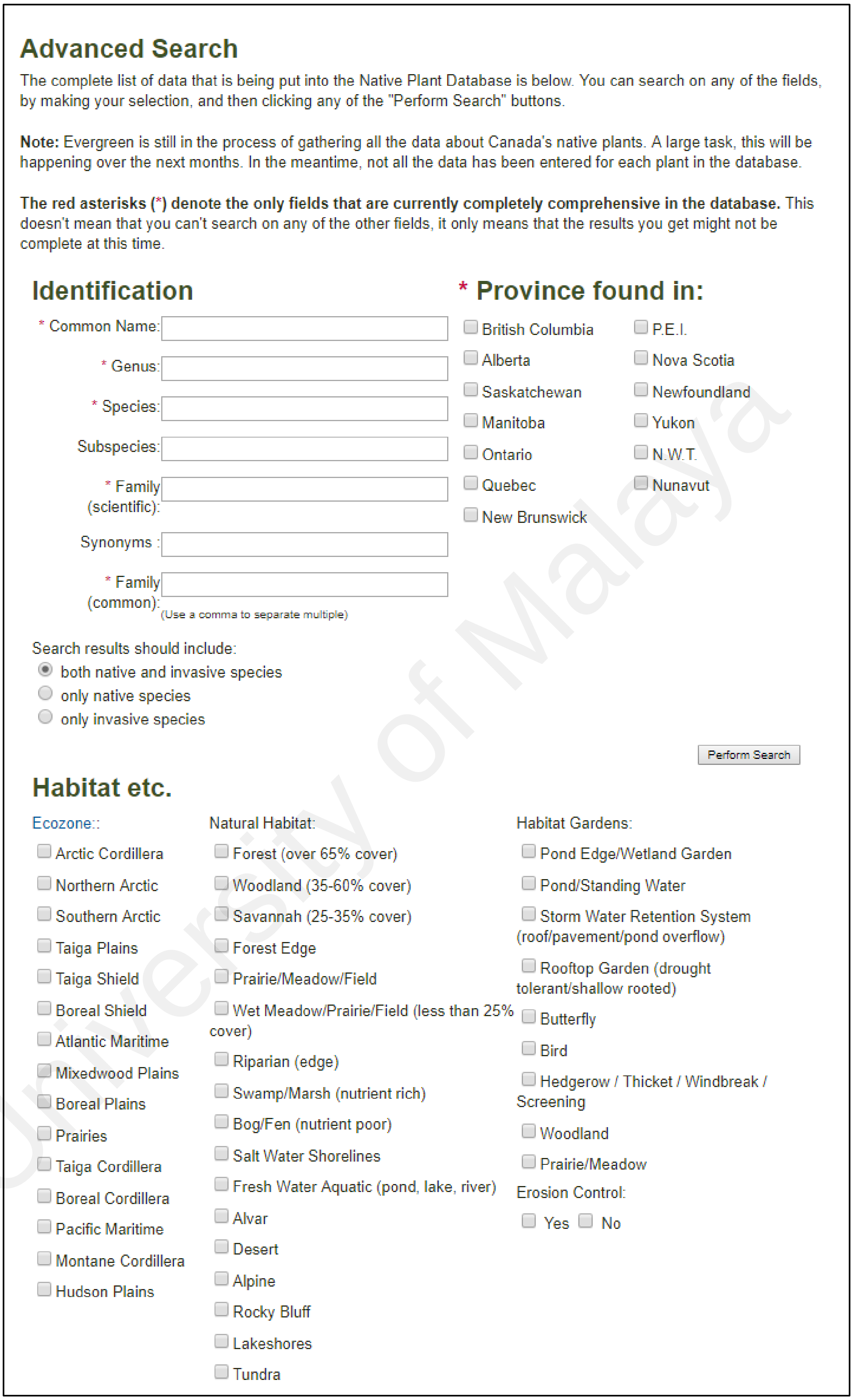
Figure 2.8: Example of PLANTS query result

iv) Native Plant Database (https://nativeplants.evergreen.ca/)

Native Plant Database (Evergreen, 2017) is another example of open access database that consists of textual and image information on native plant species including grasses and vine species located in Canada such as taxonomical, morphological and ecological data. It is useful for gardeners where it includes growing conditions of the plant species with its distribution.

The Advanced Search tool in this database allows users to query for any plant species based on a number of search parameters such as taxonomical data, region area and type of habitat per Figure 2.9. The results of the query are presented to users in tables and images. For instance, Figure 2.10 shows the result for query 'Rosa acicularis' where information of this plant species are arranged in tables and image of the plant species. By clicking on 'View more images of this plant', users are directed to another page with other images of plant species and each image is credited to the photographer as well.

A significant feature in this database is the Recommended Plant Lists tool as shown in Figure 2.11 that lists out a number of plant species suitable for plantation according to the regions in Canada and the type of ecozone. Users can personalize their own plant list where they can add information of a plant species into a list for easy reference in the future. Images provided in this database are dependent on the submission by volunteers.



Advanced Search

The complete list of data that is being put into the Native Plant Database is below. You can search on any of the fields, by making your selection, and then clicking any of the "Perform Search" buttons.

Note: Evergreen is still in the process of gathering all the data about Canada's native plants. A large task, this will be happening over the next months. In the meantime, not all the data has been entered for each plant in the database.

The red asterisks (*) denote the only fields that are currently completely comprehensive in the database. This doesn't mean that you can't search on any of the other fields, it only means that the results you get might not be complete at this time.

Identification

* Common Name: ____

* Genus: ____

* Species: ____

Subspecies: ____

* Family (scientific): ____

Synonyms : ____

* Family (common): ____
(Use a comma to separate multiple)

Search results should include:
● both native and invasive species
○ only native species
○ only invasive species

* Province found in:

□ British Columbia
□ Alberta
□ Saskatchewan
□ Manitoba
□ Ontario
□ Quebec
□ New Brunswick
□ P.E.I.
□ Nova Scotia
□ Newfoundland
□ Yukon
□ N.W.T.
□ Nunavut

Habitat etc.

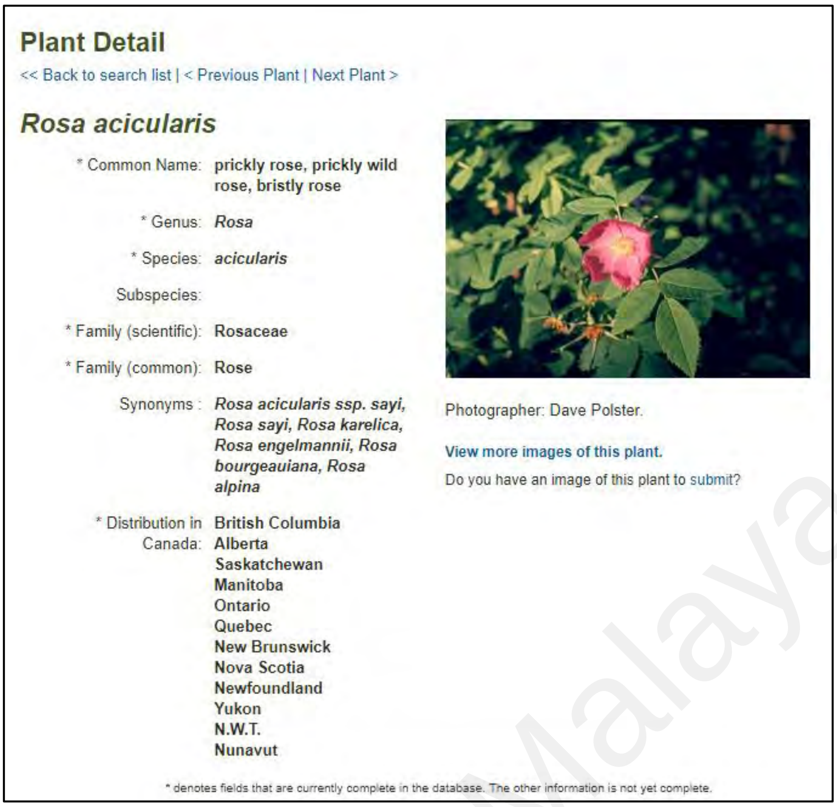
Ecozone::
□ Arctic Cordillera
□ Northern Arctic
□ Southern Arctic
□ Taiga Plains
□ Taiga Shield
□ Boreal Shield
□ Atlantic Maritime
□ Mixedwood Plains
□ Boreal Plains
□ Prairies
□ Taiga Cordillera
□ Boreal Cordillera
□ Pacific Maritime
□ Montane Cordillera
□ Hudson Plains

Natural Habitat:
□ Forest (over 65% cover)
□ Woodland (35-60% cover)
□ Savannah (25-35% cover)
□ Forest Edge
□ Prairie/Meadow/Field
□ Wet Meadow/Prairie/Field (less than 25% cover)
□ Riparian (edge)
□ Swamp/Marsh (nutrient rich)
□ Bog/Fen (nutrient poor)
□ Salt Water Shorelines
□ Fresh Water Aquatic (pond, lake, river)
□ Alvar
□ Desert
□ Alpine
□ Rocky Bluff
□ Lakeshores
□ Tundra

Habitat Gardens:
□ Pond Edge/Wetland Garden
□ Pond/Standing Water
□ Storm Water Retention System (roof/pavement/pond overflow)
□ Rooftop Garden (drought tolerant/shallow rooted)
□ Butterfly
□ Bird
□ Hedgerow / Thicket / Windbreak / Screening
□ Woodland
□ Prairie/Meadow

Erosion Control:
□ Yes □ No

Figure 2.9: Native Plant Database tool - Advanced Search



Plant Detail

<< Back to search list | < Previous Plant | Next Plant >

Rosa acicularis

* Common Name:	prickly rose, prickly wild rose, bristly rose
* Genus:	*Rosa*
* Species:	*acicularis*
Subspecies:	
* Family (scientific):	Rosaceae
* Family (common):	Rose
Synonyms :	*Rosa acicularis ssp. sayi, Rosa sayi, Rosa karelica, Rosa engelmannii, Rosa bourgeauiana, Rosa alpina*
* Distribution in Canada:	British Columbia Alberta Saskatchewan Manitoba Ontario Quebec New Brunswick Nova Scotia Newfoundland Yukon N.W.T. Nunavut

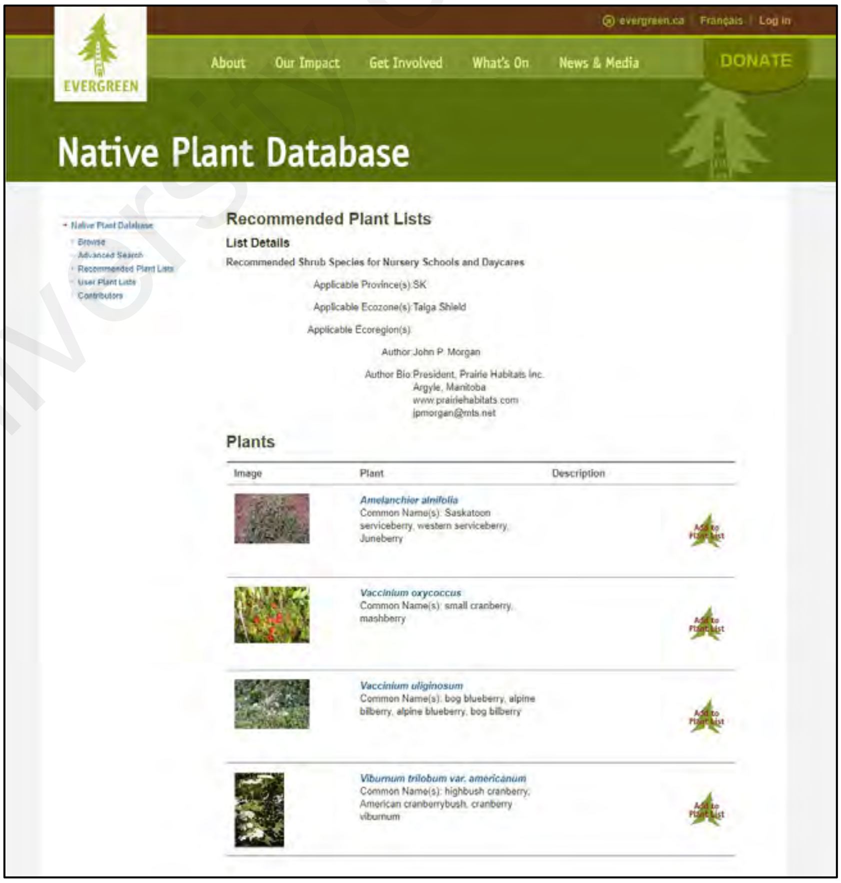
Photographer: Dave Polster.

View more images of this plant.

Do you have an image of this plant to submit?

* denotes fields that are currently complete in the database. The other information is not yet complete.

Figure 2.10: Native Plant Database query result



Recommended Plant Lists

List Details

Recommended Shrub Species for Nursery Schools and Daycares

Applicable Province(s):SK

Applicable Ecozone(s):Taiga Shield

Applicable Ecoregion(s):

Author:John P. Morgan

Author Bio:President, Prairie Habitats Inc.
Argyle, Manitoba
www.prairiehabitats.com
jpmorgan@mts.net

Plants

Image	Plant	Description
	Amelanchier alnifolia Common Name(s): Saskatoon serviceberry, western serviceberry, Juneberry	Add to Plant List
	Vaccinium oxycoccus Common Name(s): small cranberry, mashberry	Add to Plant List
	Vaccinium uliginosum Common Name(s): bog blueberry, alpine bilberry, alpine blueberry, bog bilberry	Add to Plant List
	Viburnum trilobum var. americanum Common Name(s): highbush cranberry, American cranberrybush, cranberry viburnum	Add to Plant List

Figure 2.11: Native Plant Database tool - Recommended Plant Lists

v) maizeGDB (https://www.maizegdb.org/)

maizeGDB (Andorf et al., 2016) is an online database that contains biological information of crop plant *Zea mays* or commonly known as corn. It covers from genomic data such as sequence, gene product, and functional characterization, to literature reference related to the data of corn species (Harper et al., 2016).

There are different tools provided in this database for different purposes such as SNPversity to compare different single nucleotide polymorphisms, BLAST (Boratyn et al., 2013) to find regions of similarity between biological sequences and CornCyc that provides the *Zea mays* metabolic pathways. There is also a genome browser tool as shown in Figure 2.12 that offers access to genomic data of *Zea mays* (Sen et al., 2010) where users can view, interact with, as well as perform textual search for specific regions of the sequences from different data source. The search results are shown in interactive graphical form where it allows users to export any gene regions of interest, download in *.fasta* format, hover the cursor over any coding genes for more information and zoom in or out of a gene region.

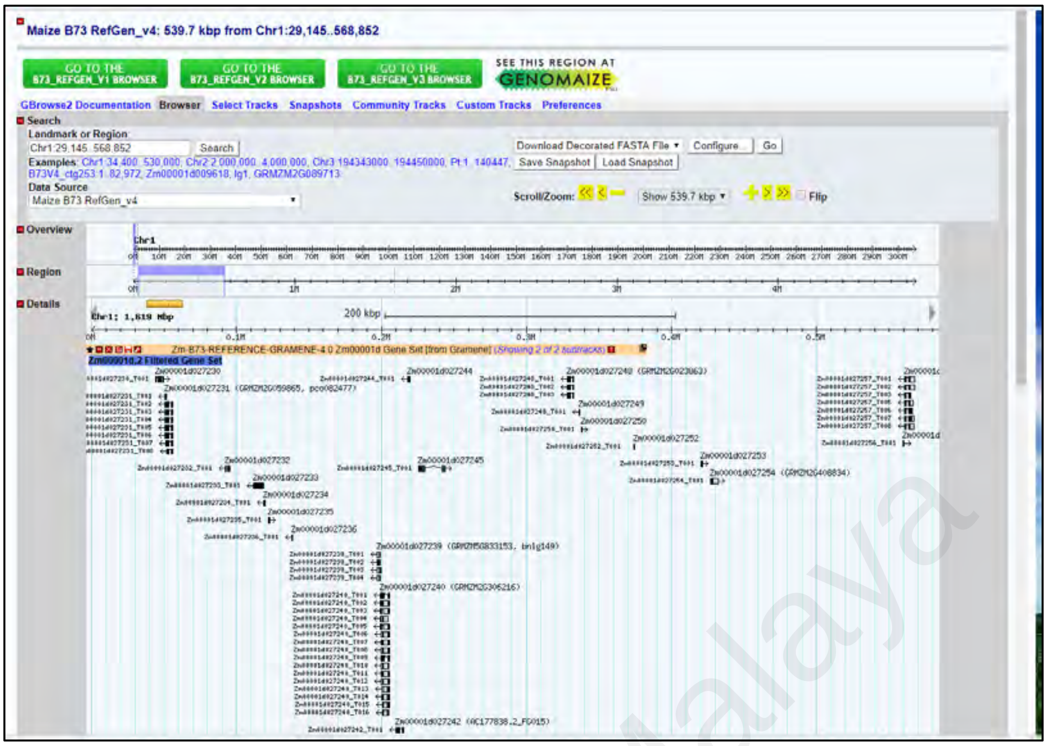

Maize B73 RefGen_v4: 539.7 kbp from Chr1:29,145..568,852
GO TO THE B73_REFGEN_V1 BROWSER
GO TO THE B73_REFGEN_V2 BROWSER
GO TO THE B73_REFGEN_V3 BROWSER
SEE THIS REGION AT GENOMAIZE
GBrowse2 Documentation
Browser
Select Tracks
Snapshots
Community Tracks
Custom Tracks
Preferences
Search
Landmark or Region
Chr1:29,145..568,852
Search
Examples: Chr1:34,400..530,000, Chr2:2,000,000..4,000,000, Chr3:194343000..194450000, Pt:1..140447, B73V4_ctg253:1..82,972, Zm00001d009618, lg1, GRMZM2G089713
Data Source
Maize B73 RefGen_v4
Download Decorated FASTA File
Configure...
Go
Save Snapshot
Load Snapshot
Scroll/Zoom:
Show 539.7 kbp
Flip
Overview
Region
Details
Chr1
Chr1: 1,819 Mbp
200 kbp
Zm-B73-REFERENCE-GRAMENE-4.0 Zm00001d Gene Set [from Gramene] (Showing 2 of 2 subtracks)
Zm00001d.2 Filtered Gene Set

Figure 2.12: MaizeGDB database tool - genome browser

vi) MyBIS (https://www.mybis.gov.my/one/)

MyBIS (MyBIS, 2017) is an open access database that acts as a centre for biodiversity information in Malaysia which covers information of animals, chromista, fungus, and plants. MyBIS has a simple basic search where users can search for any species or references by entering scientific name, common name, taxonomical name, or any keyword as shown in Figure 2.13.

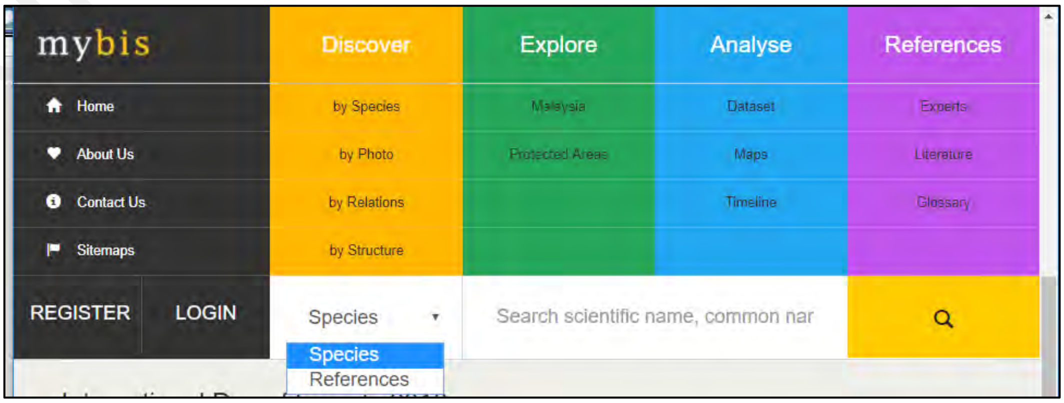
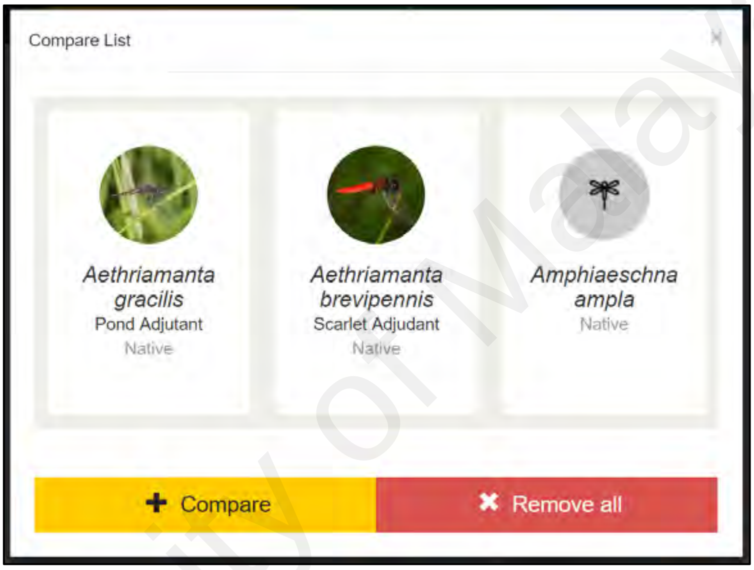


Figure 2.13: MyBIS basic search tool

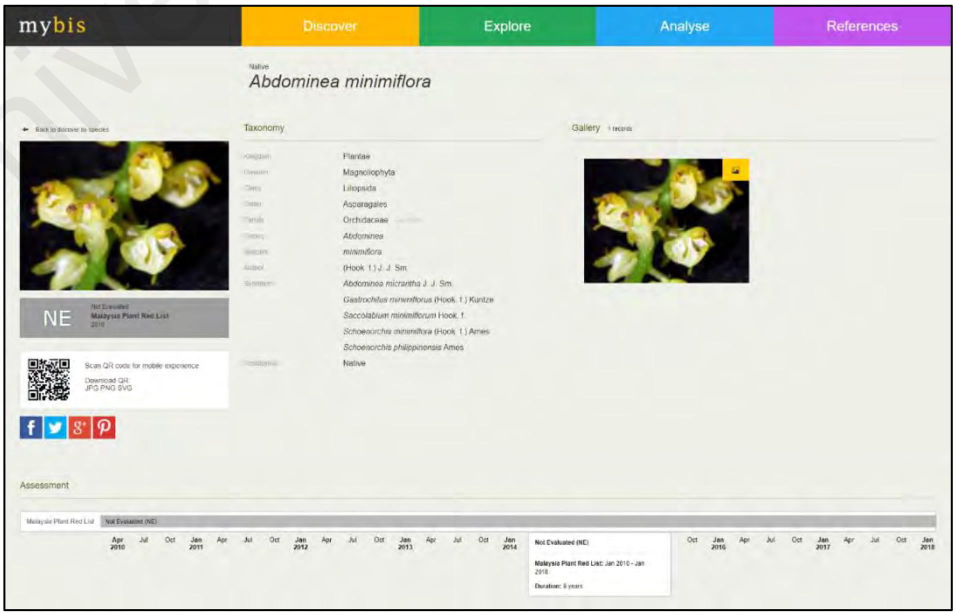
In addition, users can perform comparison between different species by clicking on the plus symbol as shown in Figure 2.14. The search results can be filtered later by parameters such as plant's habit, residential types, and Malaysian states. Figure 2.15 shows how the results are presented, which is in textual forms where taxonomical data of the plant species are displayed in tables and images.



Compare List

Aethriamanta gracilis	*Aethriamanta brevipennis*	*Amphiaeschna ampla*
Pond Adjutant	Scarlet Adjudant	Native
Native	Native	

Figure 2.14: Adding species into the Compare list



Native

Abdominea minimiflora

Taxonomy

Kingdom	Plantae
Division	Magnoliophyta
Class	Liliopsida
Order	Asparagales
Family	Orchidaceae
Genus	*Abdominea*
Species	*minimiflora*
Author	(Hook. f.) J. J. Sm.
Synonym	*Abdominea micrantha* J. J. Sm. *Gastrochilus minimiflorus* (Hook. f.) Kuntze *Saccolabium minimiflorum* Hook. f. *Schoenorchis minimiflora* (Hook. f.) Ames *Schoenorchis philippinensis* Ames
Residential	Native

NE Not Evaluated Malaysia Plant Red List 2010

Scan QR code for mobile experience

Download QR: JPG PNG SVG

Assessment

Malaysia Plant Red List | Not Evaluated (NE)

Not Evaluated (NE)

Malaysia Plant Red List: Jan 2010 - Jan 2018

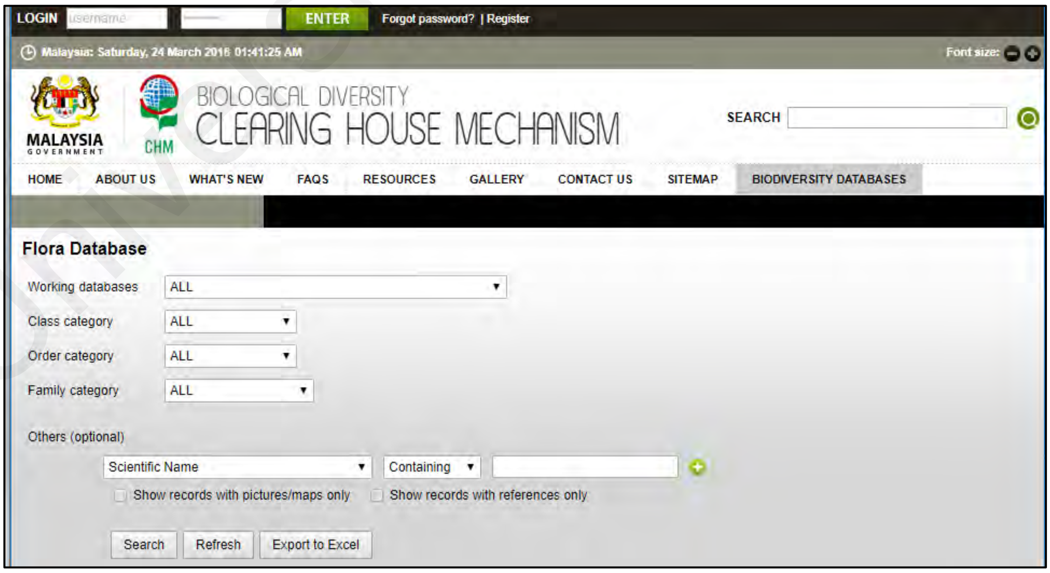
Duration: 8 years

Figure 2.15: MyBIS search result

vii) MyCHM (http://www.chm.frim.gov.my/)

MyCHM (FRIM, 2017) is another open access database developed to facilitate biodiversity information exchange and promote cooperation between different specialities. It consists of taxonomical data of plant, animal and fungi species including its location and habitat in Malaysian state. Images of the organism are also displayed to users when available. This database is branched to three different databases which are Flora, Fauna and Fungi databases.

The Flora database consists of two datasets; Provisional Checklist of the Seaweeds of Malaysia and Provisional Checklist of the Vascular Plants of Malaysia. Figure 2.16 depicts the query tool in MyCHM. To perform query in the flora database, users can query based on the taxonomical classification such as class, scientific name and synonyms; plant's conservation data such as threats and Malaysia's red list category; geological data such as habitat and distribution.



Flora Database

Working databases: ALL

Class category: ALL

Order category: ALL

Family category: ALL

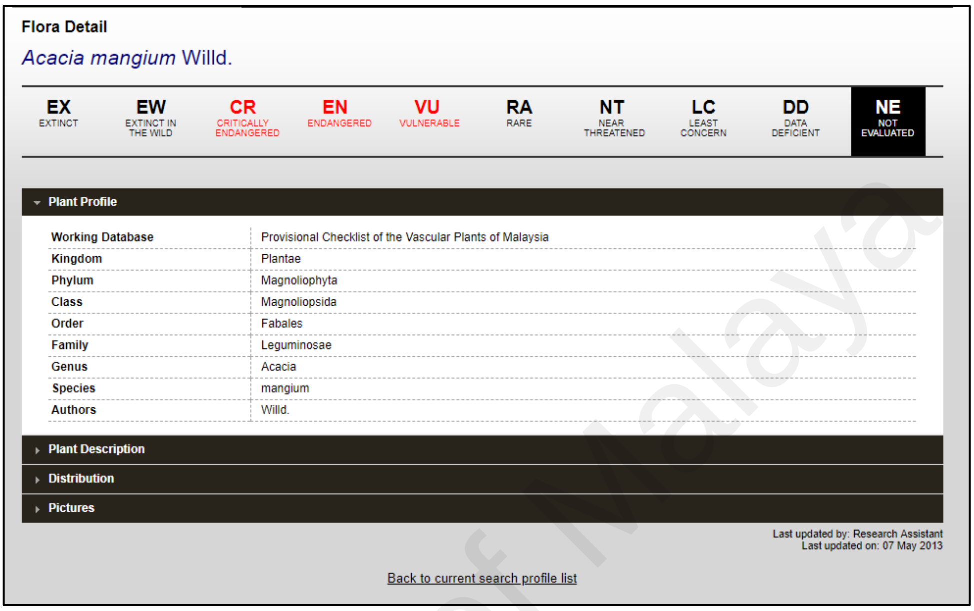
Others (optional)

Scientific Name | Containing

□ Show records with pictures/maps only □ Show records with references only

Figure 2.16: MyCHM query tool

The results of the query are in a form of simple lists of Plant Profile, Plant Description, Distribution, and Pictures as depicted in Figure 2.17. In addition, this database also provides information of biodiversity experts in Malaysia to promote participation of different specialties from different agencies and institutions.



Flora Detail

Acacia mangium Willd.

EX	EW	CR	EN	VU	RA	NT	LC	DD	NE
EXTINCT	EXTINCT IN THE WILD	CRITICALLY ENDANGERED	ENDANGERED	VULNERABLE	RARE	NEAR THREATENED	LEAST CONCERN	DATA DEFICIENT	NOT EVALUATED

Plant Profile

Working Database	Provisional Checklist of the Vascular Plants of Malaysia
Kingdom	Plantae
Phylum	Magnoliophyta
Class	Magnoliopsida
Order	Fabales
Family	Leguminosae
Genus	Acacia
Species	mangium
Authors	Willd.

Plant Description

Distribution

Pictures

Last updated by: Research Assistant
Last updated on: 07 May 2013

Back to current search profile list

Figure 2.17: MyCHM query result

viii) Malaysia Botanical Garden (http://mybotanicalgarden.my/)

Malaysia Botanical Garden (2008) is a plant-based database that consists of Malaysian plants information. Figure 2.18 shows its simple search tool where users can search for any plant species according to the category (i.e. family name). The result page that display the retrieved results is very plain and only consists a few plant data such as genus name, local name, origin and distribution, usage of plant, status, reference, and location as shown in Figure 2.19.

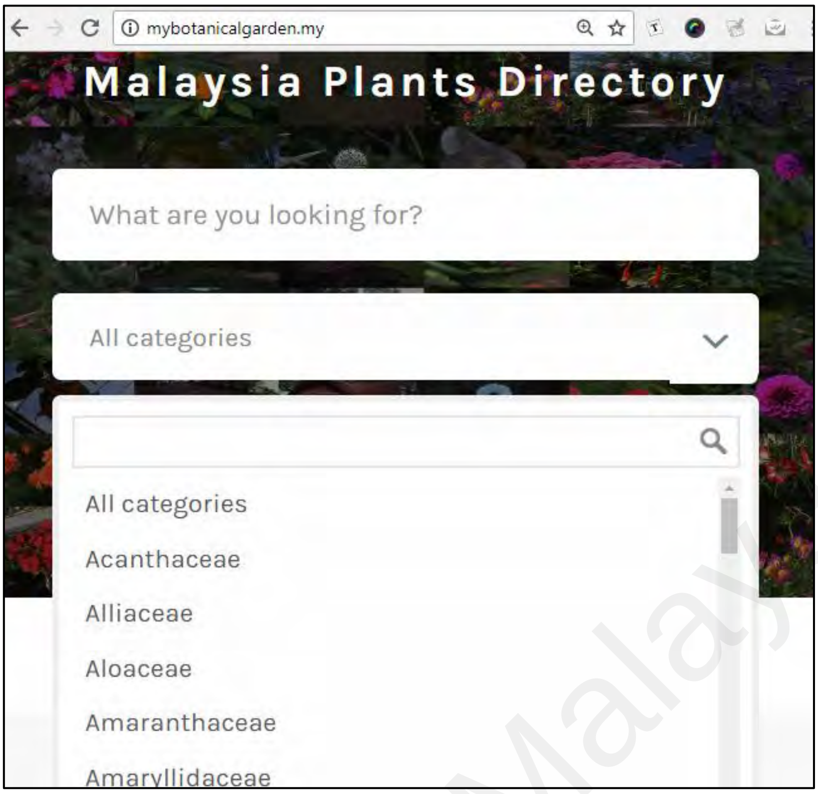
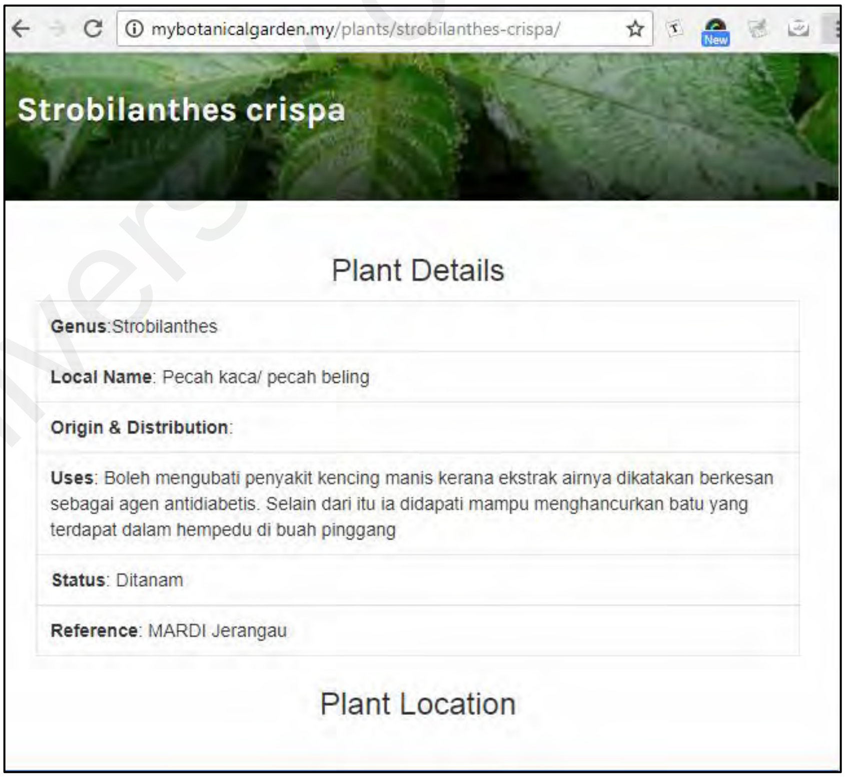


Figure 2.18: Malaysia Botanical Garden search tool



Strobilanthes crispa

Plant Details

Genus:Strobilanthes

Local Name: Pecah kaca/ pecah beling

Origin & Distribution:

Uses: Boleh mengubati penyakit kencing manis kerana ekstrak airnya dikatakan berkesan sebagai agen antidiabetis. Selain dari itu ia didapati mampu menghancurkan batu yang terdapat dalam hempedu di buah pinggang

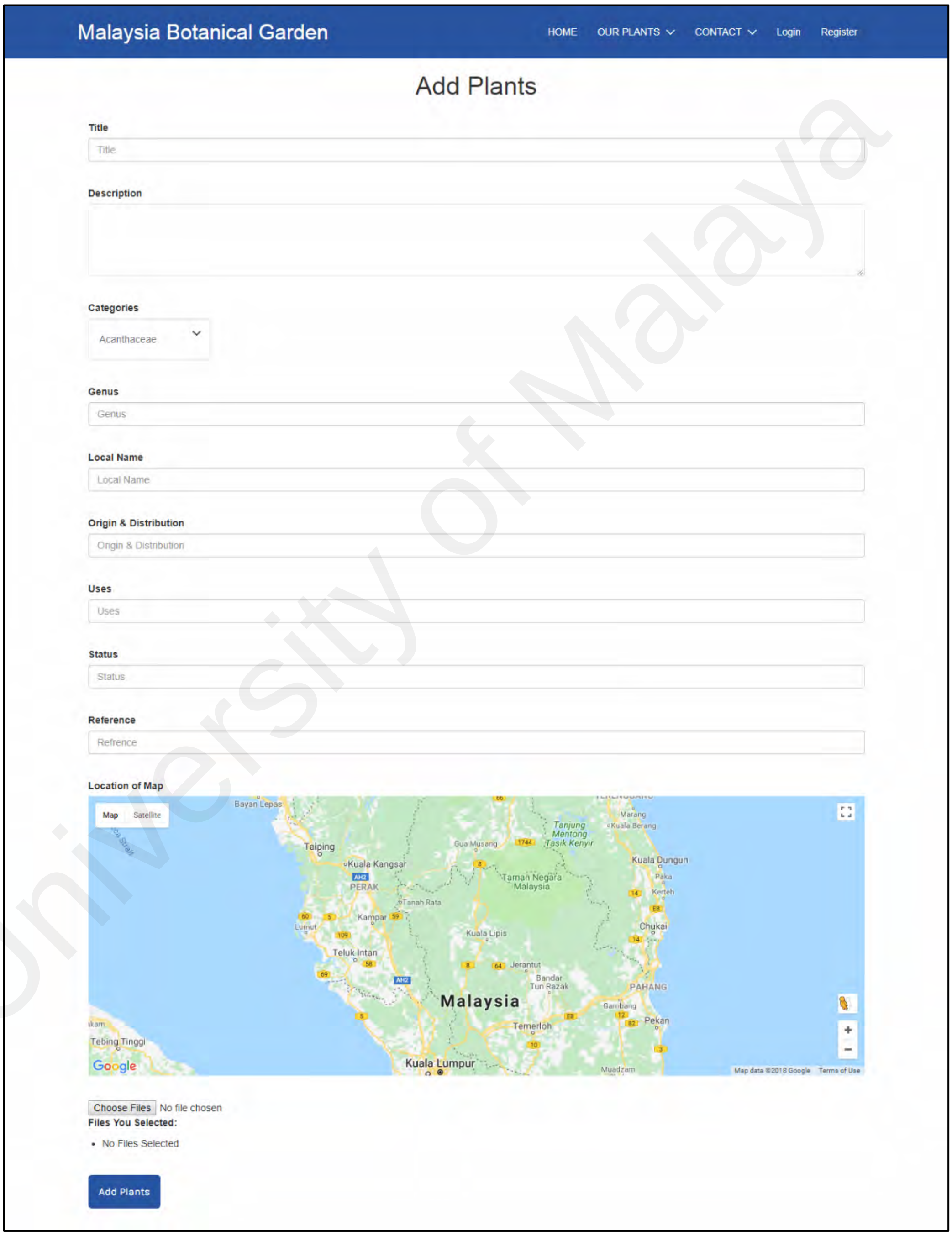
Status: Ditanam

Reference: MARDI Jerangau

Plant Location

Figure 2.19: Malaysia Botanical Garden search result

In addition, this open access database allows users to add their own data into the database whereby users can add a number of attributes regarding the plant such as description, genus name, local name, location of the plant in Malaysia, and images as shown in Figure 2.20.



Add Plants

Title

Title

Description

Categories

Acanthaceae

Genus

Genus

Local Name

Local Name

Origin & Distribution

Origin & Distribution

Uses

Uses

Status

Status

Reference

Refrence

Location of Map

Choose Files No file chosen

Files You Selected:

- No Files Selected

Figure 2.20: Malaysia Botanical Garden - Add Plant tool

2.3.2 Summary of Existing Plant Database Systems

Table 2.1 summaries the features of the current plant database systems previously described in Section 2.3.1. Based on this information, the requirements of the proposed approach are identified and explained in further detail in the Chapter 3.

From this review, plant database exists to store plant data in many domains such as plant collection, plant trait and crop. These databases provide the common features of query and search tool and present the retrieved results to the users. The retrieved results are presented in many forms such as table, list, paragraph and map in textual and image forms.

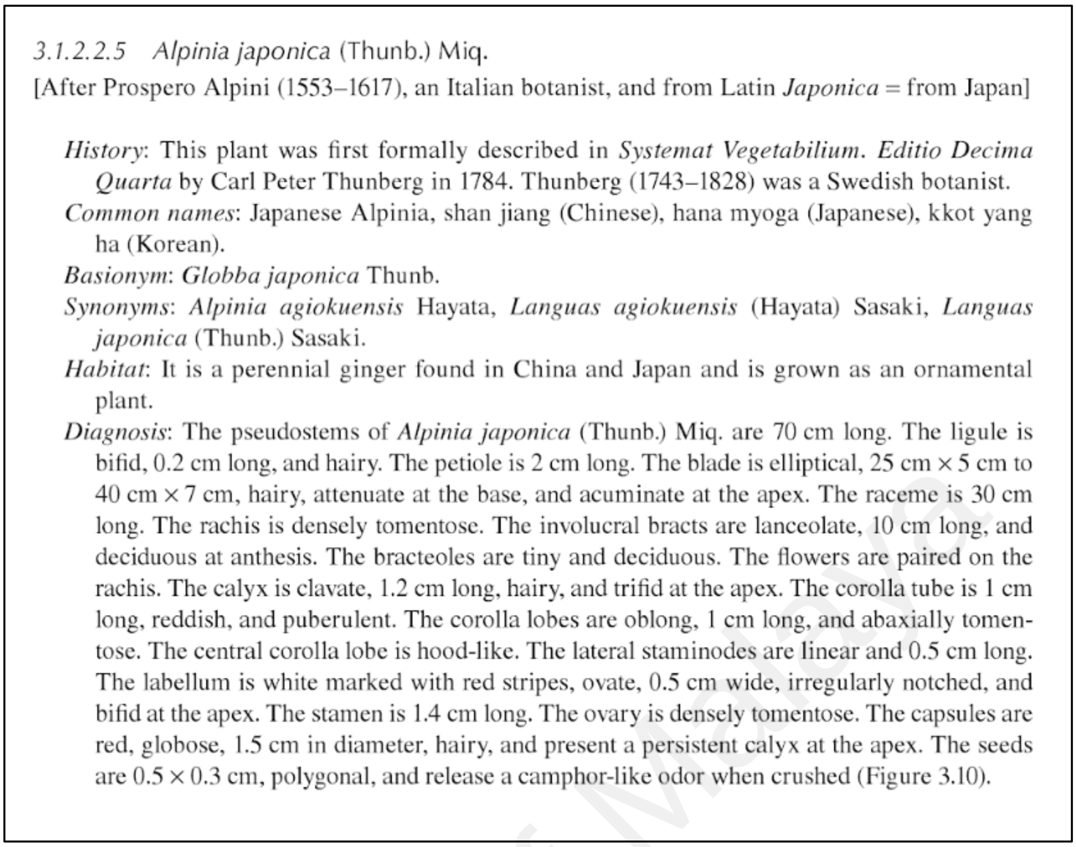
Table 2.1: Comparison of current existing plant database systems

Plant Database	Purpose	Data model	Query method	Form of information	Data presentation	Interactive feature(s)
PlantSearch	• To become the only global database of plant species in botanic gardens • To track threatened plant species which are in botanical collections	Relational	• Options • Keywords	Textual only	• Table	None
TRY Plant Trait	• To improve plant data • As a centre of other plant trait database	Relational	• Options	Textual only	• Table • Images • Map images	None
PLANTS	• To provide information of plant species found in United States territories	Relational	• Options • Keyword	Textual and images	• Table • Images • Map drawing • Illustrations	None
Native Plant	• As one-stop information centre for plant species found in Canada • As easy access for gardeners in Canada	Relational	• Keywords • Options	Textual and images	• Tables • Images	None
maizeGDB	• To provide biological information of plant species *Zea mays*.	Relational	• Options • Keywords	Textual and graphics	• Interactive graphic	• Hover over coding gene • Drag to other part of gene region • Zoom in/out
MyBIS	• As a centre for biodiversity information in Malaysia	Relational	• Keywords	Textual and images	• Table • Images	None
MyCHM	• To facilitate biodiversity information exchange • To promote cooperation between different skills	Relational	• Options • Keywords	Textual and images	• Tables • Images	None
Malaysia Botanical Garden	• To provide information of plant species in Malaysia	Relational	• Keywords	Textual only	• Tables	None

2.4 Data Presentation

There are many ways in presenting data to the users depending on the data type. Biological data are typically in a text form such as sequences, patterns, and biological literature (National Research Council Committee, 2005). Other types of biological data such as high-dimensional omics data (Wang et al., 2014) are produced in gene expression (Alba et al., 2004), geometric information of biological molecules (Dias et al., 2016), and images of natural or man-made biological entities such as digitized plant specimens (JSTOR, 2018) and micro-computed tomography images of animal tissues (Metscher, 2009) are important as well, typically presented in textual forms like paragraphs and tables or graphical forms such as statistical graphs, images, and map drawings.

Plant data, either in physical document or online database is commonly presented in text form as shown in Figure 2.21 and Figure 2.22, respectively. For example, plant species is described with taxonomical data, common names in other languages (Hussain et al., 2015; Khare, 2007) and characteristics of the plant species such as growth habit, parts of plants (e.g. leaves, flower, fruit) in paragraphs (Barrett & Tay, 2016; Hobbs & Foster, 2002). In addition, details of the plant species are also presented in tables (Normah et al., 2013; Said et al., 2001) especially in online databases where data such as taxonomical data (Bisby et al., 2010; Evergreen, 2017; FRIM, 2017) and geological data (Evergreen, 2017; Garden, 2018; USDA, 2017) are organized into tables and occasionally linked to another webpage.



3.1.2.2.5 *Alpinia japonica* (Thunb.) Miq.

[After Prospero Alpini (1553–1617), an Italian botanist, and from Latin *Japonica* = from Japan]

History: This plant was first formally described in *Systemat Vegetabilium. Editio Decima Quarta* by Carl Peter Thunberg in 1784. Thunberg (1743–1828) was a Swedish botanist.

Common names: Japanese Alpinia, shan jiang (Chinese), hana myoga (Japanese), kkot yang ha (Korean).

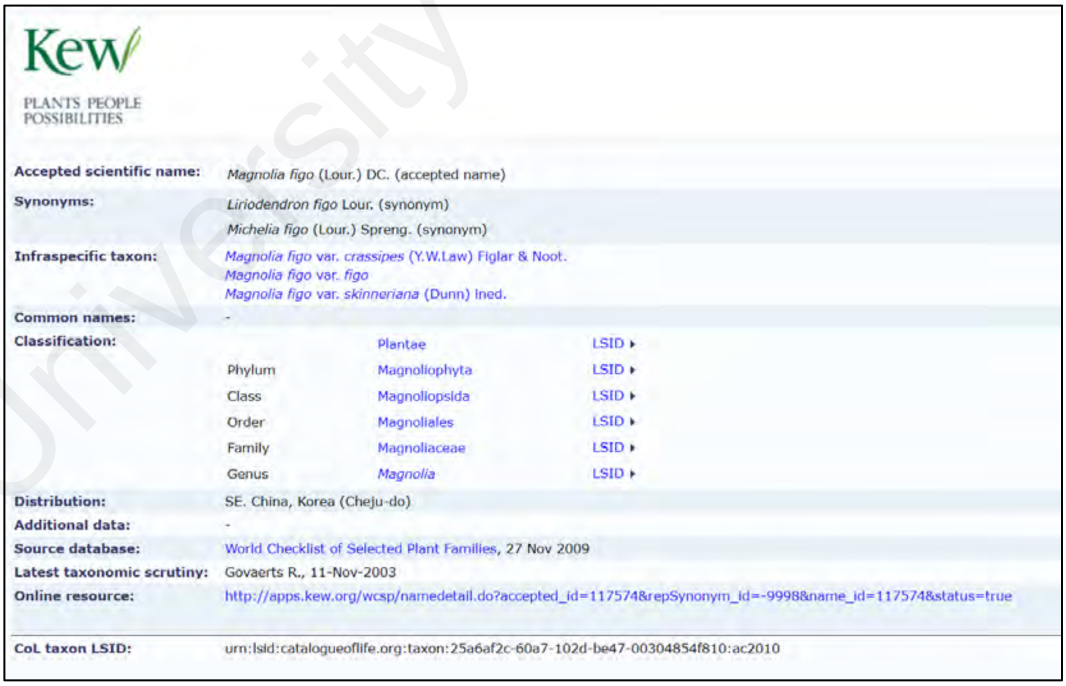
Basionym: *Globba japonica* Thunb.

Synonyms: *Alpinia agiokuensis* Hayata, *Languas agiokuensis* (Hayata) Sasaki, *Languas japonica* (Thunb.) Sasaki.

Habitat: It is a perennial ginger found in China and Japan and is grown as an ornamental plant.

Diagnosis: The pseudostems of *Alpinia japonica* (Thunb.) Miq. are 70 cm long. The ligule is bifid, 0.2 cm long, and hairy. The petiole is 2 cm long. The blade is elliptical, 25 cm × 5 cm to 40 cm × 7 cm, hairy, attenuate at the base, and acuminate at the apex. The raceme is 30 cm long. The rachis is densely tomentose. The involucral bracts are lanceolate, 10 cm long, and deciduous at anthesis. The bracteoles are tiny and deciduous. The flowers are paired on the rachis. The calyx is clavate, 1.2 cm long, hairy, and trifid at the apex. The corolla tube is 1 cm long, reddish, and puberulent. The corolla lobes are oblong, 1 cm long, and abaxially tomentose. The central corolla lobe is hood-like. The lateral staminodes are linear and 0.5 cm long. The labellum is white marked with red stripes, ovate, 0.5 cm wide, irregularly notched, and bifid at the apex. The stamen is 1.4 cm long. The ovary is densely tomentose. The capsules are red, globose, 1.5 cm in diameter, hairy, and present a persistent calyx at the apex. The seeds are 0.5 × 0.3 cm, polygonal, and release a camphor-like odor when crushed (Figure 3.10).

Figure 2.21: Example of plant data in text



Accepted scientific name:	*Magnolia figo* (Lour.) DC. (accepted name)		
Synonyms:	*Liriodendron figo* Lour. (synonym) *Michelia figo* (Lour.) Spreng. (synonym)		
Infraspecific taxon:	*Magnolia figo* var. *crassipes* (Y.W.Law) Figlar & Noot. *Magnolia figo* var. *figo* *Magnolia figo* var. *skinneriana* (Dunn) Ined.		
Common names:	-		
Classification:		Plantae	LSID ▸
	Phylum	Magnoliophyta	LSID ▸
	Class	Magnoliopsida	LSID ▸
	Order	Magnoliales	LSID ▸
	Family	Magnoliaceae	LSID ▸
	Genus	*Magnolia*	LSID ▸
Distribution:	SE. China, Korea (Cheju-do)		
Additional data:	-		
Source database:	World Checklist of Selected Plant Families, 27 Nov 2009		
Latest taxonomic scrutiny:	Govaerts R., 11-Nov-2003		
Online resource:	http://apps.kew.org/wcsp/namedetail.do?accepted_id=117574&repSynonym_id=-9998&name_id=117574&status=true		
CoL taxon LSID:	urn:lsid:catalogueoflife.org:taxon:25a6af2c-60a7-102d-be47-00304854f810:ac2010		

Figure 2.22: Example of plant data in an online database retrieved from Catalogue of Life (2018**)**

Besides that, graphical form is also used in describing plant species. For instance, statistical graphs like bar charts and scatter plots are used to describe distribution of genes (Lee et al., 2009; Petryszak et al., 2016) and results of genomic experiments (Dash et al., 2012). Figure 2.23 illustrates an example of plant data described in a form of histogram. Other than that, images of each part or whole plant species (JSTOR, 2018; MyBIS, 2017) are commonly displayed to users. Meanwhile, drawings of map are also used to highlight the distribution of specific plant species (Kattge et al., 2011; USDA, 2017).

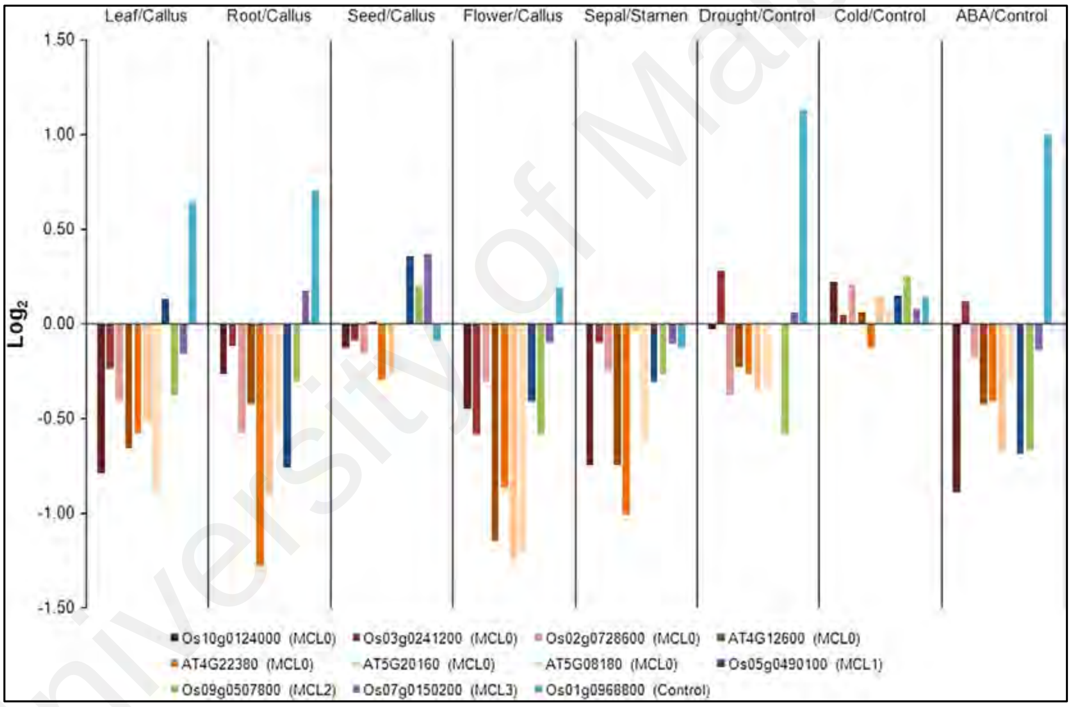


Figure 2.23: Example of plant data in a form of histogram

Briefly, plant data are heterogenous, containing complex pictures and many terminologies to describe the plants in textual and image forms. Textual data are commonly presented in paragraph, table and list for easily retrieving, viewing and gaining the information. On the other hand, the data in graphical form can deduce the knowledge from the visualised data. The following Section 2.5 provides more details on the main topic in this research; data visualisation.

2.5 Data Visualisation

Scientific field yields innumerable amount of data from previous findings, publications and researches. Moreover, newly founded research in methodologies and advancement has led to a steep increase in the number of new scientific data and scientists look into any patterns, trends, or relationships in data. As the plain view of texts or tables is insufficient in giving a clear explanation of the data, thus visualisation aids in presenting the data in various forms that suit the data flow. Data visualisation is a comprehensive field of crossover between many fields such as mathematics, computer science, cognitive and perception science, and engineering fields (Telea, 2014). Besides that, data visualisation presents data using a visual or artistic approach rather than the conventional reporting method (Yuk & Diamond, 2014). It plays an important part in many fields such as business (Tegarden, 1999), geography (Groenendyk, 2013), and biology (Chen et al., 2014; Jensen & Papin, 2014; Sedova et al., 2015).

To portray the concept of data visualisation, Figure 2.24 illustrates the main elements in data visualisation which are the messenger, the receiver, and the message (Kirk et al., 2016). The messenger conveys message in a form of data, ideas or results to receiver who is the user of the visualisation. The message in the middle is the data visualisation which is the form of communication between the messenger and receiver. Through a proper way to encode the message using data visualisation, the receiver can decode the message by interpreting it into a meaningful insight and knowledge.

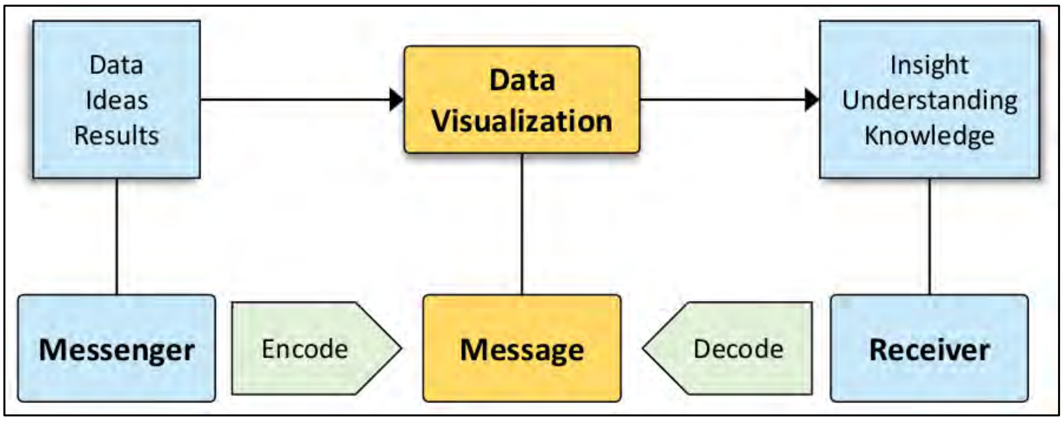


Figure 2.24: Main elements in data visualisation. Image reproduced with permission from Kirk et al. (2016)

In biological field, visualizing biological data helps researchers to view the data in a different angle to provide cognitive support and analysis (Tory & Moller, 2004). It is easier for the brain to understand an image rather than words of numbers (Cukier, 2010) whereby the numerical data may be translated using dots, lines or bars to help in presenting qualitative information (Few, 2004). Not only that, data visualisation can summarize a large amount of data into effective graphics (Ware, 2012).

Data visualisation is important in data analysis process where data are arranged and structured for clear understanding of the implications of the data as by visualizing the data, it helps in translating the data to a suitable type of visualisation such as multi-dimensional graphics (e.g. bar charts, histogram, line graphs), stacked graphics (e.g. treemap, dimensional stacking) (Keim, 2002), and network graphics (e.g. undirected, weighted, lattice) which then emphasizes on the key points contained in the data. Users can perform analytical tasks such as reconstruction of biomolecular modeling (Lučić et al., 2005), investigating biological pathways (Murray et al., 2017) and studying properties of individual species (Conesa & Mortazavi, 2014). From data analysis, users can infer new knowledge. For instance, in systems biology, the combination of data obtained from experiments, data visualisation, and statistical-model approaches allow researchers to gain new knowledge about the causal influences in cancer signalling networks (Hill et al., 2016; Iglesias-Martinez et al., 2016) or to identify biologically relevant pathways based on proteomic dataset (Mukherjee & Speed, 2008; Zuo et al., 2015).

The following Section 2.5.1 describes the six examples of data visualisation tools in biology and the main features of these tools are summarised in Table 2.2 as presented in Section 2.5.2. Based on the review done, the main components in the data visualisation development are also identified as briefly explained in Section 2.5.3.

2.5.1 Biological-based Data Visualisation

The rapid growth in both volume and variety of biological data causes an increasing challenge for researchers to fully understand these data. One of the main solutions is using data visualisation through the increasing number of approaches and systems for biological data visualisation (Czauderna & Schreiber, 2017; King et al., 2015; Kleiberg et al., 2001) and has increasingly become a fundamental aspect in biology field (O'Donoghue et al., 2010). A few examples of biological based data visualisation are explained as the following:-

i) Circos

Circos (Krzywinski et al., 2009) is a visualisation tool to assist users in the analysis of genomic data comparison. This tool can be used on Unix, Mac OS and MS Windows operating systems, and also available online at *http://mkweb.bcgsc.ca/tableviewer/*. It works well for genomic data and common data as long as a relationship between the two elements exists. Circos visualizes the data in circular layout which is ideal for exploring connections between data.

It also supports various plot types such as scatter plot, histogram, and heat map. Users can define their own style to individual elements such as colours and position.

Besides that, this tool is unique as it supports global and local zooming feature which means users can enlarge and/or compress to the whole regions or only to individual region as shown in Figure 2.25. Regions A and B are zoomed to '10x' magnification while regions 'J' and 'K' are zoomed to '20x' magnification. With this feature, regions of interest can be shown in detailed while keeping the rest of the data in view.

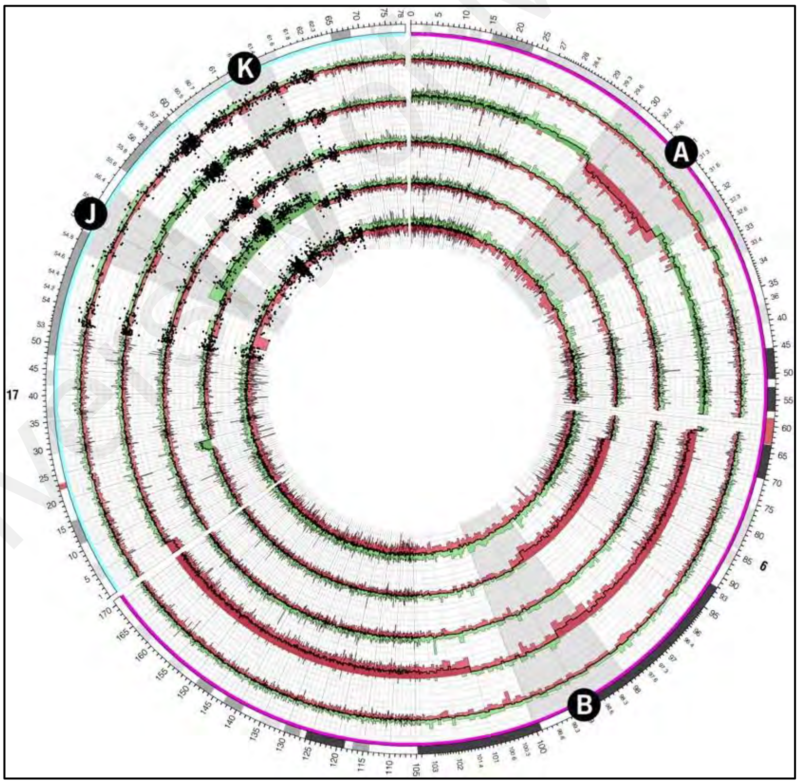
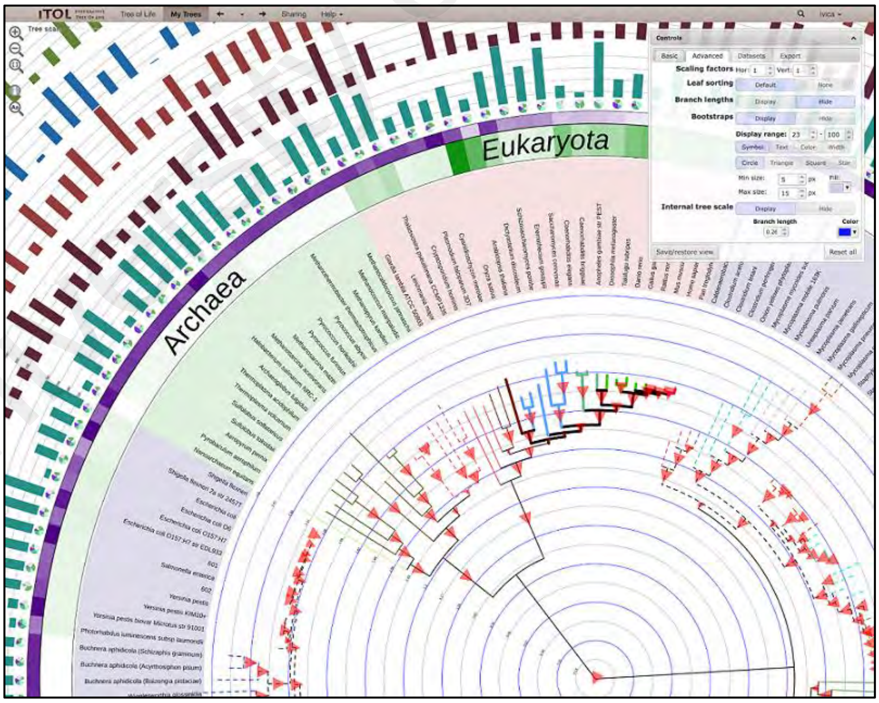


Figure 2.25: Example of data visualisation using Circos. Image reproduced with permission from Krzywinski et al. (2009)

ii) Interactive Tree of Life (iTOL)

This is a web-based tool to display, annotate, and manage phylogenetic tree of genomic sequences. Figure 2.26 shows an example of iTOL's user interface where phylogenetic tree of Tree of Life (Ciccarelli et al., 2006) is annotated with various datasets at one time. Other than basic functions such as various tree display formats, delete or move nodes, and re-root tree, there are other additional functions which allows users to customize their tree displays in different approaches. For instance, users can perform pruning; a process of creating a smaller tree by selecting one or several branches from original tree. Users can also annotate external data on a tree and customize styles for branch, label, and tree size. The phylogenetic tree created can be exported to different file formats such as PDF, PNG, and EPS file formats which give flexibility for users.



Scaling factors Hor: 1 Vert: 1
Leaf sorting: Default / None
Branch lengths: Display / Hide
Bootstraps: Display / Hide
Display range: 23 - 100
Symbol / Text / Color / Width
Circle / Triangle / Square / Star
Min size: 5 px
Max size: 15 px
Fill:
Internal tree scale: Display / Hide
Branch length 0.26
Color

Figure 2.26: Example of phylogenetic tree generated by iTOL tool. Image reproduced with permission from Ciccarelli et al. (2006)

iii) BioVis Explorer (http://biovis.lnu.se/)

BioVis Explorer (Kerren et al., 2017) is an interactive web-based visualisation tool which provides reviews of a collection of visualisation tools or methods used in presenting biological data. The visualisation is based on multidimensional scaling where distance between the visualisation tools and information such as publication year, authors, and categories are calculated using mathematical formula. For instance, distance between a pair of techniques is computed based on the difference in the assigned categories using Jaccard index (Kerren et al., 2017). Figure 2.27 shows the visualisation in BioVis Explorer, which supports user interactions including zooming and panning. Users can obtain detailed information by clicking on a thumbnail, view comparison measurement between related visualisation tools by hovering on a thumbnail, as well as filtering tool that allows users to filter based on certain parameters such as keyword, types of data, and tasks used in the visualisation tool.

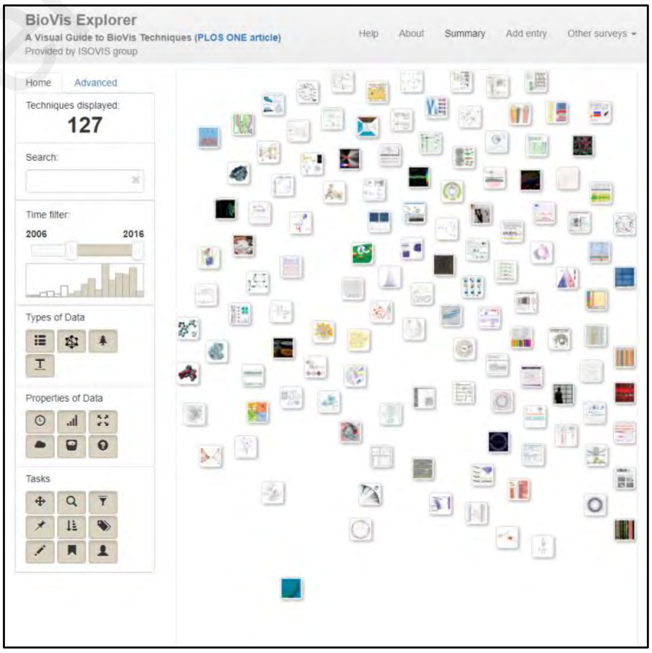


Figure 2.27: Visualisation in BioVis Explorer. Image reproduced with permission from Kerren et al. (2017)

iv) Dendroscope

Dendroscope (Huson et al., 2007) is an interactive phylogenetic tree viewer that is available as a software written in Java programming language which views variety analyses of molecular data sets. Dendroscope allows the phylogenetic tree to be displayed in a number of views such phylogram, cladogram, or unrooted tree. Figure 2.28 shows an example of phylogenetic tree generated using Dendroscope where taxon Homo sapiens is highlighted in red box. Other features such as rotate, magnify, search tool, and export diagram to another file format are also available. In addition, this software is available in different platforms such as Linux, MacOS X, and Windows XP. These exceptional features have made this software one of the popular tree viewers.

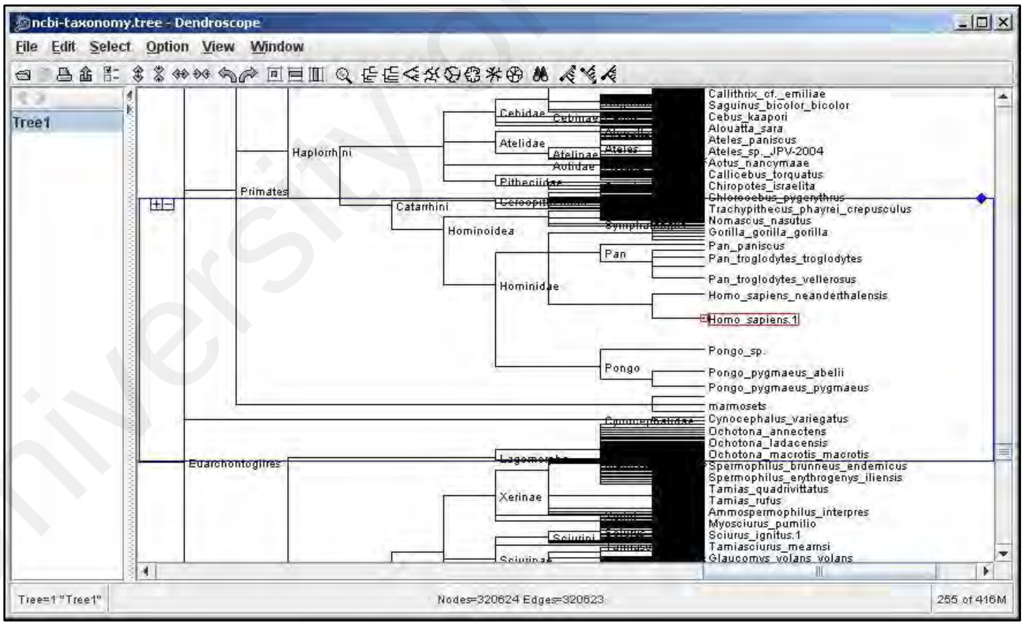
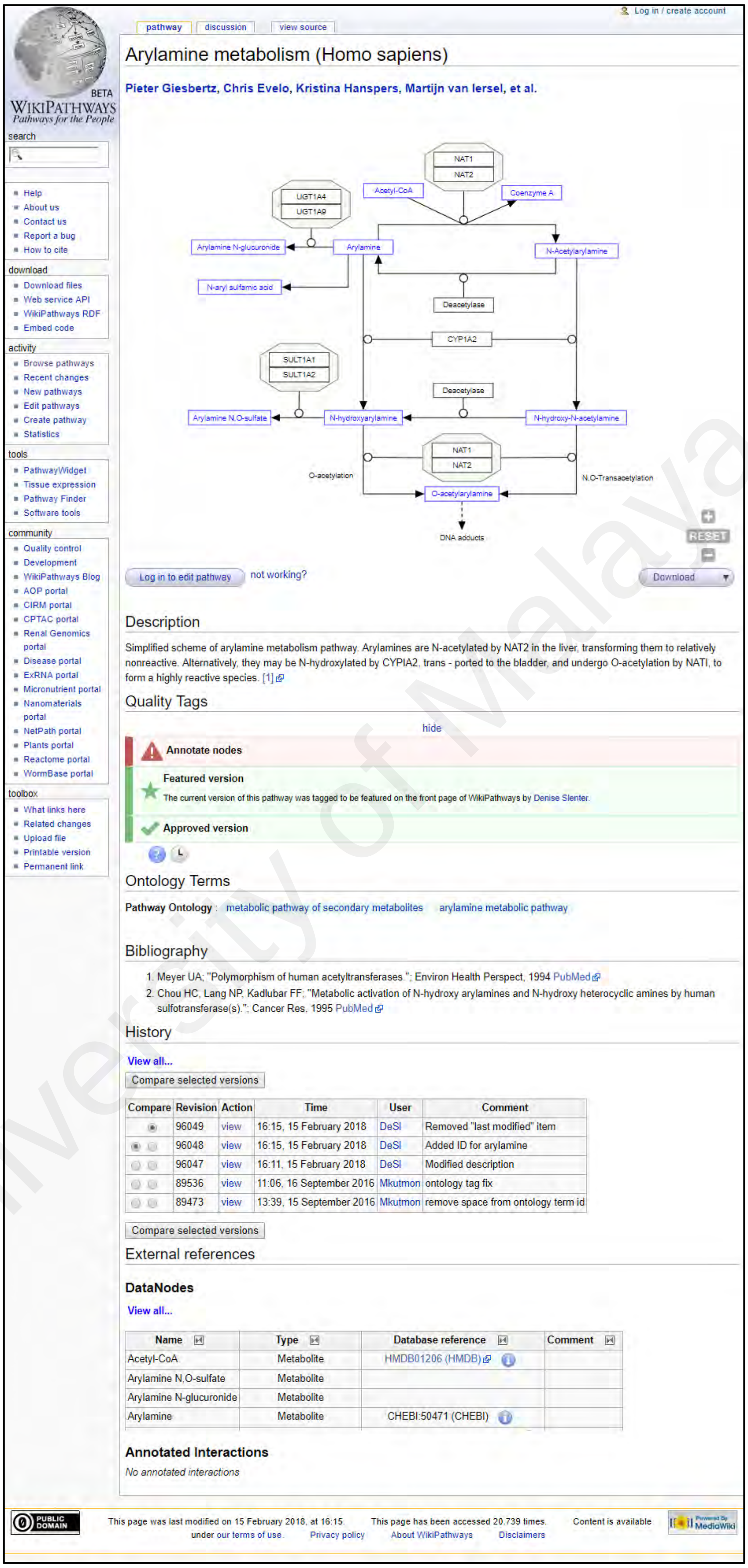


Figure 2.28: Interactive phylogenetic tree viewer in software Dendroscope. Image reproduced with permission from Huson et al. (2007)

v) WikiPathways

WikiPathways (Kelder et al., 2009; Slenter et al., 2018) is designed to aid in the contribution and maintenance of biological pathways' information. It is an open platform for the curation of biological pathways while collaborating with other researchers. WikiPathways owns a custom graphical editing tool and includes other databases that cover gene, proteins, as well as small-molecule systems. Each biological pathway in this web-based system has a dedicated wiki page that displays detailed information such as description, references, export options, and lists of component gene and protein (see Figure 2.29).



Arylamine metabolism (Homo sapiens)

Pieter Giesbertz, Chris Evelo, Kristina Hanspers, Martijn van Iersel, et al.

Description

Simplified scheme of arylamine metabolism pathway. Arylamines are N-acetylated by NAT2 in the liver, transforming them to relatively nonreactive. Alternatively, they may be N-hydroxylated by CYPIA2, trans - ported to the bladder, and undergo O-acetylation by NATI, to form a highly reactive species. [1]

Quality Tags

hide

Annotate nodes

Featured version

The current version of this pathway was tagged to be featured on the front page of WikiPathways by Denise Slenter.

Approved version

Ontology Terms

Pathway Ontology : metabolic pathway of secondary metabolites arylamine metabolic pathway

Bibliography

1. Meyer UA; "Polymorphism of human acetyltransferases."; Environ Health Perspect, 1994 PubMed
2. Chou HC, Lang NP, Kadlubar FF; "Metabolic activation of N-hydroxy arylamines and N-hydroxy heterocyclic amines by human sulfotransferase(s)."; Cancer Res, 1995 PubMed

History

View all...

Compare selected versions

Compare	Revision	Action	Time	User	Comment
●	96049	view	16:15, 15 February 2018	DeSl	Removed "last modified" item
● ○	96048	view	16:15, 15 February 2018	DeSl	Added ID for arylamine
○ ○	96047	view	16:11, 15 February 2018	DeSl	Modified description
○ ○	89536	view	11:06, 16 September 2016	Mkutmon	ontology tag fix
○ ○	89473	view	13:39, 15 September 2016	Mkutmon	remove space from ontology term id

Compare selected versions

External references

DataNodes

View all...

Name	Type	Database reference	Comment
Acetyl-CoA	Metabolite	HMDB01206 (HMDB)	
Arylamine N,O-sulfate	Metabolite		
Arylamine N-glucuronide	Metabolite		
Arylamine	Metabolite	CHEBI:50471 (CHEBI)	

Annotated Interactions

No annotated interactions

This page was last modified on 15 February 2018, at 16:15. This page has been accessed 20,739 times. Content is available under our terms of use.

Figure 2.29: Visualisation of biological pathways in WikiPathways. Image reproduced with permission from Giesbertz et al. (2018)

vi) UGENE

UGENE (Okonechnikov et al., 2012) is another software that works on Windows, MacOS X, and Linux. This open source software helps researchers to analyse different biological genetics data such as sequences, alignments, and phylogenetic trees. GUI provided in this software as shown in Figure 2.30 which consists of project viewer and sequence viewer enable users with no skill in computer programming to use tools in the software easily. Another key feature of this software is that it does not only visualize for a specific type of biological data but also a number of different visualisation formats available such as phylogenetic tree, chromatogram, and multidimensional structure.

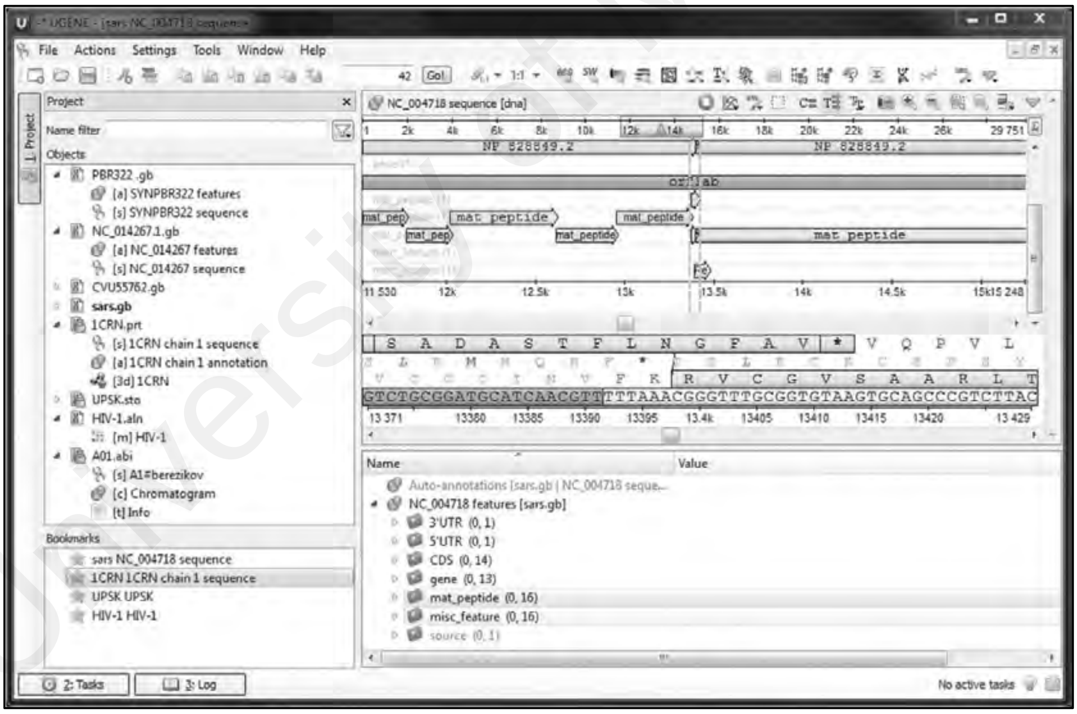


Figure 2.30: GUI of UGENE software. Image reproduced with permission from Okonechnikov et al. (2012)

2.5.2 Summary of Biological-based Data Visualisation

Table 2.2 is a summary of the biological-based data visualisation as discussed in the Section 2.5.1. It can be seen that there are many types of data visualisation in presenting biological data. Each type is suitable for different input data. Besides that, an interactive element is the key of data visualisation because it provides a mechanism for the users to interact and manipulate the data. The data flow is also easy to understand with the support of colouration scheme and text labelling.

Table 2.2: Summary of biological-based data visualisation

Name	Purpose	Type of system	Input	Type of visualisation	Main feature(s)
Circos	To visualize comparisons of genomic and general data	• Standalone software • Web-based system	Textual	• Circular layout	• Support various plot types in circular layout • User-defined styles • Support global and local zooming
iTOL	To provide viewer for phylogenetic tree of genomic sequences	• Web-based system	Molecular sequences	• Phylogram • Cladogram • Radial	• Annotate external datasets • Pruning • User-defined styles • Export options
BioVis Explorer	To provide review of a visualisation approaches used in presenting biological data	• Web-based system	Textual query	• Network graph	• Support basic user interactions such as zoom and pan • Filtering tool
Dendroscope	To visualize analyses of molecular data sets	• Standalone software	Molecular sequences	• Phylogram • Cladogram	• Different visualisation options • Support basic user interactions such as rotate and magnify • Support multi platforms
WikiPathways	To provide information of curated biological pathways	• Web-based system	Textual query	• Mind map	• Custom graphical editing tool • Dedicated wiki page for each biological pathway
UGENE	To provide analysis of biological genetics data	• Software	Molecular sequences	• Phylogram • Cladogram • Chromatogram • Chord diagram	• Support multi platforms • Different visualisation options • Project viewer • Sequence viewer

2.5.3 Main Components of Data Visualisation Development

Data visualisation is visual communications that make complicated data understandable and visually engaging so that users can easily observe the information and make inferences. From the review done in Sections 2.5.1 and 2.5.2, there are three main components in data visualisation development that have been identified. One of the main components is the importance of understanding the logical flow of the data beforehand. Therefore, the type of data modelling used is essential in visualizing data as it determines the flow of data retrieval to generate the visualisation. This component is discussed further in Section 2.6.

The next main component is type of visualisation used to present the data. It plays a major role in presenting data to users as different visualisation emphasize on different types of data. For example, line chart is used to show trends and patterns such as to show how data changes over a period of time and the number of algae found over a number of data collection in different treatments (Bhardwaj, 2017). Meanwhile, bar chart emphasizes on the differences between data such as comparing number of trees in residential area (Nitoslawski & Duinker, 2016). Other types namely multidimensional graphics such as histogram, and treemap; hierarchical graphics such as dendrogram, radial tree, and hyperbolic tree; and network graphics such as network graph, alluvial diagram, and circular hierarchy graph (Zoss, 2018) are also used to present the data. The types of data visualisation are discussed further in Section 2.7.

The last component is the tools used in creating the data visualisation. These tools are capable to process different types of data and present the data to users in graphical form. The tools are available in open-source and proprietary software or visual library that is made up of a set of programming languages which is easy for users to integrate them into their systems. In biological database, data visualisation is often used to visualize complex dataset such as ETE that visualize phylogenetic trees and multiple sequence alignments (Huerta-Cepas et al., 2016) and BioPlex Display which visualize protein-protein interaction (Schweppe et al., 2018). Section 2.8 discusses this component in detail with some examples of the available tools.

2.6 Data Modelling for Database Design and Development

Database design and development are crucial to determine the structure of the database as a whole system. It involves data modelling technique that depends on the database system we need. Data modelling is important in determining the logic of data flow. The following Section 2.6.1 presents the common data modelling techniques namely relational, object-oriented and graph data models. Since this research focuses on ontology which is based on the graph data model, Section 2.6.2 describes more detail on ontology topics including a few examples on biological based ontologies, and ontology development methodology, tools and evaluation.

2.6.1 Data Modelling

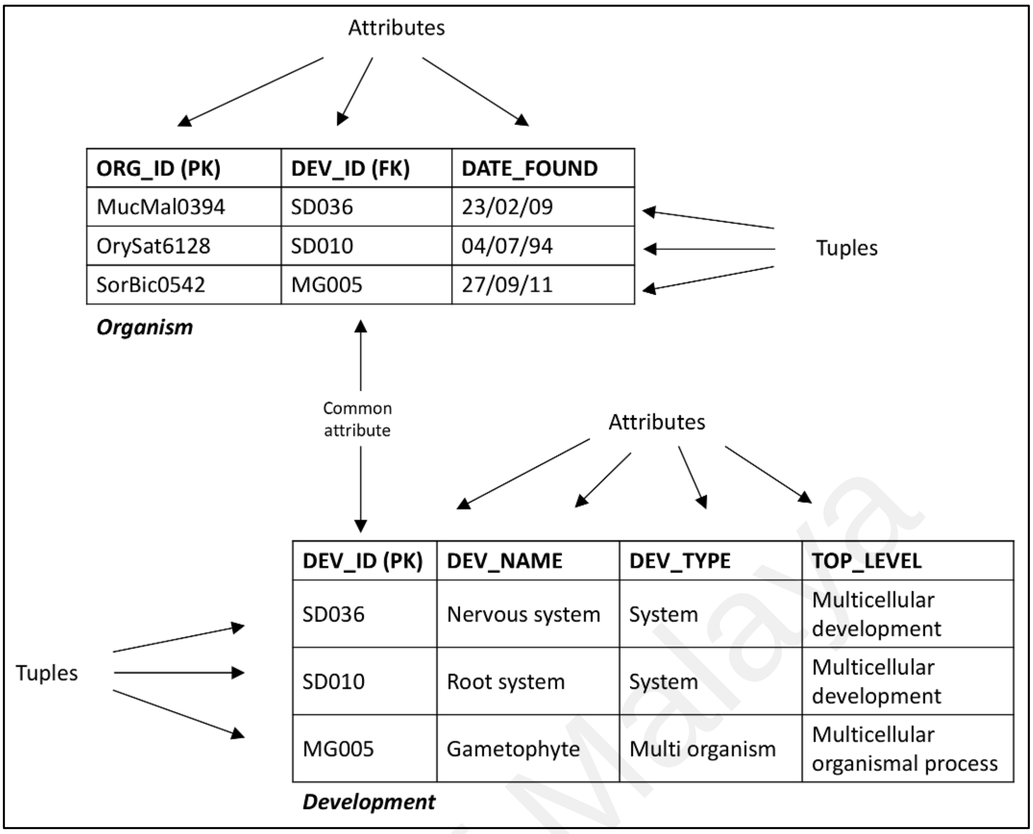
The purpose of data model is to validate that all data objects required by the database are completely and accurately represented and it includes three levels; conceptual data modelling, logical data modelling, and physical data modelling (Siricharoen, 2008). It is illustrated in a data model notation which is often in graphical format (McCaleb, 1999). Database model is a type of data model that defines the logical structure of a database. It also describes the processing and storing of data that occur inside the system. Common data models include relational model (Codd, 1970), network model (Limited, 2010), object-oriented model (Rumbaugh et al., 1991), and graph data model (Kunii, 1990).

However, in this research, only relational model, object-oriented model, and graph data model are elaborated in detail.

i) Relational model

Relational model is a method of structuring data using relations that consist of columns and rows which are also commonly known as table. Relational model provides a means of describing data with its natural structure without covering any additional structure for machine representation purposes (Codd, 1970).

Figure 2.31 illustrates two tables namely Organism and Development that consist of basic components of relational model database. In this model, relation is referred as two-dimensional structure that consists of intersecting rows and columns; each row is tuple and each column represents an attribute. Table Organism has four attributes which are 'ORG_ID', 'DEV_ID', and 'DATE_FOUND' and has three tuples of data while table Development has three attributes which are 'DEV_ID', 'DEV_NAME', 'DEV_TYPE', and 'TOP_LEVEL'. In relational model database, there can be more than one relation or known as table and there is at least one common attribute between tables which creates the relation. Other components in relational model are *primary key* (PK) and *foreign key* (FK). PK is a column in a particular table that contains unique value which exclusively identifies each row, while FK is a column in a table that links to primary key of another table, therefore establishing a connection between them. As illustrated in Figure 2.31, column 'ORG_ID' is labelled as 'PK' and column 'DEV_ID' is labelled as 'FK' in table 'Organism'. Column 'DEV_ID' table 'Organism' links to 'PK' of table 'Development' which is column 'DEV_ID'.



ORG_ID (PK)	DEV_ID (FK)	DATE_FOUND
MucMal0394	SD036	23/02/09
OrySat6128	SD010	04/07/94
SorBic0542	MG005	27/09/11

Organism

DEV_ID (PK)	DEV_NAME	DEV_TYPE	TOP_LEVEL
SD036	Nervous system	System	Multicellular development
SD010	Root system	System	Multicellular development
MG005	Gametophyte	Multi organism	Multicellular organismal process

Development

Figure 2.31: Basic components of relational model database

Relational model is user-friendly as it provides understanding to users in a simple way through the structure of the data that avoids intricacy (Sumathi & Esakkirajan, 2007). This model also provides a flexible structure for database with changing requirements and increasing amounts of data (Gupta & Mittal, 2009) by allowing database developers to easily change the database structure without directly affecting the data. Relational database allows the usage of Structured Query Language (SQL) as the query language to retrieve data from the database. In addition, as this model is based on mathematical concept, users can use mathematical operations without understanding the physical storage or data structure (Ponniah, 2007).

The main disadvantage of this model is that it only focuses on the data structure and not the meaning of the relationships between data (Singh & Gupta, 2014; Sumathi & Esakkirajan, 2007). As an example, while FK does tell users how these tables are connected, but it does not describe it in semantic context. Other than that, flexibility of the database is decreased as the complexity of the data increases (Gupta & Mittal, 2009) due to the need to link all the tables, which can get very confusing due to the huge amount of data and complicated linking between them. Furthermore, it also affects machine performance to respond to a data query if the number of tables is too large.

ii) Object-oriented model

Object-oriented model defines a database as a collection of objects in form of programming languages. In general, objects consist of attributes that define the characteristics of an object, and methods which define the behaviour of an object (Dietrich & Urban, 2011). As an example, if 'plant' is an object, the attributes of 'plant' can be scientific name, plant type, and organism identifier. The methods that can be applied to 'plant' such as assigning 'plant' into its geological distribution, obtaining 'plant' common name, and so on. Objects with the same attributes and methods are known as a class. A class defines the type of object where each object is viewed as an instance or individual of the class. Figure 2.32 shows the basic components of object-oriented model. Object 'MucMal', 'OrySat', and 'SorBic' are instances of class 'Organism' while object 'SD036', 'SD010', and 'MG005' are instances of class 'Development' as each object in respective class consists of the same attributes and methods. Objects in class 'Organism' have attributes 'org_id', 'dev_id', and 'date_found', and methods 'getOrgId()', 'getDevId()', 'getDevName()', and 'getDate()'. While objects in class Development have attributes 'dev_id', 'dev_name', 'dev_type', and 'top_level', and methods 'getDevId()', 'checkDevId()', and 'getTopLevel()'.

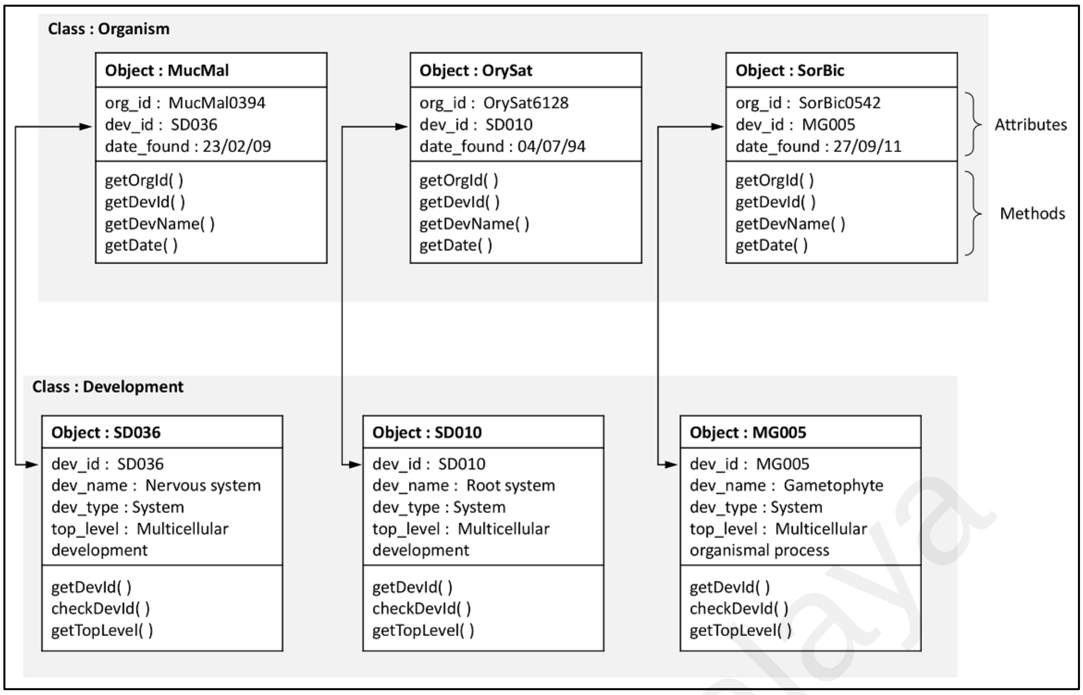


Figure 2.32: Basic components of object-oriented model

This model is typically used for data with complex relationships demanding high performance. It also requires less maintenance as it represents the real world better than other database model. Besides that, object-oriented model emphasizes on the objects rather than on the data itself (Rumbaugh et al., 1991) as the model is based on objects and each object is associated with methods. In addition to that, this model has high flexibility as new methods can be created from existing objects as it can easily be accessed, hence new objects can be created at any time (Prabhu, 2011). New objects may also obtain the attributes from other objects without affecting the structure of the model (Dietrich & Urban, 2011).

Nevertheless, this model also has its disadvantages. It is not suited for all kind of data. It is best suited for dynamic, interactive environments such as for computer aided design (CAD) software (Kim et al., 1990; Liang et al., 1998) and engineering design systems (Kim, 1990; Senturia et al., 1992). Some information system applications for transactional system such as accounting system may not be suitable for object-oriented model due to the nature of the its mechanism. Other than that, there is no universal model for object-oriented database that can satisfy different database system (Ray, 2009), as object-oriented database is developed using programming languages (Yazici & George, 2013) and hence, there is no standardized query language as well.

iii) Graph data model

Graph data model is represented by graph structure (Angles & Gutierrez, 2008). This model is almost similar to another data model, the network model. While both data model represent database in generic graph, yet network model lacks a good abstraction level where it is difficult to separate the model from the actual implementation. In addition to that, the structures of the data are less flexible.

Graph data models are based on graph theory where it applies the usage of nodes, edges and properties (Robinson et al., 2015). Nodes represent entities or instances of a domain. Edges are connections between related instances in which it shows the relationships between them. It is a level of abstraction that only exists in this data model. Properties are attributes of nodes where it gives information of the nodes and it can reside with the nodes and/or the edges. Figure 2.33 illustrates the basic components in graph model; node 'Organism' has two properties which are 'Org_ID' and 'Org_Name' and node 'Development' has three properties which are 'Dev_ID', 'Dev_Name', and 'Date'. Node 'Organism' is connected to node 'Development' by edge 'undergo' while node 'Development' is connected to node 'Organism' by edge 'occurred_in'.

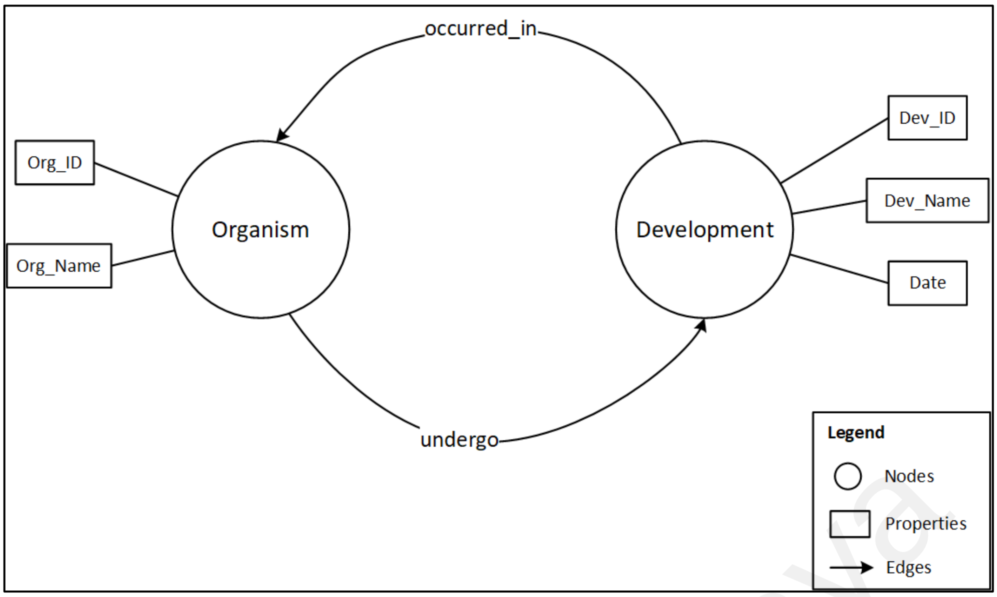


Figure 2.33: Basic components of graph model

Graph data model in semantic technology uses the components of graph theory to form 'triple' (Sakr & Gaber, 2014); a statement composed of a subject, a predicate, and an object in subject-predicate-object arrangement. Triples represent the simplest statements and they describe the relationships of data by the data itself thus providing flexible schemas (Segaran et al., 2009). In general, subjects serve as entities of a domain. Predicates are attributes of the entity and objects can act as subjects of other triples or have literal values such as strings or number (Fowler, 2015). The usage of this type of statement can be seen in many notable graph-based databases such as ArangoDB (ArangoDB, 2014), MarkLogic (MarkLogic, 2017), and OrientDB (Tesoriero, 2013).

Graph data model expresses relationships and connections between data which is ideal for modelling and querying hierarchies within the data. Such feature is very helpful for business related analysis such as master data management which is an extensive method that creates a master reference source for all of business analytic data (Dreibelbis et al., 2008; Robinson et al., 2015).

Other than that, due to the inherent structure in this model (Chen et al., 2004), it enhances querying process where it increases the capability to perform complex query and decrease time needed to complete a query (Robinson et al., 2015) as developers can structure their data including how data are connected to one another and defining its metadata without any restriction. Moreover, the structure of the model can easily be change such as adding new nodes or edges without disrupting any other part of the model (Sikos, 2015). However, while graph data model might be more advance compared to earlier data modelling, it is not useful for any operational-based system as it cannot efficiently process high volume of single task that span the entire database (Robinson et al., 2015).

2.6.2 Ontology

Throughout years of data modelling progress, ontology, which is based on graph data model, has appeared as an alternative in database design (Martinez-Cruz et al., 2012) as it requires a fortified definition of the data. Ontology concept is actually borrowed from the field of philosophy that refers to the subject of existence. In the computer science field, ontology is the s*pecification of conceptualization*; descriptions of the concepts and relationships and a set of definitions of formal vocabulary (Gruber, 1993). *Conceptualization* is expressing the knowledge in terms of entities and *specification* is the representation of the said concept in a concrete form. In other words, ontology refers to a formalization of the domain knowledge (Ahmad, 2012). A body of formally represented knowledge is based on a conceptualization where the objects, concepts, and other entities assumed to exist in some area of interest and the relationships among them (Gómez-Pérez et al., 2007). A conceptualization is an abstract, simplified view of the world to represent. Every knowledge-based system is committed to some conceptualization. Ontology is not exactly a form of database model since it does not provide computational specification of a database system (Martinez-Cruz et al., 2012).

Ontology specifies the concepts, relationships, and other distinctions relevant in representing a domain. The specification is in the form of structural vocabularies such as for classes and relations where it provides meanings and formal constraints on its coherent use (Noy & McGuinness, 2001). In such an ontology, definitions associate the names of entities in the universe of discourse such as classes, relations, and functions with human-readable text that describes its meaning and formal axioms constraining the interpretation and well-formed use of these terms (Sage & Rouse, 2009).

Ontology is common in the biological field in providing a better understanding of the data across different domains. For example, Gene Ontology (GO) (Ashburner et al., 2000; GO Consortium, 2017) is a database comprises of the GO ontologies that includes the annotations of genes and gene products to ensure consistent description of gene products across databases. GO provides standardized vocabularies that represent gene product properties in three domains which are cellular component, molecular function, and biological process. Figure 2.34 shows an example of a set of term related to term '*system development*' where the most general term which is the domain '*biological process*' located at the top of the graph.

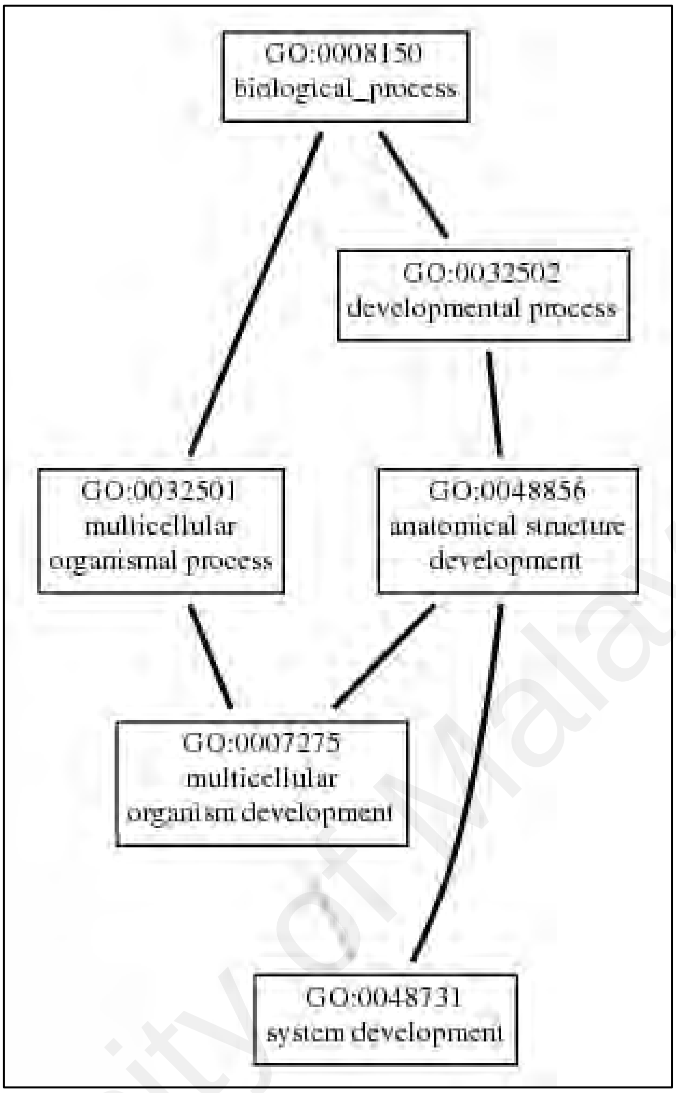


Figure 2.34: Annotation of terms in Gene Ontology. Image reproduced with permission from Carbon et al. (2009)

Another example is Plant Ontology (Avraham et al., 2008; Cooper et al., 2018) which is an open access ontological database that provides controlled vocabularies for plant anatomy and developmental stage. It involves the participation of other collaborating databases which encourages the use of controlled vocabularies across other databases as attributes. Figure 2.35 shows an example of a set of term related to term '*brittle endosperm*' is shown in hierarchy where specific terms are located at the bottom of the graph. However, it is not limited to one-way direction which is similar to GO.

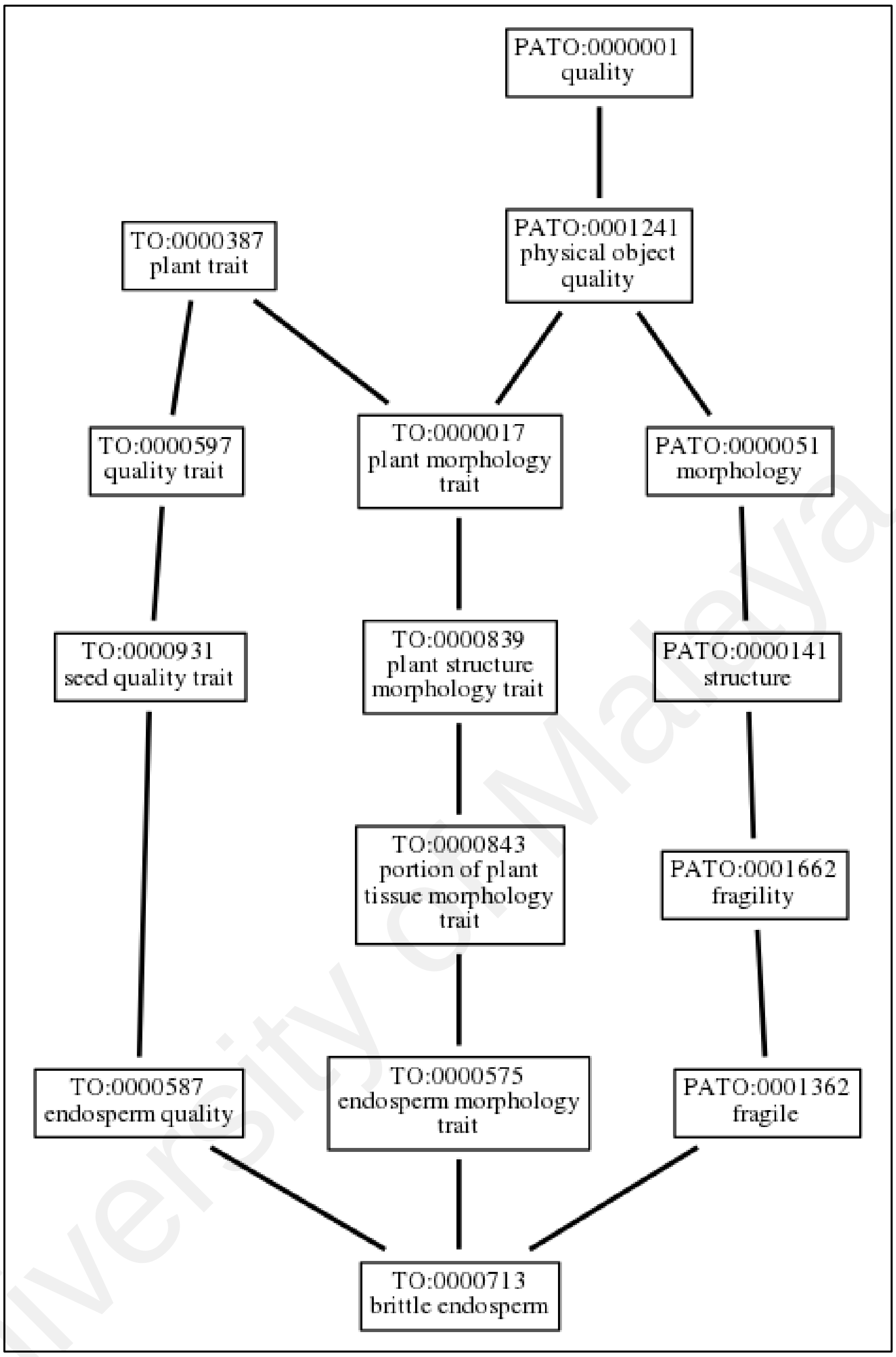


Figure 2.35: Annotation of terms in Plant Ontology. Image reproduced with permission from Cooper et al. (2018)

Components of the ontology make it useful in developing the database. It provides the overall view of the classes in the database to the user. Its main components are concepts, classes, relations, restrictions, and axioms (Chen, 2008; Noy & McGuinness, 2001). Gargouri (2010) describes each components of the ontology whereby a concept represents a set of class within a domain. It represents a group of different individuals that share common characteristics. A class is a conceptual grouping of similar terms which contain individuals, other classes, or combination of both. Members of a class are individuals or instances; they are the base unit of the ontology and contain relations and specific attributes which is in form of data value. Relations describe the interactions between concepts, classes and individuals. Although it is dependent on the ontology language, it is often possible to express different categories of relationships between concepts. Meanwhile, axioms constrain the values for classes or instances stated in a logical form that comprise of the overall theory of the ontology of a domain.

Figure 2.36 illustrates the association between components of the ontology. Within a concept, there are a number of classes that consist of individuals. Each individual contains attributes in form of data value. Furthermore, there are relations that link together individuals, classes as well as concepts which then form the domain. Axioms are generally ruling in logical form that made up the overall theory of the ontology such as datatype definitions, declarations about classes, and assertions (Bock et al., 2012).

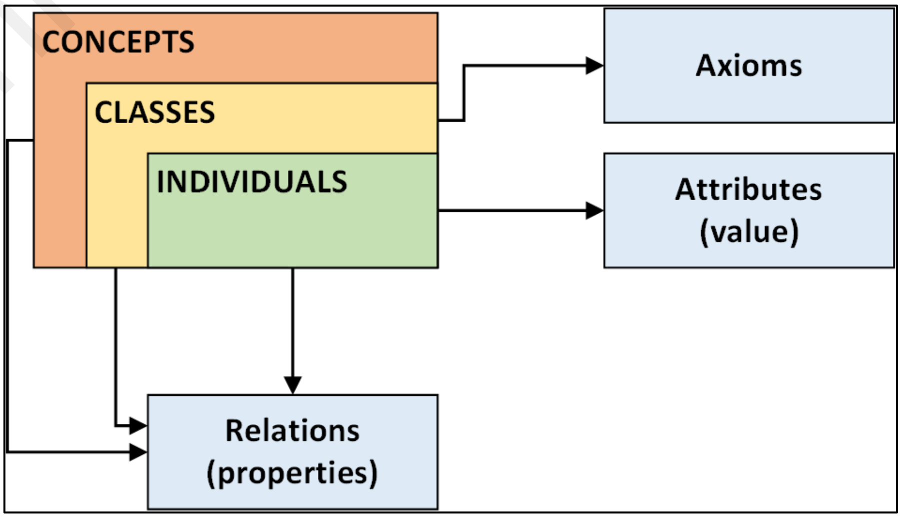


Figure 2.36: The components of an ontology

Ontology has become popular on the World Wide Web. It ranges from large classifications of websites' categories (Song et al., 2006; Zhou et al., 2018) to products for sale and their features (Hepp, 2008). According to Noy and McGuinness (2001), there are five main reasons that influence the usage of the ontology. First, ontology works very well in sharing common understanding of the structure of information among people or users. For example, when different systems use the same underlying ontology, users can extract and gather information from these different websites to answer queries or as input data in another application. Second, ontology allows reuse of domain knowledge, thus permitting users to reuse ontology with detailed vocabulary for their own domain. Several existing ontologies can also be combined into an ontology such as Generalized Upper Model developed from other ontologies which are Penman Upper Model and Merged Upper Model (Bateman et al., 1994). Third, it is possible to change premises of ontology when the knowledge about the domain is changed. Fixed coding of the data into database system would make it a difficult task to change the premises of the domain knowledge and difficult to understand. Moreover, clear and detailed specifications of domain knowledge are useful for new users. Fourth, the domain knowledge can be separated from the operational knowledge. For instance, further investigations and experiments can be done based on different biomedical ontologies such as BioPortal (Whetzel et al., 2011), Bio2RDF (Belleau et al., 2008), and OBO (Smith et al., 2007), or Open PHATCS project (Williams et al., 2012) that focuses to solve specific problems in drug discovery research. Lastly, it provides a way to analyse the domain knowledge once a declarative specification of the vocabularies used is available. It is a very important process for reusability or extension of existing ontology (Lu & Jin, 2002).

The following Section 2.6.2.1 introduces few existing biological based ontologies. Besides that, Sections 2.6.2.2, 2.6.2.3 and 2.6.2.4 describe in detail the methodologies, tools and evaluation in the ontology development. Based on the review, their practices are identified to assist in identifying the requirements of the proposed approach in this research.

2.6.2.1 Examples of biological based ontologies

Ontology is well known in biological field whereby there are a number of biological based ontologies to define the basic terms and relations in biological domains and shared among users in the community as the main reference. Moreover, it is also the foundation for integration and exchange of biological data. The creation and usage of biological ontologies have emerged recently as an important issue in biological community. A few examples of biological based ontologies are described in detail below.

i) Gene Ontology (GO)

GO (Ashburner et al., 2000) provides the most extensive resource regarding the functions of genes and gene products and it is available to be used for computing knowledge. GO defines concepts used to describe gene function and their relationships. It classifies gene functions based on three aspects: (i) molecular function, (ii) cellular component, and (iii) biological process. GO annotation is evidence-based statements that relates specific gene product to a specific ontology term. This ontology is readily available to be exported in different file format such as in OBO flat file format or RDF/XML file format.

ii) Plant Ontology (PO)

Plant Ontology (Avraham et al., 2008) is a collaboration of several plant databases and experts in plant systematics, botany, and genomics with robust and extensible controlled vocabularies that represent the biology of plant structures and developmental stages. Two plant-specific knowledge domains which are the anatomical entities of plants, and growth stages in various plants including their relationships are emphasized in this ontology. One of the main purposes of PO is to develop a standardized data annotation to allow data reusability and sharing among scientific community. Moreover, PO provides a querying process at different level of abstraction to fully facilitate the use of controlled vocabularies.

iii) IDOMAL

IDOMAL (Topalis et al., 2010) is a malaria ontology that covers different aspects of malaria such as clinical, epidemiological, and biological data. It also includes interference attempts to control the disease. It is developed using OBOEdit2 software (Day-Richter et al., 2007); based on Basic Formal Ontology and is a part of Open Biology and Biomedical Ontology (OBO) Foundry. Besides that, this ontology is constructed in the frame of Infectious Disease Ontology where it forms the top-level of the IDOMAL. Furthermore, it is developed by stages where the first version of IDOMAL consists of clinical and epidemiological features of the disease and further extended for other additional components such as immunology and other vector-borne diseases.

iv) Saliva Ontology (SALO)

Saliva Ontology (SALO) (Ai et al., 2010) is a web-based ontology that contains a controlled vocabulary of terms and relations related to salivaomics and diagnostics. This ontology also includes saliva-relevant literature to assist in identifying terms, synonyms, and definitions. In addition, SALO is linked to other external ontologies such as GO, the Protein Ontology (PRO) (Natale et al., 2007), and Chemical Entities of Biological Interest (ChEBI) (Degtyarenko et al., 2008), hence providing better coverage of the ontology.

v) Textpresso

Textpresso (Müller et al., 2004) collects the terms used in scientific literature where the current ontology consists of 33 categories of terms. The categories are divided to classes of biological concepts, relations between two objects, and description of each concept. The whole corpus of articles including abstracts is marked to identify terms within the categories. It also extends the range of the data by including categories from GO database. It is a useful curation tool to search for any biological articles.

2.6.2.2 Ontology development methodology

Ontology development is an important process as it sets the fundamental structure of a knowledge domain. However, there is no one correct way for developing ontology as the best solution depends on the type of application that users need (Noy & McGuinness, 2001). There are a growing number of methodologies that address the ontology development and maintenance. Although these methodologies may be different at some stages, most of them have skeletal processes which are as explained in detail below.

i) Ontology requirement

It refers to what we need in the ontology and prerequisite of the ontology. In this process, developers state the purpose of the development of the ontology, its intended users, and other technical and requirement details (Suárez-Figueroa et al., 2009).

ii) Ontology implementation

It refers to the process of executing the ontology design into life. It involves tools required to develop the ontology. The formal language used in implementing the ontology is also determined by ontology developers. This topic is described in detail in Section 2.6.2.3.

iii) Ontology evaluation

It refers to the process whereby the developed ontology is evaluated to ensure that what is built meets the requirement of the application. It is important in cases where the ontology is automatically populated from different resources that might not be the same which leads to repetitive instances or instances that are clustered according to their sources in the same ontology. In both cases, it may decrease the usefulness of the ontology. This process prompts the next important process, which is ontology refinement. Further explanation of this process is described in Section 2.6.2.4.

iv) Ontology refinement

It refers to the process of refining and improving the developed ontology based on the evaluation result for it to fits the objective of the application better. Ontology implementation, evaluation, and refinement are more or less a cycle stage because developers will keep repeating these three processes to have the desired ontology.

Despite that, ontology development methodologies can also be categorized based on the type of approaches. The variety in approaches for ontology development does not affect the quality of the ontology as it only eases workflow of the process. Application-driven approach is an example of methodology in which the ontology is built for a specific purpose. Examples of this type of methodology are TOVE (Grüninger & Fox, 1995) and by KACTUS project (Fernández-López & Gómez-Pérez, 2002). Another type of methodology is ontology extension where it extends or adds to existing ontology to form a new ontology such as Generalized Upper Model (Bateman et al., 1994) and SENSUS (Swartout et al., 1996); or using tools to facilitate the creation of specification document or integration of brainstorming processes into relevant structures such as OntoEdit (Sure et al., 2002) and CODA (Fiorelli et al., 2010).

Table 2.3 shows the comparison of workflow by these methodologies. Nonetheless, these approaches agree on one thing; processes that take place are clearly defined and can be used in developing any kind of ontology depending on the type of approach that fit the main objective of the ontology.

Table 2.3: Comparison of ontology development methodologies

Type of methodology	Project Name	Type of Approach	Workflow	Ontology Life Cycle	Taxonomy Design
Application driven	TOVE	Middle-out approach	1. Creation of motivating scenarios 2. Formulation a set of competency questions 3. Specification of the terminology Establishment of conditions for the completeness of the ontology	No	Yes
	KACTUS	Top-down approach	1. Specification of the application 2. Development of a preliminary design based on top-level ontological categories Ontology refinement and structuring	Yes	No
Ontology extension	Generalized Upper Model	-	Addition of new classes to existing ontology	No	No
	SENSUS	Bottom-up approach	1. Collection of a series of terms 2. Addition of all concepts in the path from the terms to the root Addition of relevant terms to the domain	No	Yes
Tool-based	OntoEdit	-	1. Requirement specification 2. Refinement of the semi-formal description of the ontology Evaluation of the formal ontology	No	No
	CODA	-	1. Ontology learning 2. Population of ontology with new data Linguistic enrichment of ontology by external resources	Yes	No

2.6.2.3 Ontology development tools

In the ontology development, using the tools that are available either commercial or open source speed the process of designing and developing the ontology. In fact, ontology development tool can be used in all phases of the development; from the creation to the maintenance of ontologies. Currently, there are a number of tools that have been developed to implement the metadata of ontologies using semantic markup languages recommended by W3C such as Resource Description Framework (RDF), combination of DARPA Agent Markup Language (DAML) and Ontology Inference Layer (OIL), and Web Ontology Language (OWL). Some examples of ontology development tools that are currently available are described below.

i) Protégé

Protégé (Protégé, 2017) is an open source tool that allows developers to create and to manage terminologies and ontologies. It provides a platform for developers to use the terminologies in end-user applications. Other than that, this software also provides several features that is very useful such as graphical user interface (GUI) for navigating the graph, visualisation components to view individuals' relationships and programming interface so that users can create terminology-based applications. It also has an extensible collection of plugins where users in the community can design and share their own defined plugins. Protégé helps developers through the process of system development and allow users to save and share their ontologies in owl file format.

ii) NeOn Toolkit

NeOn Toolkit (Haase et al., 2008) is another open source ontology tool that works on different platforms. It supports the development of ontology in Web

Ontology Language version 2 (OWL2) languages. This tool is accessible from the Eclipse platform which is an integrated development environment (IDE). In spite of that, NeOn Toolkit provides a set of plugins that covers a great number of ontology engineering tasks including annotation, ontology evaluation, and ontology reasoning. The main significance of this tool is the initial set of plugins available in Eclipse has extension mechanism of other plugins, preventing any problem in using additional new plugins.

iii) Top Braid Composer

Top Braid Composer (TBC) (TopQuadrant, 2018) is an ontology editor that supports Semantic Web standards such as RDF and OWL. It provides visual editing support as well as reasoning function. GUI provided in this ontology editor is sophisticated; it offers many features such as drag-and-drop, viewing and editing ontologies in different serialization formats, and visual editors for RDF graphs and concept diagrams. This tool comes in three different versions; TBC Free Edition is the basic tool with limited features, TBC Standard Edition includes all features available and TBC Maestro Edition is almost similar to TBC Standard Edition but with advanced support for other features such as Top Braid Live, SPARQLMotion and pre-built web services.

iv) Knoodl

Knoodl is a cloud-based open source ontology tool for creating, managing, analysing as well as visualizing ontology in RDF/OWL descriptions. Some highlight features of this tool are ontology import/export, enable query remote from SPARQL endpoints, dashboard view to aid users for quick and flexible display of query results, and visualize ontology using built-in Google add-ons.

v) Swoop

Swoop is an open source ontology editor that offers various OWL presentation syntax views. It also supports reasoning and ontology comparison where different ontologies can be compared based on their Description Logic-based definitions. Unlike other example of ontology tool mentioned previously, Swoop is a web-based tool and stores the ontology as HTML models. Other than that, Swoop supports W3C standards such as RDF(S), OIL and DAML.

Table 2.4 summaries the findings on the ontology development tools. Findings and analysis help to decide the tools used in the framework of this research.

Table 2.4: Summary of ontology development tools

Features	**Protégé**	**NeOn Toolkit**	**Top Braid Composer**	**Knoodl**	**Swoop**
License type	Open source	Open source	Commercial	Open source	Open source
Semantic web architecture	Standalone	As Eclipse plugins	Standalone Eclipse plugins	Cloud-based	Web-based
Extensibility	Plugins	Plugins	Plugins	Plugins	Plugins
Ontology storage	Files DBMS (JDBC)	Files	DBMS	-	Files
Reasoning	Yes	Yes	Yes	No	Yes
Import format	XML, RDF(S), OWL, text file, Excel, etc	OWL, RDF, XML	RDFa, OWL, RDF(S), UML, text files, etc	-	OWL, XML, RDF, text formats
Export format	XML, RDF(S), OWL, Turtle, JSON-LD, OBO, etc	OWL2, RDF, OWLX, Turtle, OMN	XML, RDF(S), OWL, HTML, F-logic, etc	-	RDF(S), OIL, DAML

2.6.2.4 Ontology evaluation methodology

Ontology evaluation is performed during the process of ontology development to enable a wide adoption of the ontology in the semantic-related application or system. In addition, ontology developer can recognize areas that need editing or showing some parts of the ontology that might cause issues in the future. Besides that, this allows ontology to act as a shared knowledge base in the ontology engineering community.

Ontology evaluation can be done in different types of approaches depending on the kind of ontologies evaluated and its purpose. In general, most evaluation approaches can be categorized to application-based, criteria-based, data text-based, and calculation-based as described below. Besides that, these approaches can also be grouped based on the level of evaluation which are lexical, hierarchy, relations, context, and structure, as many evaluation approaches focus on certain levels of the ontology only.

i) Application-based

Application-based evaluation is used for ontology developed for an application by uploading developed ontology into a built application that is usually a web based, and it is calculated using algorithm and it may involve several ontologies to determine which would best suit a particular purpose (Porzel & Malaka, 2004). An example of this type of evaluation is as done by Porzel & Malaka (2004) where performance- and task- based evaluations were conducted on the ontology.

ii) Criteria-based

Criteria-based evaluation is done by satisfying certain proposed criteria of different dimensions such as based on semiotics' theory (Burton-Jones et al., 2004) or adaptation from an ISO standard (Duque-Ramos et al., 2013).

iii) Data text-based

Data text-based evaluation is an approach that compares ontology with a source of data in which terms are usually extracted from a corpus and number of terms overlapping between the ontology and the corpus is counted. For instance, evaluation approach by Maedche & Staab (2002) measured the similarity between strings and compared it to a golden standard that is based on other ontology. Another example is by evaluating using related topic models where discrepancies in the semantic structure of the ontology are identified (Gangopadhyay et al., 2012). This type of evaluation is more suitable to check the coverage of the ontology. However, it can be said that there are not many ontologies that focused on every area of its knowledge domain.

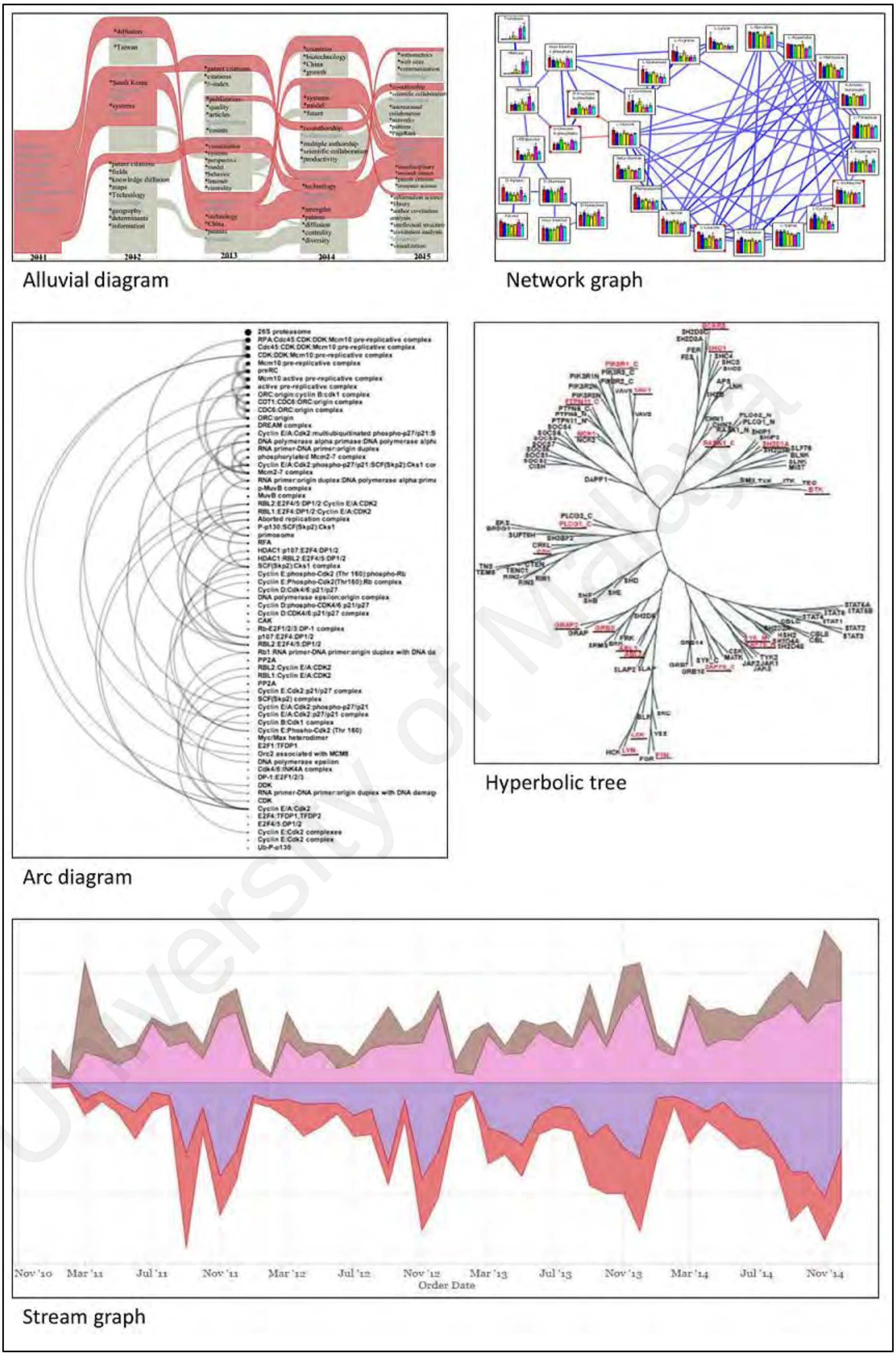
iv) Calculation-based

Calculation-based evaluation applies statistical algorithms such as clustering algorithm (Brewster et al., 2004) or precision and recall method (Euzenat, 2007) for determining the probabilistic model or other mathematical related statement.

2.7 Types of Data Visualisation

One of the essential components in visualisation is type of visualisation that can present the data. Yuk & Diamond (2014) list several general types of visualisation such as graph, timeline, and flow chart. Graph involves the usage of x- and y- axis and generally represents the relationships between two variables, for example the relationships of different gene expression in different conditions (Anders & Huber, 2010). Timeline is a type of visualisation that shows changes of something which is illustrated on a graph. For instance, timeline is used to illustrate the growth of sample available in Gene Expression Omnibus (GEO) database (Barrett et al., 2011) and phenotypic differences in biological systems (Secrier et al., 2012). Other than that, flow chart represents workflow, algorithm or process of how something works. For example, it is used to depict process of defining important areas of valuable plant diversity (Blasi et al., 2011) or to highlight components of a biological tool (Brohée et al., 2008).

Besides that, visualisation can also be classified based on data taxonomy and number of dimensional of the data which are one-dimensional, two-dimensional, three-dimensional, temporal, multidimensional, hierarchical, and network (Shneiderman, 1996). These data can be presented in many types of visualisation. A few types of visualisation as shown in Figure 2.37 are commonly used in presenting biological data such as alluvial diagram (Eren et al., 2015; Ruan et al., 2017), node-link diagram or typically known as network graph (Cline et al., 2007; Junker et al., 2006; Wang et al., 2013). There are also other types of visualisation which are more complicated such as arc diagram (Dang et al., 2015; Wu & Bello, 2010), stream graph (Aldinucci et al., 2011), and hyperbolic tree (Bingham & Sudarsanam, 2000; Manning et al., 2002).



Alluvial diagram

Network graph

Arc diagram

Hyperbolic tree

Stream graph

Figure 2.37: Types of visualisation

2.8 Data Visualisation Tools

A good visualisation combines functioning and effective visualisation tools. The goal of data visualisation tools is to create graphics of data that can be easily interpreted to gain knowledge and insights (Soukup & Davidson, 2002). Data visualisation tool refers to visual library made up of a set of programming languages to design desired visualisation. Some of the available data visualisation tools are described as follow:

i) D3.js

D3.js (Teller, 2013) is a JavaScript-based library to create dynamic and interactive data visualisation in web browser. D3 refers to Data-Driven Documents and this open source library heavily utilizes Cascading Style Sheets (CSS), Hypertext Markup Language (HTML), and Scalable Vector Graphics (SVG) standards which provide controls to user over the final result. In addition, there are many libraries built using D3.js such as (i) *d3sparql* (http://biohackathon.org/d3sparql/) that transforms SPARQL query results in form of JavaScript Object Notation (JSON) format that is applicable for D3 layout; (ii) *C3.js* (http://c3js.org/) a D3-based library for reusable charts such as time series chart, step chart, and pie chart; and (iii) *Dimple* (http://dimplejs.org/) an object-oriented API for business analysis.

ii) Cytoscape.js

Cytoscape.js is an open source library for graph analysis and visualisation. This library focuses on graph theory or also known as network graph. In addition, the library also contains many useful functions in graph theory and can be used together with Node.js; a platform built on Google Chrome’s JavaScript runtime for easy and fast building and scalable network applications.

iii) Leaflet.js

Leaflet.js is (Agafonkin, 2017) another open-source JavaScript visual library specifically for mobile-friendly interactive maps. It is very lightweight with only 38 KB of JavaScript; it contains most mapping features for developers. This library can also be extended with external plugins as well as utilizes CSS3 features.

2.9 Critical Analysis on Plant Data Sources and Presentation

Based on the discussion on the existing plant database systems in Section 2.3, there are four common features that can be found in plant database systems which are query method, form of information available to users, data presentation and interactive feature. Query method refers to the method of requesting information from the database. Form of information refers to the format of the data presented to users. Data presentation is the way data retrieved from the database is presented to users. Interactive feature is any other additional element in the query method or data presentation that can help users in the process of gathering information and analysing the data as well as providing a good user experience in using the system.

Based on Table 2.1, many of the existing plant database systems use the same query methods which are either options where fixed selections are given to users to choose or keywords where users can enter keyword of interest into a search box. Regularly many plant database systems use both methods where users need to choose one of the given options and input a keyword based on that selection. For example as shown in Figure 2.6, users can choose option ‘Scientific Name’ and enter a scientific name of interest into the search box. The purpose of this combination of methods is to speed up the querying process. The next two aspects are related where it is shown that many plant database systems present their data in textual format that is in tables or lists. It is also observed that data from these plant database systems are presented separately where data of each plant species are displayed individually, therefore users would require longer time to gather information that they need. Moreover, this method of data presentation is fixed and unchangeable which only allow users to view the data in one perspective only and prevent them from making an inference. Particularly for plant database with distinct data (e.g. Malaysia Botanical Garden that focuses on plant species in Malaysia) where users would want to analyse the relationships between data, for instance, the distributions of the plant species and its location but it is impossible to do so due to the static presentation of the data in tables and lists and thus it devalues the information presented to users.

Nevertheless, there are several plant database systems that provide data in form of images and illustrations where images of the plants and maps that show the locations of the plants are included in the result. It is interesting to note that for a plant database that provides molecular level of biological data such as maizeGDB that contains genomic datasets, it uses interactive graphic where users can navigate through the different datasets by using tools. This type of graphic is also commonly known as genome browser. Data in forms of images or graphics also offer more information for users given that each image is annotated accurately. The last common aspect in plant database systems is interactive feature in which most of the plant database systems do not have. It is an additional element to the systems yet it can improve user experience in using the system as well as help users in the data analysis process. However, for plant database such as maizeGDB which consists of different type of genomic data, interactive features are important to allow users to explore between different data. For instance, the interactive feature of hovering over coding gene allows users to view the information can reduce the time needed for users to find the information they wanted.

As mentioned previously in Section 2.5, at the present time scientists rely on the relationships in data due to the increasing amount of new scientific data. The conventional approach in presenting data in texts or tables is irrelevant anymore as it does not help to relay the significant associations between data. Hence, it is appropriate to use data visualisation to present the data to users as it enables system developer to organise the data in a way by using suitable type of visualisation and provide a perceptive presentation of the data to users. The summary of biological-based data visualisation in Table 2.2 shows that there is a wide range of type of visualisation that suits for different types of biological data. For instance, a circular layout is useful to make comparisons of different biological data while phylogram is useful in presenting the genomic data such as molecular sequences. In addition, it can increase understanding of complex data as it can combine two or more data in a visualisation. For example, the phylogenetic tree generated iTOL tool (see Figure 2.26) consists of sequences from different kingdoms and its genome sizes and therefore allow users to observe the differences in genome size and the evolutionary relationships of different species. This is because users can process visual information more easily as complex data are summarized in the same visual. Moreover, unlike static data presentation such as table, data visualisation can show changes between data in timely manner where it encourages users to explore and manipulate the data which leads to better analysis. Some of the data visualisation systems also allow users to view the data in different types of visualisation in which it aids users to view the data in different perspectives.

Besides that, the abundance of data visualisation tools makes it easier to develop a functional data visualisation system. Examples of data visualisation tool in Section 2.8 show the variety of tools available freely for developers where each tools specialise for different purpose. Based on the reviews of data visualisation tools in Section 2.8, there are tools that can generate many types of visualisation such as D3.js where users can create dynamic visualisation specifically for web browser meanwhile the other tools have distinct purposes where Cytoscape.js is designated for generating network graph and Leaflet.js emphasises on generating interactive maps. Furthermore, data visualisation tools lessen the processes required in the development of data visualisation where it is easy to use these tools as it implements universal programming languages such as JavaScript, HTML and CSS. The ability of different tools to collaborate should be taken into consideration as well. For instance, Cytoscape.js can be used together with Node.js which assists in synchronisation between client- and server- sides and therefore creating a fast and efficient application.

2.10 Summary

This chapter provides the findings from the literatures to identify problems focused in this research. Particularly, Malaysian plant databases are rich in various data. However, the data retrieved from these databases are typically presented in table form which is suitable for viewing and gaining the information only. Data visualisation is then introduced to enhance the method in delivering the data and information up to the knowledge enrichment. The details of problem definition are described in Chapter 3. Besides that, the literature review assists in identifying the requirements of the proposed solution as presented in Chapter 3.

CHAPTER 3: PROBLEM DEFINITION AND SOLUTION

3.1 Introduction

All information regarding this research are gathered and analysed to identify the problems and the solutions are proposed. This chapter defines few issues described in Section 3.2. In general, this research focuses on the data presentation, and data modelling issues especially on complex plant data. Based on these problems, the need of an alternative approach of data presentation by integrating data visualisation into the database system is justified as described in Section 3.3. The proposed solution and user requirements are explained further in Section 3.4 and Section 3.5 respectively.

3.2 Problem Definition

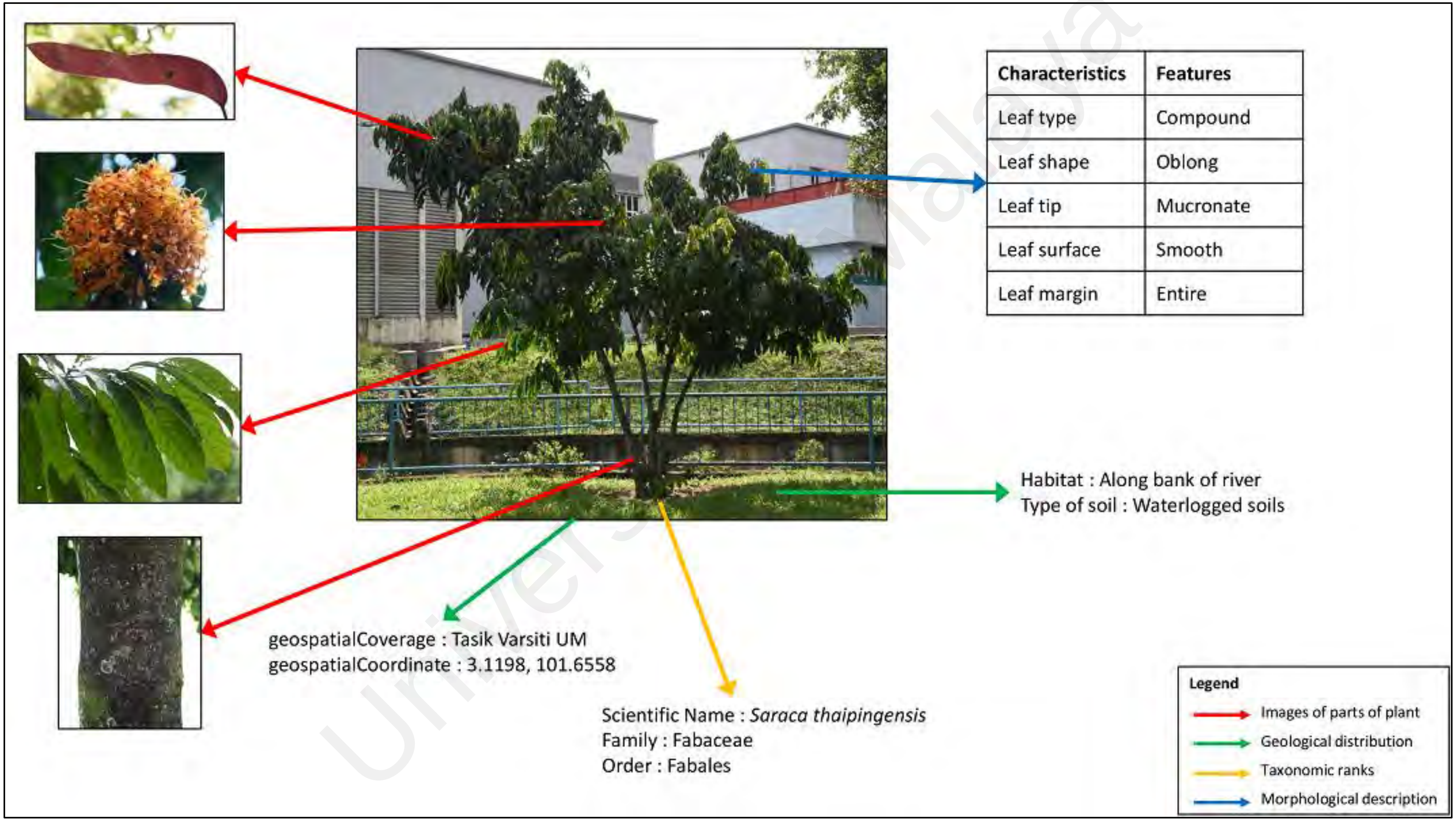
Based on the literature review done in Chapter 2, there are four main issues derived in this research as the following:

3.2.1 Plant Data Description

The amount of plant data is enormous; there are many types of data that can be extracted even at the top abstraction level. For instance, morphological characters that describe structures of plants such as stems, leaves, and flowers; taxonomic data identifies and classifies plant species according to the ranks in taxonomic study; geological distribution defines the attributes of the plant species habitat; and images of morphological form of plant species.

For instance, Figure 3.1 shows how a plant species can be described by different type of data. As shown in this figure, data such as morphological description of the plant, taxonomical data, and geological distribution can be obtained from a plant species.

Additionally, images of each part of the plant such as bark, leaf, flower, and fruit can be used for further analysis where from these images, patterns or shapes of the part can be distinguished and digitized. This information are very valuable for different purposes such as digital measurement of plant species (Easlon & Bloom, 2014; Jin et al., 2015; van Stan et al., 2010), detection of plant diseases (Arivazhagan et al., 2013; Rumpf et al., 2010; Singh & Misra, 2017) and identification of plant species (Carranza-Rojas et al., 2017; Cope et al., 2012; Kumar et al., 2012).



Characteristics	Features
Leaf type	Compound
Leaf shape	Oblong
Leaf tip	Mucronate
Leaf surface	Smooth
Leaf margin	Entire

Habitat : Along bank of river
Type of soil : Waterlogged soils

geospatialCoverage : Tasik Varsiti UM
geospatialCoordinate : 3.1198, 101.6558

Scientific Name : *Saraca thaipingensis*
Family : Fabaceae
Order : Fabales

Legend
- Images of parts of plant
- Geological distribution
- Taxonomic ranks
- Morphological description

Figure 3.1: Plant data description in textual and image forms

Furthermore, different type of data can be associated with one another to improve the interpretation and understanding of the plant species. For instance, Figure 3.2 shows a conceptual map of the data association based on Figure 3.1. As shown in Figure 3.2, *Theobroma cacao*'s geological distribution has ecological attributes. Therefore, it can be concluded that the location of *Theobroma cacao* which is Tasik Varsiti UM has waterlogged soils (see Figure 3.1). These findings from the data association will help researchers in making analysis and new insight about the plant.

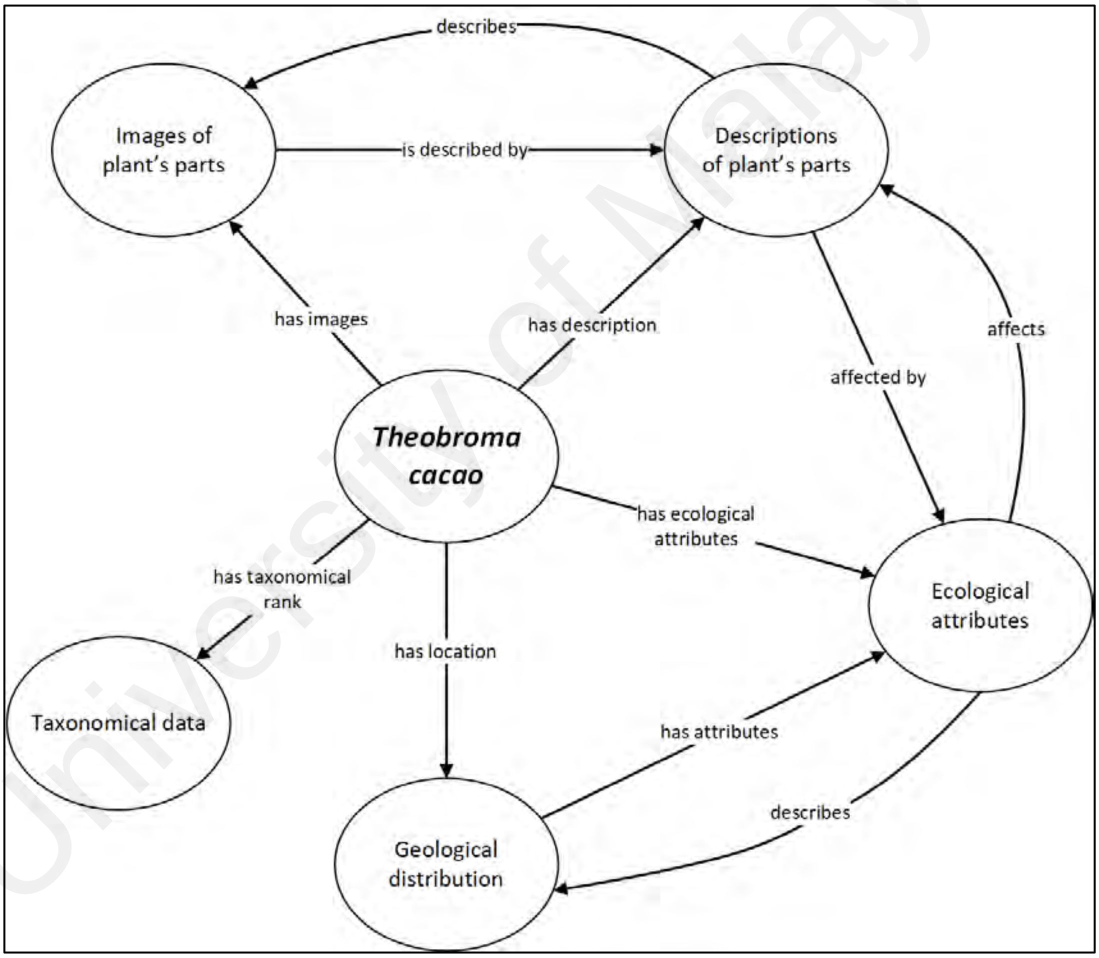


Figure 3.2: Conceptual map of data association between plant data

One of the important objectives in plant biology is to understand the mechanisms and functions of characteristics of plant species that can assist researchers to perform detailed study or comparison study between plant species (Council, 1989; Grierson et al., 2011). Apart from the physical characteristics of plant species, there are also plant genomic data from experiments such as gene expression profiling (Alba et al., 2004; Dash et al., 2012) and genome sequencing (Abe et al., 2012; Dong et al., 2004) which means there are even more terms required. However, there is an issue where some plant structures are described by their species-specific terms (Ilic et al., 2007). For example, fruit of Arabidopsis genus are often referred to as silique (Bates et al., 2013), fruit of *Oryza sativa* it is called grain or caryopsis (Kourmpetli & Drea, 2014), and fruit of *Zea mays* is known as kernel (Eckhoff et al., 2003). Hence, due to species-specific terms, it constrains researchers in correlating data between plant species.

Furthermore, inconsistent terminologies used in different databases have led to ambiguous meaning and confusion among users especially for novice. Efforts done by Biodiversity Information Standards organization that developed and maintained standardized vocabularies commonly known as TDWG standards (TDWG, 2018) encourages uniformity in terms used to describe plant species. Another concern is that with the rapid progress in technologies, terms used in plant biology undoubtedly will expand and evolve over time. This will lead to complicated association between terms which compromise its usefulness if there is no action taken. Consequently, more robust data modelling approach are demanded to cater future needs.

3.2.2 Plant Data Presentation

Generally, data presentation provides information to the users and as discussed in Section 2.4, different approaches in data presentation emphasize certain part of the data. While data presentation in textual form is the conventional approach, its inability to allow users to perform further analysis has devalued the information as the static presentation of data in tables only allows users to view the data in one perspective only. As shown in Table 2.1, many plant database systems commonly present data in form of textual and lengthy descriptions. It is inconvenient for users as it requires extra time and work for users to read through those texts to find information that they need. Furthermore, plant data are commonly presented to users individually, meaning that the data of each plant species are displayed separately. Therefore, the textual form of the plant data cannot demonstrate how a plant species is related to another which means that this approach does not offer the opportunity for users to analyse the data or to make an inference.

Data visualisation is an alternative approach in presenting the data to users and is gaining popularity in recent years. In biological field, it is also used to understand biological systems and omics data. Examples include BioVis Explorer (Kerren et al., 2017), Dendroscope (Huson et al., 2007), and WikiPathways (Slenter et al., 2018) utilize visualisation tool and its features to present their data to users. By visualizing the data, it organizes and structures the data to convey understanding and internalize cognitive references to users (Rowley & Hartley, 2008) which then allow users to analyse information given and deduce new inference in which it leads to new knowledge finding.

3.2.3 Other Issues

In addition to the main issues described above, most plant databases exist separately such as textual database or image database. This causes users to alternately view different databases to obtain and gather information which requires more work. Furthermore, image database typically lack description of image which is not helpful for data analysis.

The query method is an issue too as querying is a vital process involved in database management. A query retrieves data from the database. Querying process can be done by text (FRIM, 2017; Kattge et al., 2011; The Plant List, 2013), image (Şaykol et al., 2005; Tsai et al., 2010), and audio (Lei et al., 2008).

Many plants databases favour textual querying. Based on the general discussion of plant-based databases in Section 2.3, databases such as PlantSearch (Botanic Gardens Conservation International, 2017), Native Plant Database (Evergreen, 2017), maizeGDB (Andorf et al., 2016), MyBIS (MyBIS, 2017) and MyCHM (FRIM, 2017) require users to search for any keyword based on query parameters available while databases such as TRY Plant Trait Database (Kattge et al., 2011) and PLANTS (Natural Resources Conservation Service, 2017) allow users to query by choosing from the given options. It is common that textual querying involves query using the terminologies such as scientific name, habitats and distribution locations. This should be considered when designing query tool as it influences the flow of data storage and retrieval.

3.3 Needs for Data Visualisation to Present Plant Data

Plants help to create an environment suitable for human habitation; it plays a role in regulating global climate, as well as provides food and other practical applications

(National Research Council Committee, 1992). Therefore, it is crucial for researchers to perform more studies on plant biology.

Even though Galperin et al. (2017) state that there are only a small number of plant databases compared to other kinds of biological databases, but these plant database systems still exist in many places in this world, as well as in Malaysia which focus on different data perspective as described in Section 2.3.1. Besides that, these databases provide the same common features which are the query and searching tool, and present the retrieved data in many forms as described in Section 2.4 to the users for viewing and gaining the information purposes, but not for knowledge enrichment. This knowledge enrichment can be supported through data visualisation.

Particularly for Malaysian plant databases, there is no database providing data visualisation with interactive features to present their plant data in textual and image forms. Thus, this research focuses on the data visualisation as an alternative approach in data presentation to the plant for better efficiency in plant data management and hence, encouraging researches on plant data. Data visualisation elucidates the relationships between plant data to obtain new insight from the presented data.

Presenting plant data in visualisation form provides a better understanding of the data as it clarifies the main point of the data in different angles. Furthermore, interactive features that come with data visualisation such as zoom (Kerren et al., 2017), filter tool (Huson et al., 2007), export tool (Slenter et al., 2018), and display options (Dash et al., 2012; Okonechnikov et al., 2012) allow users to manipulate the data to their liking before making any analysis. Data visualisation also present large amount of data clearly without clutters of texts, making it easier for users to draw inferences. In addition, relationships and patterns between data are easily recognizable from the visualisation of extensive amount of the data through choosing the right type of visualisation.

To use data visualisation, there is a need to look into data modelling as it determines the organization of the data into a logical structure as any data collected can be meaningless without a proper method to organize it into significant details. Relational model is the simplest data model that structures data in a manner that avoids intricacy and allows users to change requirements of its structure. However, it only focuses on the data structure in a logical way without considering its structure in semantic context. In addition, as the complexity of the data increases, it decreases the flexibility to alter the structure of the data. Relationships within and among tables are implemented using the key values, yet it does not describe the relationships between data and thus, lost its meanings.

Meanwhile, object-oriented model treats data as a collection of objects where they are organized in classes and each object is linked to its attributes and methods. It does not have a rigid structure as it can be updated and extended. Its flexibility gives it much power such as higher productivity and better quality. Furthermore, it is possible to reuse the same data model as operations of the data model are coded internally which means less programming code is required for an external application. However, this model focuses heavily on the objects rather than the data in programming context. It is also mainly suitable for dynamic and interactive systems.

On the other hand, graph data model represents the data structure in the form of graphs. This model mainly focuses on the semantic context where the graph can be represented in the form of triple statement that describes the relationships of the data and hence, provides flexible structure. Ontology is an alternative database design based on the graph data model. Ontology technique requires solid definition of the data and provides better understanding of the data across different domains (Cooper et al., 2018; GO Consortium, 2017; Smith et al., 2007). As described in Section 2.6.2, the advantages of using ontology in developing domain knowledge are common understanding among users, reusability, and flexibility of the ontology prove that it is an alternative approach that should be considered in representing plant data in semantic context.

There are a few factors that are important in the development of the plant ontology and the plant data visualisation system such as plant data representation, annotation of plant images, type of visualisation, and expected output of the data visualisation. Suitable vocabularies are important to describe the plant data and to annotate plant images so that plant data are represented in a meaningful, flexible and accurate manner so that any additional vocabulary can be included in the future without contradicting the previous vocabularies used and does not need to change the whole data structure. Moreover, the selection on type of visualisation depends on how the data will be presented to users. Different types of visualisation have different purposes in delivering the information to the users. Last but not least, the output of the data visualisation should be functioning well and efficiently produce the correct output.

3.4 Proposed Solution

Figure 3.3 illustrates the framework of the proposed solution in this research. It consists of two main components, on the server-side and client-side.

On the server-side, the database acts as data storage to store all data in textual and image forms. Thus, this plant database consists of ontological plant data and image database locally stored in the server file system. The database design is described further in Section 4.5. The data are then retrieved and visualised through a web-based system, further described in Section 4.6.

As on the client-side, the graphical user interface (GUI) is used for users to interact directly with the system on the server-side. From the GUI, users can send a text query to the system. In the server, ontological plant data and images are retrieved using SPARQL query language and the retrieved results are then presented to users in a visualisation form. The overall system functional and non-functional requirements are as defined in Section 3.5.

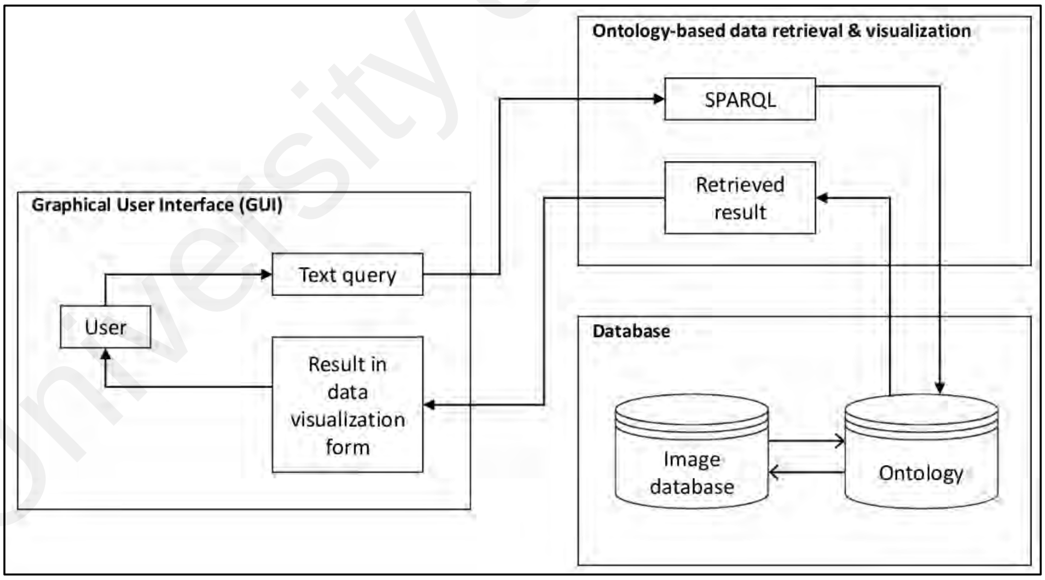

Ontology-based data retrieval & visualization
SPARQL
Retrieved result
Graphical User Interface (GUI)
Text query
User
Result in data visualization form
Database
Image database
Ontology

Figure 3.3: The proposed solution

3.5 User Requirements

Based on the analysis of current existing database systems in Section 2.3 and Section 2.6, user requirements are determined to set the scope and expectation of the system (Isson, 2015). User requirements are classified to two categories which are functional requirements that concentrate on the functionality of the system (Aurum & Wohlin, 2005), and non-functional requirements that serve as selection criteria for operation support of the system (Chung et al., 2012). Users of the system can be system developers and end users such as botanists, researchers, students and laymen.

There are four functional requirements for the system as described below:-

i. **Data storage**

   It is where the data of plant species and its samples are stored in ontological form and images of plant samples are stored in the sever file system directory. This requirement only involves system developers where from the data storage, the process of data retrieval to visualisation can be determined.

ii. **Querying process for data retrieval**

   This allows users to query data from the database using the GUI. In this research, the retrieved data is displayed in a visualisation form.

iii. **Visualisation of the relations between plant data**

   The relationships between data are crucial to highlight data association in the knowledge domain. Three types of relationships are considered:-:

   i) Relationships between taxa, for example a taxon is linked to one another by the taxonomical ranks

   ii) Relationships between samples where a sample is related to another as they are obtained from the same location

   iii) Relationships between taxa and its sample in which each taxon has three samples and each sample has a unique identifier.

iv. **Interactive features in the data visualisation**

It is to assist users to explore the visualisation dynamically and manipulate the displayed data for analysis.

On the other hand, non-functional requirements of the system establish the quality of the system. A few criteria are chosen whereby they act as the attributes of the system and can be clearly expressed in the system to ensure that the objectives of the system development are successfully fulfilled. The three non-functional requirements for the system are:-

i. **Usability**

Usability of the developed system is set to be efficient and easy to use especially for users without computer skill. It can be used by the users with different background such as botanists, researchers, as well as layman users with interest in plant data. A well-written manual on how to use the system is prepared for first-time users who may need guidance.

ii. **Accessibility**

The developed system is accessible openly as a public web-based visualisation system. Users can access the system from anywhere without any hassle as long as an internet connection is available.

iii. **Extendibility**

The system is designed with extendibility function for future advancement. The developed ontology can be modified without changing the overall structure of the ontology. Modifications such as adding new classes or changing the metadata of existing classes and properties can be easily done to adapt to any changes in the future. In addition, parameter in querying process can be modified or added in the future for advanced query tool.

3.6 Summary

This chapter provides a summary of the identified problems in plant data presentation especially on the online database data presentation. Data that are presented in table or list are suitable for viewing and gaining the information only, but not for deducing the knowledge, which is important for data interpretation and analysis. This problem however relates to the data modelling issue in a way that to give a meaning to the data. In addition to that, the overview of the proposed solution in which all materials and methods used for the system development are described. In general, the proposed solution consists of database to store plant data in textual and image forms using ontology; the system for ontology-based data retrieval and visualisation; and the GUI for the users to communicate with the system.

CHAPTER 4: RESEARCH METHODOLOGY

4.1 Introduction

This chapter further discusses the proposed solution which includes the research methodology and activities that took place to achieve the objectives of this research. Furthermore, the whole system design and implementation using the development tools are described. The system testing is performed to confirm whether the defined users' requirements mentioned previously in Section 3.5 are achieved or not to fulfil the objectives of this research. Figure 4.1 shows the main research activities that took place in solving the problems defined in this research and each activity is elaborated in more detail in the following sections.

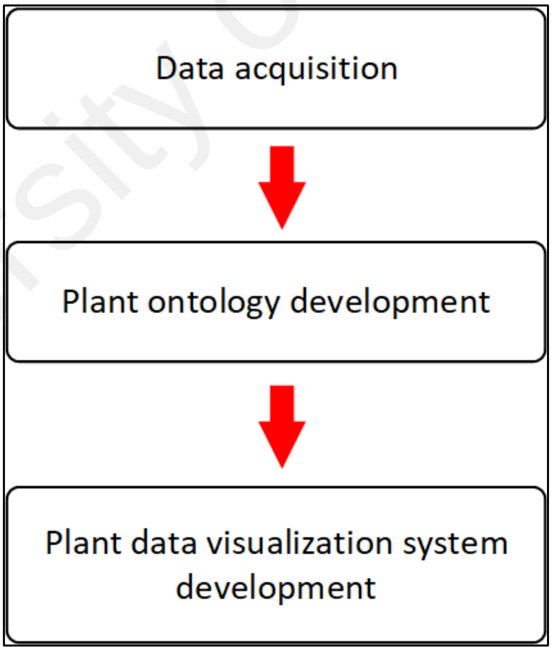

Data acquisition
Plant ontology development
Plant data visualization system development

Figure 4.1: Main research activities

4.2 System Design

System design is the process of defining the structure of the system where each element involved in the system is described in detail (Sage & Rouse, 2009) to satisfy the specified user requirements. It includes the system architecture, graphical user interface and data modelling.

System architecture is the conceptual model that describes the structure, behaviour and views of a system (Jaakkola & Thalheim, 2011). It is the fundamental arrangement of a system in which its components work together in a way or another and the standards to control its design and development. Figure 4.2 illustrates the three-tier system architecture in this research namely data-tier, application-tier, and presentation-tier that hosted in server environment. Data-tier contains the backend plant database which consists of ontology-driven plant data management and image database; Application-tier involves the processes for data retrieval and visualisation; and Presentation-tier consists of user interface which is responsible for displaying data and interactions with the users.

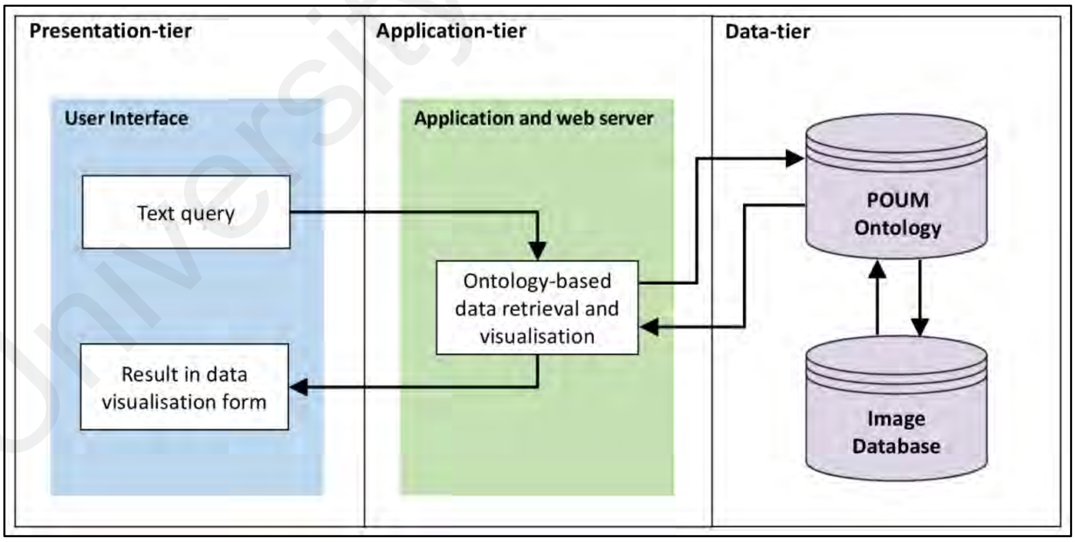

Presentation-tier
Application-tier
Data-tier
User Interface
Text query
Result in data visualisation form
Application and web server
Ontology-based data retrieval and visualisation
POUM Ontology
Image Database

Figure 4.2: The system architecture

i) Data-tier

Images of plant samples are organized in separate data storage to allow image retrieval and for visual display purposes. The ontology contains the textual annotations of the plant images and the plant species. Generally, the vocabularies or terminologies used in biological field are always evolving over time. Hence, new additional vocabularies might be needed in the future. For this reason, the process of ontology development is based on the evolutionary prototype model (Hyvönen et al., 2002) as shown in Figure 4.3. This model allows ontology to be refined or updated in the future without the need to adjust the whole data structure and ontology testing can be done as well to improve the fundamental structure of the ontology.

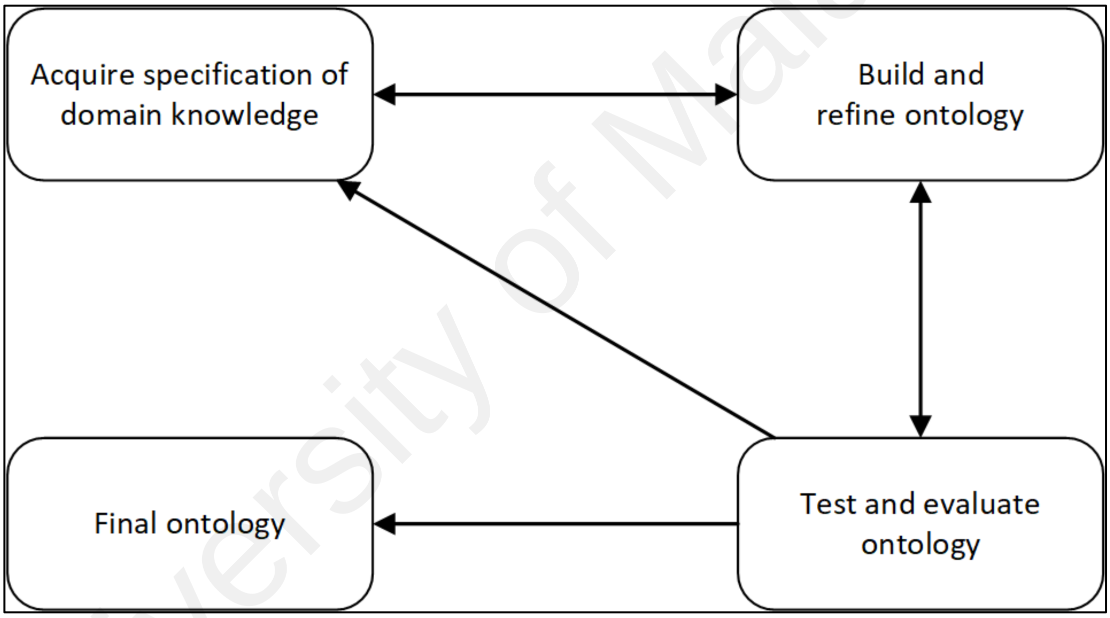

Acquire specification of domain knowledge
Build and refine ontology
Final ontology
Test and evaluate ontology

Figure 4.3: Ontology evolutionary prototype model adapted from Styrman, 2005

ii) Application-tier

In application-tier, users' query is passed to ontology-based data retrieval where it is further processed by retrieving data from the ontology and image database in data-tier. The retrieved data is then visualised to user through user interface.

iii) Presentation-tier

User interface provides a medium for interaction between the users and the system to allow effective operation and control of the system whilst the system simultaneously provides information need for any decision-making process. In this research, the user interface is designed to avoid the need for the users to have to communicate directly with the database system. There are two types of user interface created which are interface to get input from users and the output interface to display the retrieved data from database. The input interface requires users to choose from the given query parameters whereby the emphasis is given to the following parameters namely 'Scientific Name', 'Family Name', 'Location', and 'Water Usage'. The output interface shows the retrieved data in visualisation form along with interactive features to encourage communication between the users and the visualised data.

4.3 Development Environment

In this research, the development tools are defined on the software specification and hardware specification needed for system development and the Java 1.8 programming language is chosen as the main programming language for the system development.

Figure 4.4 presents the software development tools environment for building the system. There are two important tools involved which are Protégé 5.2 as the ontology editor and Eclipse Luna as the main code editor.

Protégé 5.2 is an open source software widely used as the ontology editor as it has a simple interface yet provides extensive support such as definition of axioms, merging of ontologies, and plugins. Additionally, for the ontology testing, Pellet API

(Pellet, 2017) is used as the Reasoner. Pellet is an open source reasoner for OWL language and based on Description Logic (DL). Pellet provides functionality to check the consistency of the ontology, calculate the hierarchy of the classification and explanation of the inferences. Meanwhile, Eclipse Luna provides support for any Java based development tools, plugins, and web tools platform. Both software have huge community support which is helpful to find any solutions to problems occurred during the development.

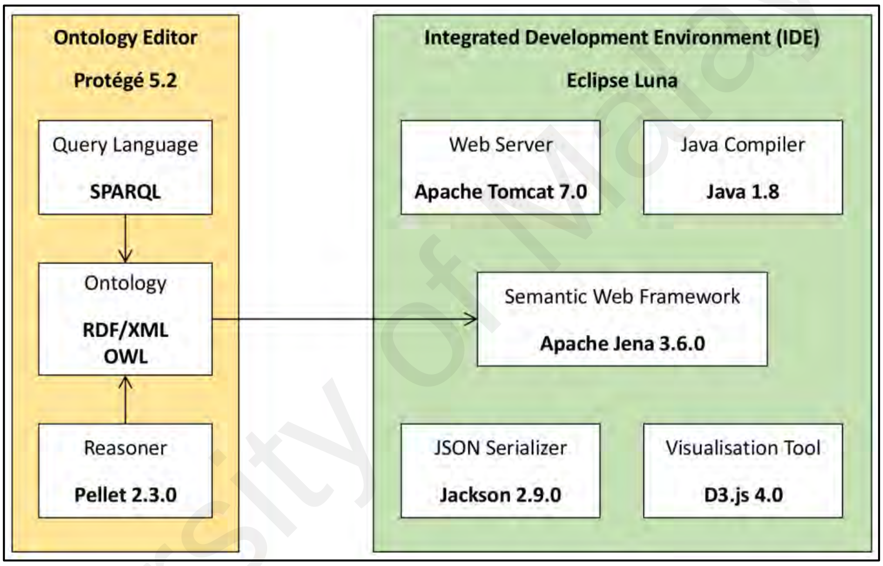


Figure 4.4: The software development tools environment

The RDF, RDFS and OWL are chosen as the meta-languages to implement the ontology. RDF is defined as a language for expressing data models using triple statements. In addition, RDFS and OWL are used to add semantics and more description in the ontology. RDFS is a general-purpose language that provides specific vocabulary for RDF to describe classes and properties while OWL adds semantics to the domain knowledge where it describes relation based on description logics (Domingue et al., 2011). The ontology is then presented in RDF/XML serialization format.

Furthermore, the other support libraries are also used and plugged-ins into the IDE Eclipse Luna. As the designed system is a web-based system, therefore Apache Tomcat 7.0 is used for the web server service. To manipulate the ontology using programming languages, a semantic web programming framework is needed as a medium to communicate. The selection of tool is dependent on the development language used to develop the system and the features that are provided. Therefore, Apache Jena 3.6.0 is chosen to support semantic web framework. Moreover, it supports different database platforms such as MySQL, DB2, and PostgreSQL; provides predefined reasoners such as transitive reasoner, generic rule reasoner, and RDFS rule reasoner; and supports SPARQL query language (Apache Jena, 2018). As for data visualisation tools, D3.js 4.0 is used as the tool to develop the visualisation for the data retrieval from the graph data model and Jackson 2.9.0 is used to serialize textual result from the query into a JSON format.

The other internet programming languages such as Hypertext Markup Language (HTML), JavaScript, Cascading Style Sheets (CSS) and Scalable Vector Graphics (SVG) are also used for creating the client-side user interface; while Java Server Pages (JSP) is used for processing the query on the server-side.

As for the hardware specifications used during the system development, the hardware on the server-side and client-side are as listed in Table 4.1.

Table 4.1: Server- and client- side hardware tools

Category	Hardware tools	
	Server-side	Client-side
Processor	Intel® Core™ i5-6300HQ CPU @ 2.30GHz 2.30 GHz	Intel® Core™ i7-3517U CPU @ 1.90 GHz 2.40 GHz
RAM	4.00 Gb	4.00 Gb
Hard disk space	470 Gb	230 Gb
Internet	1.0 Gbps	100 Mbps

4.4 System Implementation

System implementation explains in detail about the processes that took place to solve the issues as previously discussed in Section 3.2. The system is built according to the system design defined in Section 4.2 and using the hardware and software specifications as described in Section 4.3. The following Section 4.5 explains the processes that took place in the plant database design and development, while Section 4.6 explains the processes involved in the data retrieval and visualisation system design and development.

4.5 Plant Database Design and Development

Database design in this research focused on the relationships between the plant data, i.e. plant species and their samples. Hence, the plant data description and the relationships between the data are carefully defined with appropriate vocabularies so that the represented data have meaning as described in Section 4.5.1. Once the plant data are defined, the sampling is performed for data acquisition as described in Section 4.5.2 to develop the plant database that consists of image database and ontology-driven plant data management as explained in Sections 4.5.3, 4.5.4, 4.5.5, 4.5.6 and 4.5.7.

4.5.1 Plant Data Description

As discussed in Section 3.2.1, there are many kinds of data that can be extracted from plant. Therefore, a few criteria for selecting data were set beforehand. The first criterion is the plant data should be a characteristic of the plant that can be observed easily without any need of tool. Other than that, the data should also be available in other public plant databases to ensure consistency and for extendibility in the upcoming research if required. Furthermore, comparison between data from public plant database and data obtained from data acquisition process can be performed which is useful for future work. In addition to that, because of the plant's environment may influence its characteristics, any observation of the location of the sampling should be taken into consideration.

In this research, plant species is described properly and data obtained in this process is organized into the ontology for further processes such as data querying and retrieval. Moreover, images of parts from the plant species are also used as a part of plant data description as it illustrates the morphological characteristics of plant species in digitized form.

Figure 4.5 shows an example of data that can describe a plant species such as taxonomical classification, morphological characteristics, ecological attributes, geographical distribution, and images of plant species. Taxonomical classification from species level to kingdom level is described for each plant species. Plant morphological characteristics represent features of parts of plant species such as bark, leaves, and flower (Mishra, 2004). In this research, morphological characteristics of plant species and plant samples are described separately due to different habitat conditions. Morphological characteristics of plant samples are acquired based on images of plant species which depicts parts of plant species such as bark, fruit, flower, leaf as well as the whole tree.

Besides that, ecological attributes and geological distributions of each plant samples are depending on the locations of plant samples. This is because different locations of plant samples as shown in Figure 4.6 have distinct environmental conditions which affect the appearances of the plant samples. Ecological attributes of plant species refer to the level of water usage for plant growth, type of soil, and type of habitat while geological distribution determines the location of the plant species.

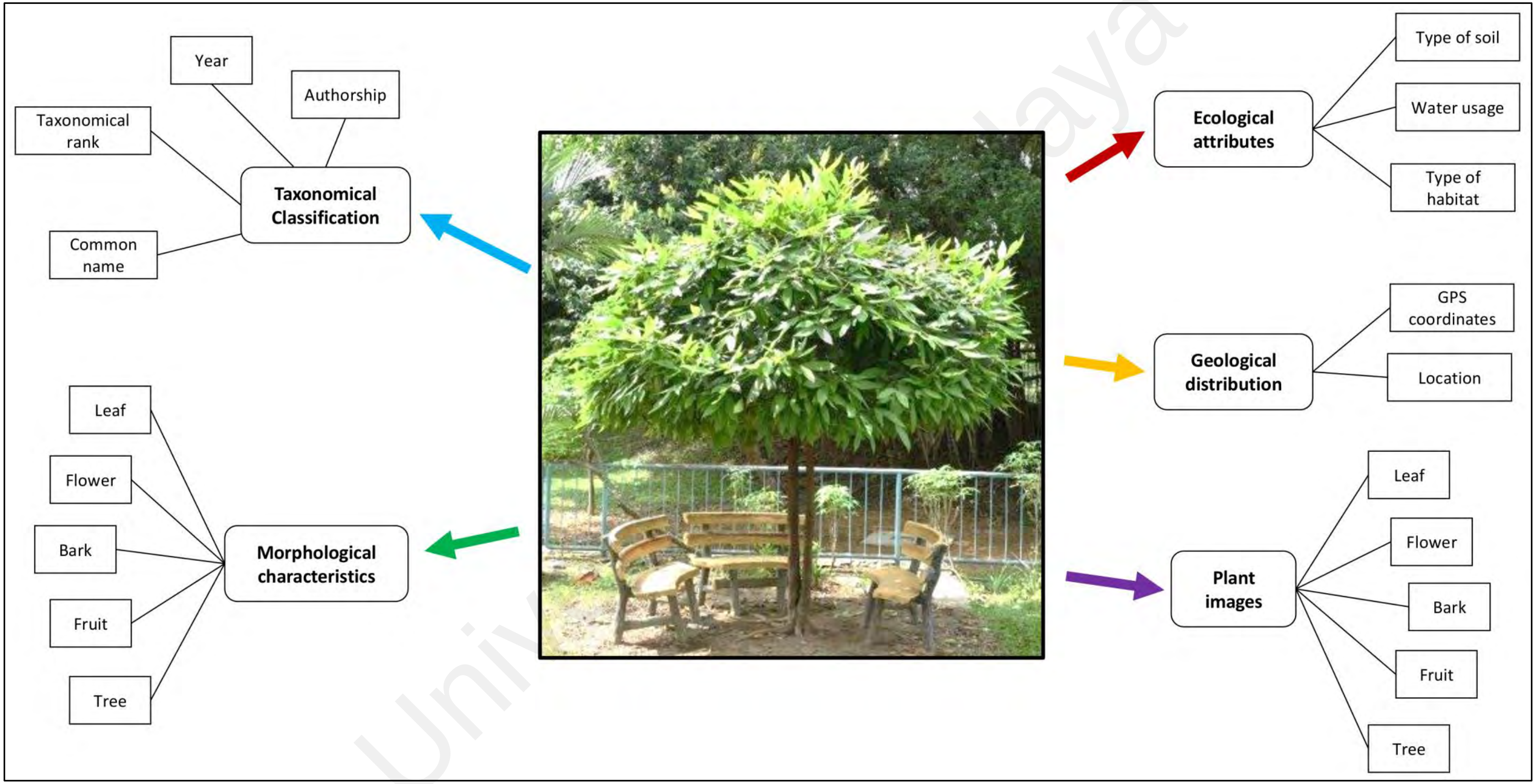


Figure 4.5: Plant data description and the relationships

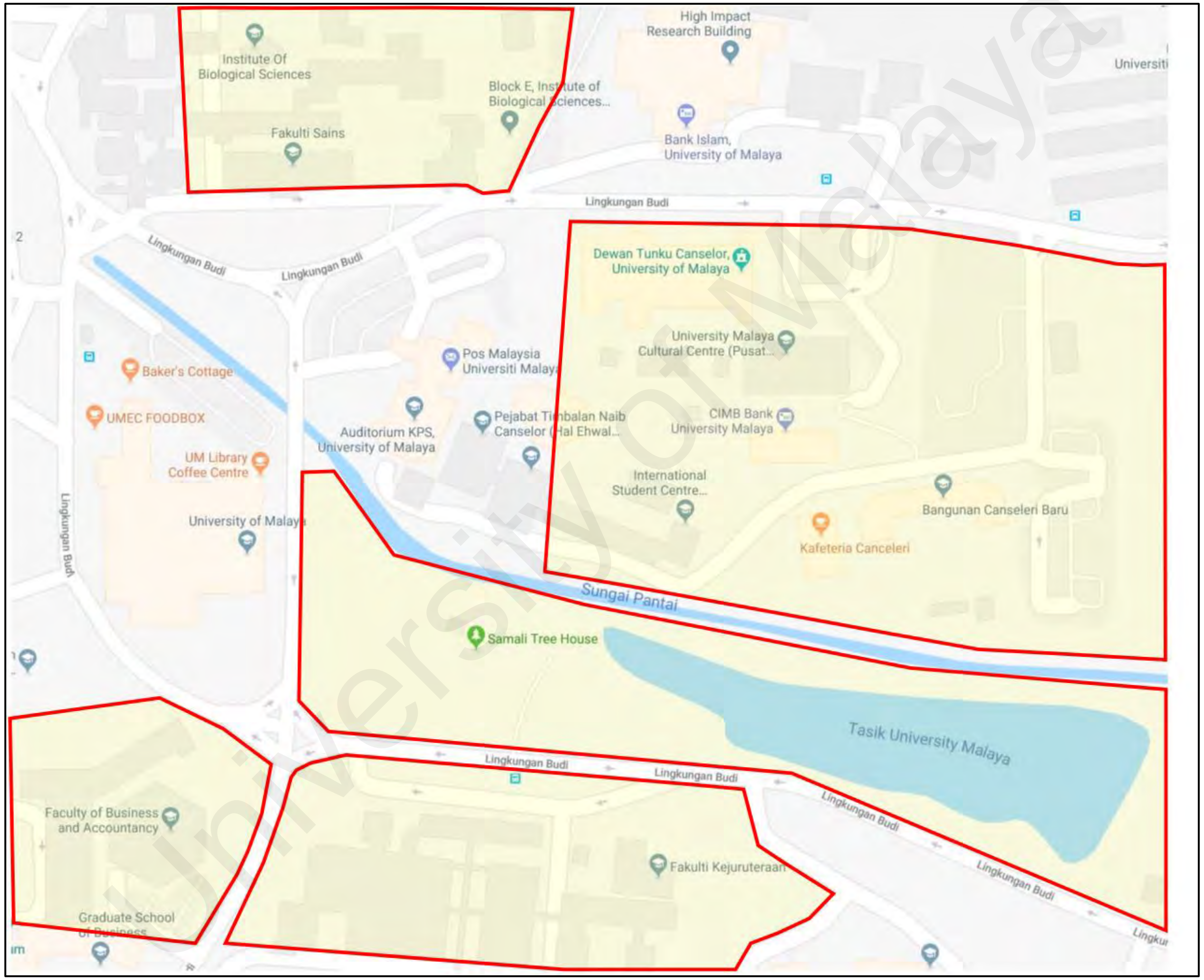


Figure 4.6: Locations of plant sampling in UM (courtesy of Google inc.)

4.5.2 Data Acquisition

Two types of plant, namely tree and shrub are used in this research. Tree and shrub are almost looking alike but in general, both types of plant can be distinguished where a matured tree is a woody plant that consists of one perennial stem or trunk. Meanwhile, shrub is a woody plant with a few of perennial stems and is usually smaller in size than tree (Cullina, 2002; Harris et al., 2003; Jones & Wofford, 2013). The plant species and their samples belong to these plant types are collected and acquired, in the forms of textual and image.

Plant samples are collected from four locations in UM such which are Varsity Lake, Faculty of Science, Faculty of Business and Accountancy, Faculty Engineering, Dewan Tunku Canselor, and Main Library (see Figure 4.6). For instance, Figure 4.7 shows the location in UM for the sample of *Saraca thaipingensis.* Images of plant samples are acquired using camera Nikon DSLR D750. Figure 4.8 shows an example of sample *S. thaipingensis* in which images of the plant species' fruit, leaves, flower, bark, and tree are taken. In addition, data of the geographical distribution of plant samples such as GPS coordinates and the name of location are collected. For example, the GPS coordinates for *S. thaipingensis* sample are 3.1198°N, 101.6558°E and located at Tasik Varsiti UM.

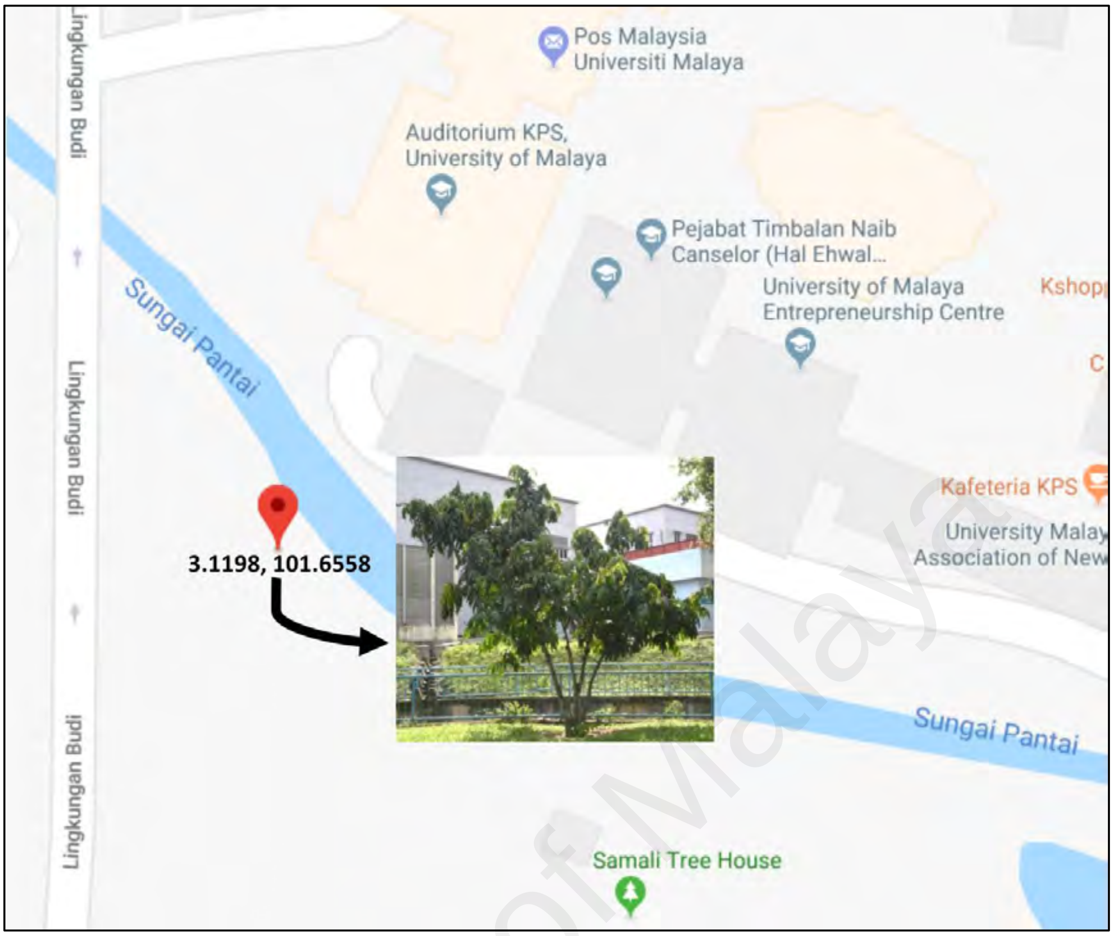

Pos Malaysia Universiti Malaya
Auditorium KPS, University of Malaya
Pejabat Timbalan Naib Canselor (Hal Ehwal...
University of Malaya Entrepreneurship Centre
Kshop
Kafeteria KPS
University Malay Association of New
Sungai Pantai
Lingkungan Budi
3.1198, 101.6558
Sungai Pantai
Samali Tree House

Figure 4.7: Location of plant sample for *S. thaipingensis* species in UM (courtesy of Google inc.)

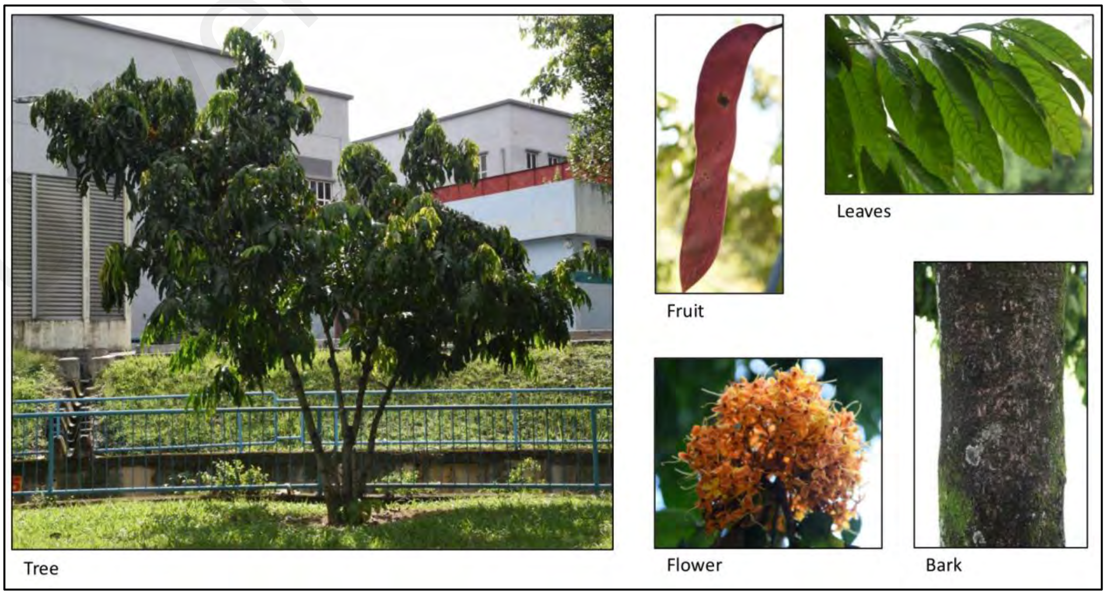
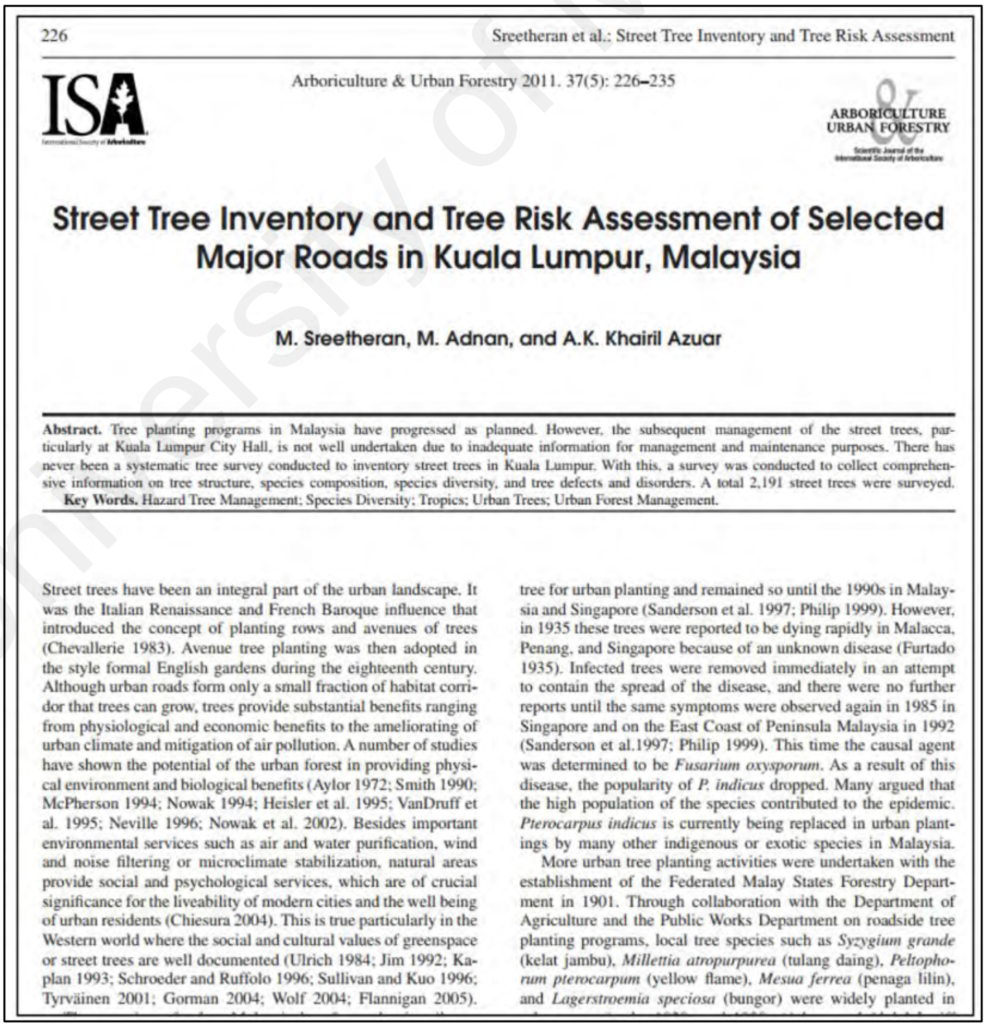


Figure 4.8: Plant sample images for *S. thaipingensis* species

Furthermore, species of the sample is identified with the help from the botanist. Information obtained is further extracted from manuscripts such as books (Boo et al., 2014; Gardner et al., 2011; Said et al., 2001) and journals as shown in Figure 4.9 (Sreetheran et al., 2011; Webb, 1998); and online public databases as shown in Figure 4.10 (Evergreen, 2017; FRIM, 2017; NParks Flora & Fauna Web, 2013; The Plant List, 2013). The reason of this process is because some of the plant samples are not natively grown. Thus, comparison of certain data such as characteristics of the plant species and its sample can be made. For instance, Table 4.2 shows the information of taxonomical classification and ecological attributes of *S. thaipingensis* while Table 4.3 shows morphological characteristics of *S. thaipingensis*.



226 Sreetheran et al.: Street Tree Inventory and Tree Risk Assessment

ISA
International Society of Arboriculture

Arboriculture & Urban Forestry 2011. 37(5): 226–235

ARBORICULTURE & URBAN FORESTRY
Scientific Journal of the International Society of Arboriculture

Street Tree Inventory and Tree Risk Assessment of Selected Major Roads in Kuala Lumpur, Malaysia

M. Sreetheran, M. Adnan, and A.K. Khairil Azuar

Abstract. Tree planting programs in Malaysia have progressed as planned. However, the subsequent management of the street trees, particularly at Kuala Lumpur City Hall, is not well undertaken due to inadequate information for management and maintenance purposes. There has never been a systematic tree survey conducted to inventory street trees in Kuala Lumpur. With this, a survey was conducted to collect comprehensive information on tree structure, species composition, species diversity, and tree defects and disorders. A total 2,191 street trees were surveyed.

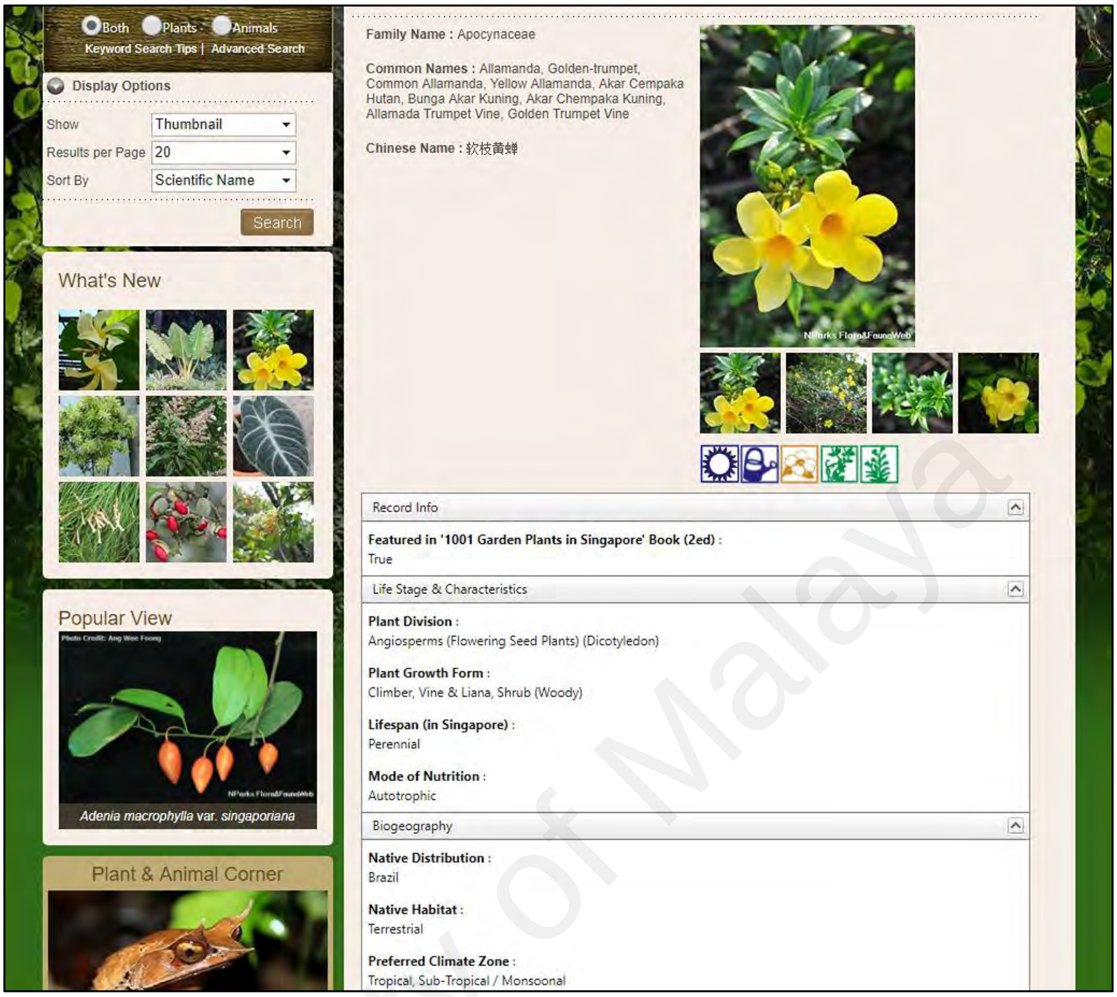
Key Words. Hazard Tree Management; Species Diversity; Tropics; Urban Trees; Urban Forest Management.

Street trees have been an integral part of the urban landscape. It was the Italian Renaissance and French Baroque influence that introduced the concept of planting rows and avenues of trees (Chevallerie 1983). Avenue tree planting was then adopted in the style formal English gardens during the eighteenth century. Although urban roads form only a small fraction of habitat corridor that trees can grow, trees provide substantial benefits ranging from physiological and economic benefits to the ameliorating of urban climate and mitigation of air pollution. A number of studies have shown the potential of the urban forest in providing physical environment and biological benefits (Aylor 1972; Smith 1990; McPherson 1994; Nowak 1994; Heisler et al. 1995; VanDruff et al. 1995; Neville 1996; Nowak et al. 2002). Besides important environmental services such as air and water purification, wind and noise filtering or microclimate stabilization, natural areas provide social and psychological services, which are of crucial significance for the liveability of modern cities and the well being of urban residents (Chiesura 2004). This is true particularly in the Western world where the social and cultural values of greenspace or street trees are well documented (Ulrich 1984; Jim 1992; Kaplan 1993; Schroeder and Ruffolo 1996; Sullivan and Kuo 1996; Tyrväinen 2001; Gorman 2004; Wolf 2004; Flannigan 2005).

tree for urban planting and remained so until the 1990s in Malaysia and Singapore (Sanderson et al. 1997; Philip 1999). However, in 1935 these trees were reported to be dying rapidly in Malacca, Penang, and Singapore because of an unknown disease (Furtado 1935). Infected trees were removed immediately in an attempt to contain the spread of the disease, and there were no further reports until the same symptoms were observed again in 1985 in Singapore and on the East Coast of Peninsula Malaysia in 1992 (Sanderson et al.1997; Philip 1999). This time the causal agent was determined to be *Fusarium oxysporum*. As a result of this disease, the popularity of *P. indicus* dropped. Many argued that the high population of the species contributed to the epidemic. *Pterocarpus indicus* is currently being replaced in urban plantings by many other indigenous or exotic species in Malaysia.

More urban tree planting activities were undertaken with the establishment of the Federated Malay States Forestry Department in 1901. Through collaboration with the Department of Agriculture and the Public Works Department on roadside tree planting programs, local tree species such as *Syzygium grande* (kelat jambu), *Millettia atropurpurea* (tulang daing), *Peltophorum pterocarpum* (yellow flame), *Mesua ferrea* (penaga lilin), and *Lagerstroemia speciosa* (bungor) were widely planted in

Figure 4.9: Example of journal containing plant data retrieved from Sreetheran et al. (2011)


○ Both ○ Plants ○ Animals
Keyword Search Tips | Advanced Search
Display Options
Search
What's New
Popular View
Photo Credit: Ang Wee Foong
NParks Flora&FaunaWeb
Adenia macrophylla var. singaporiana
Plant & Animal Corner
NParks Flora&FaunaWeb

Show: Thumbnail

Results per Page: 20

Sort By: Scientific Name

Family Name : Apocynaceae

Common Names : Allamanda, Golden-trumpet, Common Allamanda, Yellow Allamanda, Akar Cempaka Hutan, Bunga Akar Kuning, Akar Chempaka Kuning, Allamada Trumpet Vine, Golden Trumpet Vine

Chinese Name : 软枝黄蝉

Record Info

Featured in '1001 Garden Plants in Singapore' Book (2ed) :
True

Life Stage & Characteristics

Plant Division :
Angiosperms (Flowering Seed Plants) (Dicotyledon)

Plant Growth Form :
Climber, Vine & Liana, Shrub (Woody)

Lifespan (in Singapore) :
Perennial

Mode of Nutrition :
Autotrophic

Biogeography

Native Distribution :
Brazil

Native Habitat :
Terrestrial

Preferred Climate Zone :
Tropical, Sub-Tropical / Monsoonal

Figure 4.10: Example of online database containing plant data retrieved from NParks Flora & Fauna Web (2018)

Table 4.2: Taxonomical classification and ecological attributes of *S. thaipingensis* species

Type of data	Data	Data value
Taxonomical classification	Scientific name	Saraca thaipingensis
	Common name	Bunga Asoka
	Authorship	Cantley, Nathaniel
	Year published	1897
	Species	Saraca thaipingensis
	Genus	Saraca
	Family	Fabaceae
	Order	Fabales
	Class	Magnoliopsida
	Phylum	Tracheophyta
	Kingdom	Plantae
Ecological attributes	Habitat type	Along bank of rocky streams or on dry grounds
	Water usage	Moderate
	Type of soil	Waterlogged soils

Table 4.3: Morphological characteristics of *S. thaipingensis* species

Part of plant	Characteristics	Data value
Leaf	Type	Compound
	Shape	Oblong
	Venation	Pinnate
	Arrangement	Alternate
	Margin	Entire
	Tip	Mucronate
	Base	Acute
	Width	8 cm
	Length	23 cm
	Surface	Smooth
Fruit	Colour	Purple
Flower	Colour	Orange
	Inflorescence type	Corymb
	Petal number	Petalless
Bark	Surface	Smooth
Tree	Height	7 - 20 m

4.5.3 Image Database

Images of plant samples that are obtained during data acquisition process are stored in a local server. To encourage uniformity and consistency of the data for future work, a set of rules is set in naming the instances of these images as the names should be reflective of the data they represent in the domain. Table 4.4 lists a set of rules in naming each instance of plant samples' images. The naming of the images starts with the scientific name of the plant species, followed by type of sampling, parts of plant, number of the plant sample, and optional naming for object of image in frame and image of compound leaves. For an example, based on the set of rules, an image of a leaf in full frame from plant species *Delonix regia* is named as 'DELREGO-L001-FF'. Besides that, all images of plant samples are stored in JPEG format as it is viewable by all internet browsers as well as to reduce the storage size of the system.

Table 4.4: Set of rules in naming the plant samples' images

Order	Naming	Description	Example
1	Scientific name	The first three alphabet of the scientific name in uppercase	Delonix regia → e.g. ***DELREG***
2	Type of sampling	The type of sampling done to obtain the image	If from outdoor sampling, e.g. ***DELREGO***

3	Part of plant	Abbreviation of part of plant	Flower = R
			Bark = B
			Leaf = L
			Whole tree = W
			e.g. ***DELREGO-B***

4	Sample number	Starts with 001	e.g. ***DELREGO-B001***
5	Object frame	Whether the object of the image is in full frame or not.	Full frame = FF Not full frame = NF e.g. ***DELREGO-L001-FF***
6	Compound image	If image of compound type leaves is shown as the whole or a single unit of compound leaves	Whole compound = CW Single unit = CS e.g. ***SWIMACO-L001-CW***

4.5.4 Conceptual Framework of the Proposed Ontology

In this research, plant data are annotated in the form of ontology. Plant species data in textual are obtained from literatures, while plant samples data in textual and image are attained based on observations made as illustrated in Figures 4.5 and 4.6. Based on plant data description and relationships as explained previously in Section 4.4.1, the structured vocabularies are defined to represent the data as described in Section 4.4.4.1. These structured vocabularies are then used to represent the conceptualization of the data as described in Section 4.4.4.2.

4.5.4.1 Structured vocabularies

To avoid any misunderstanding in vocabularies in describing the plant data, a set of standardized vocabulary is designed in this research whereby the vocabularies are adapted from existing schema which is the Biodiversity Information Standards (TDWG,

2018) that consists of a number of specific biodiversity data standards such as Life Sciences Identifier (LSID), Darwin Core, and TDWG Access Protocol for Information Retrieval (TAPIR). Besides that, there are also a few newly defined vocabularies. These standardized structured vocabularies represent the concepts, and concepts' properties and relationships so that the meaning of the data is accurate and explicit to ensure same understanding and data sharing among users as well as maximizing the reusability in a wide range of contexts.

i) Defining the concepts

There are different approaches in determining the concepts such as top-down approach, bottom-up approach, and middle-out approach (Fernández-López, 1999; Hare et al., 2006; Uschold & Gruninger, 1996). Top-down approach starts from the most familiar concepts in the domain to the consequent distinct concepts (Prieto-Diaz, 2003). Bottom-up approach begins from the most specific concepts to more general concepts (Grewe et al., 2011). Middle-out approach is the combination of the previous two approaches where it starts from a number of concepts and proceeds to higher and lower level of these concepts (Sure et al., 2004; Uschold & Gruninger, 1996). Top-down approach may not include every general term which leads to inaccuracy (Vet & Mars, 1998) and bottom-up approach may produce too many excessive terms (Uschold & Gruninger, 1996; Zhou, 2007). Middle-out approach, however, balances in specifying details of each level. Only necessary details are obtained when each concept is expanded (El Ghosh et al., 2016; Uschold & Gruninger, 1996). In addition to that, it describes each concept in two ways; to higher and lower level at the same time while other approaches only work in one way in which there is a high possibility of missing out important concepts and definitions (El Ghosh et al., 2016;

Francesconi et al., 2010). Thus in this research, middle-out approach is used to define concepts of the domain.

Nine main concepts are described from the plant data used in this research; ‘PlantSample’, ‘Species’, ‘TaxonRank’, ‘PublicationCitation’, ‘Habitat’, ‘Distribution’, ‘Parts’, ‘ImgProperties’, and ‘Description’. Five sub-concepts are also described which are ‘Flower’, ‘Leaf’, ‘Bark’, ‘Fruit’, ‘Whole’. ‘PlantSample’ represents the sample of plant species. ‘Species’ represents the scientific name of plant species. ‘TaxonRank’ represents the taxonomical rank of the scientific name and its hierarchy. ‘PublicationCitation’ represents the source of the plant data collected from online databases or articles. ‘Habitat’ represents the details of the plant species natural environment area. ‘Distribution’ represents the information of the plant samples’ location in UM. ‘Parts’ represents description of the parts of plant species. ‘ImgProperties’ represents the basic attributes of the images of the plant samples. ‘Description’ represents the details of the plant samples. As for each sub-concept, it describes the parts of the plant. All sub-concepts are placed under main concepts ‘Parts’ and ‘Description’.

ii) Defining the concepts’ properties and relationships

In this process, properties and relationships are determined to bind all concepts together and to ensure ontology can describe relationships between data accurately. There are two types of properties for the semantic representation which are object and datatype properties. Both properties help in expressing the definition and flow in the database. Object property is the relationships between two individuals of different classes. It links an individual to another individual.

Datatype property is the relationships between an individual in the class and its data values.

The details of the properties and relationships of the nine concepts are as shown in Appendix A.

4.5.4.2 Proposed ontology schema

Plant data description and relationships as illustrated in Figure 4.5 is then translated into proposed ontology schema as shown in Figure 4.11, the ontology in a graph format. The oval shape represents the entity (the concept / class), the square represents the data value and the line represents the property and relationship (object properties and datatype property). Lines with arrowhead are the object properties that connect between the concepts while solid lines link concepts to their datatype properties that have data value. In the semantics context, this graph can be represented in form of triple statement (subject – predicate – object) that describes the relationships of the data.

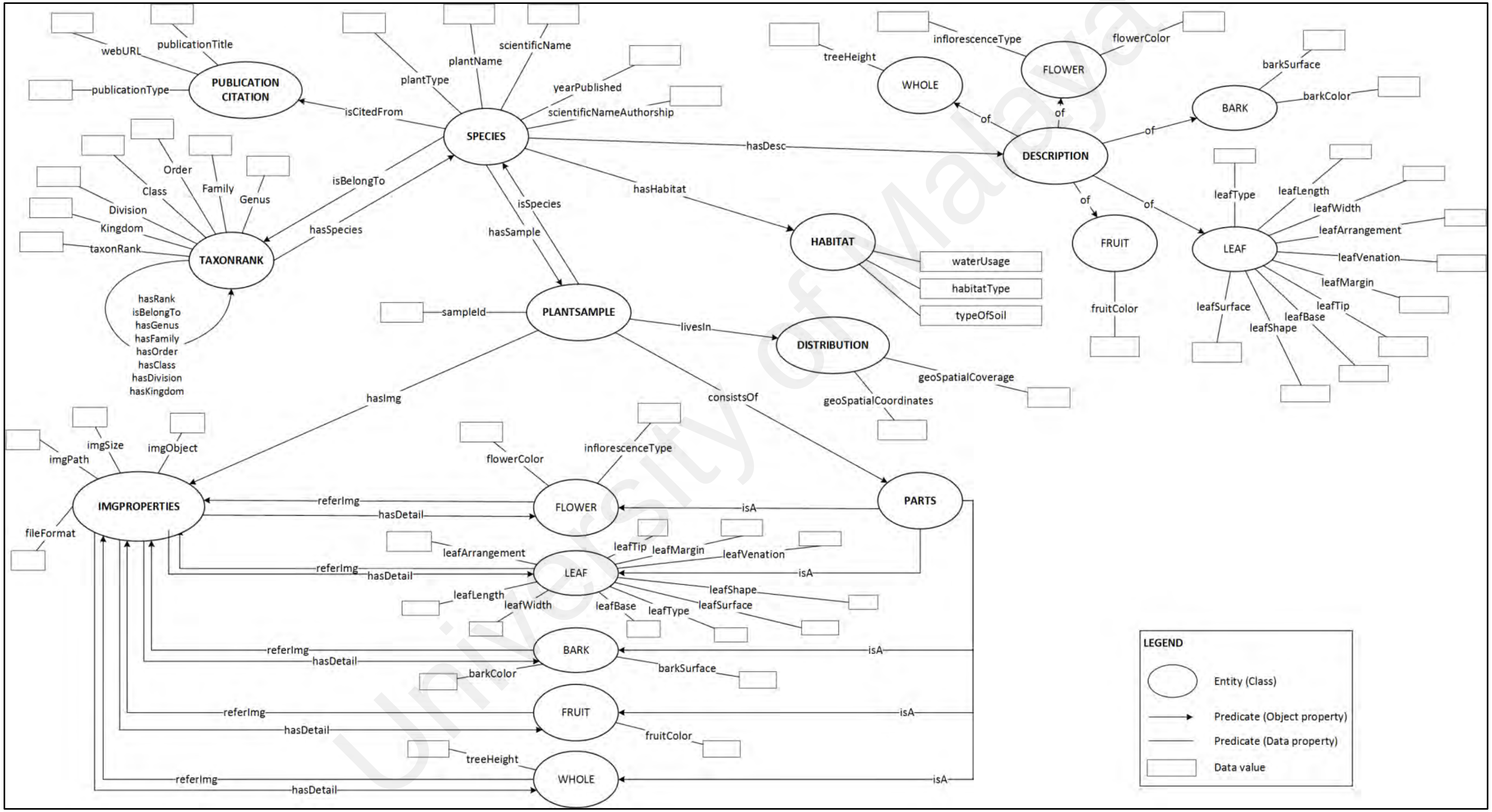


Figure 4.11: The proposed ontology schema in a graph format

Figure 4.12 exemplifies the triple statement formation. Instance ‘SDelReg001’ is the subject, ‘isSpecies’ is the predicate, and ‘DelReg_ScName’ is the object, thus in triple statement it will be written as ‘SDelReg001 isSpecies DelReg_ScName’. Predicates link concepts and concept to data values together. Another example is predicate ‘scientificName’ links the instance ‘DelReg_ScName’ to a data value of ‘Delonix regia’.

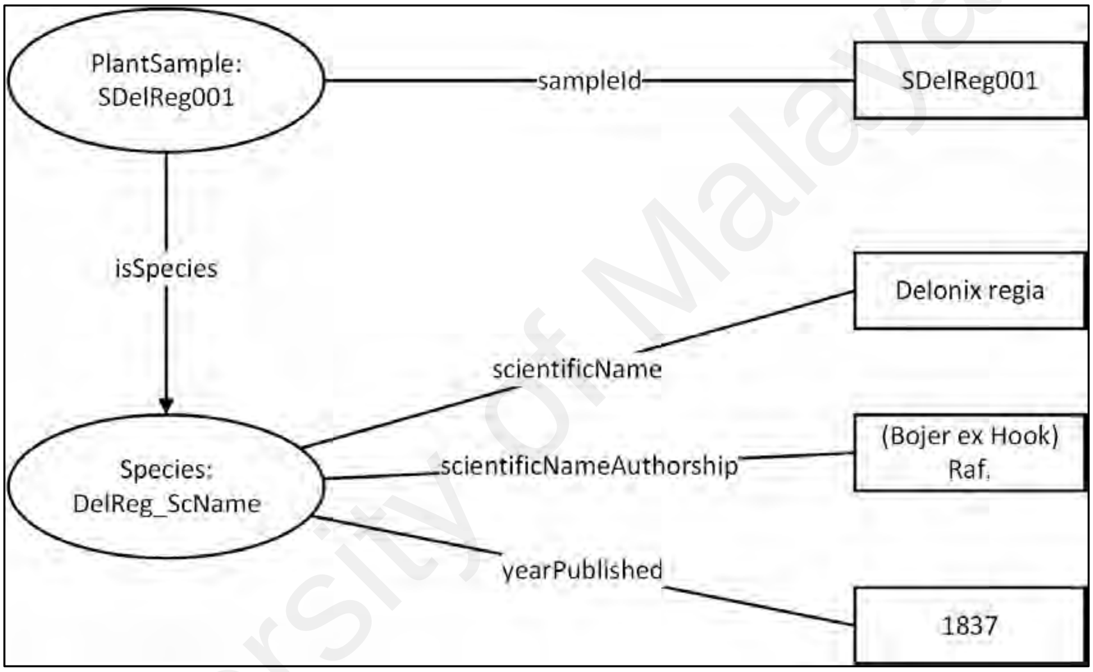


Figure 4.12: Example of triple statement formation

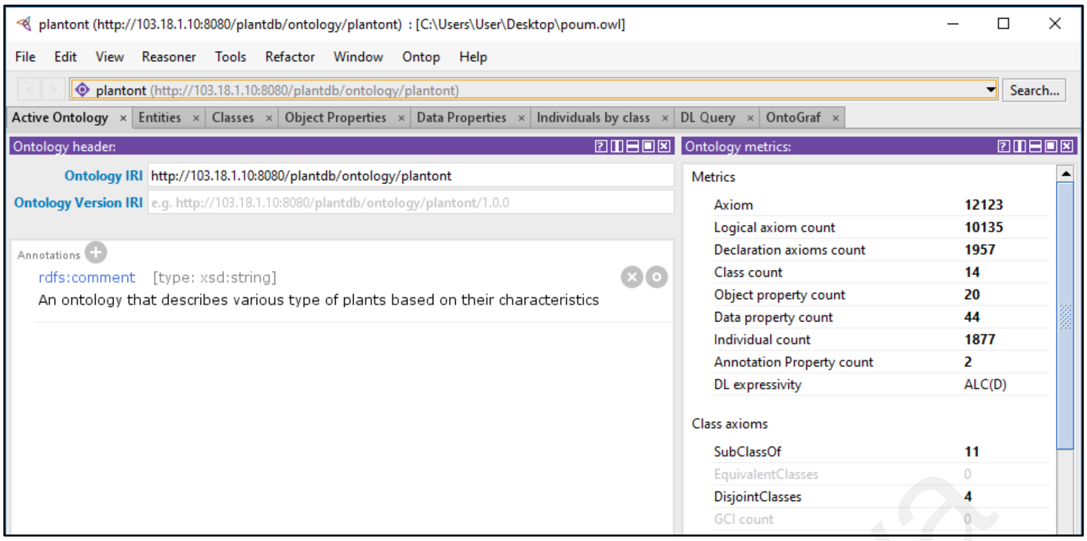
The ontology schema is then converted to ontological form so that it is in a machine-readable format specification and preceded with reasoning process to complete the process of defining and describing the data using the ontology editor software and the reasoning plug-ins.

Ontology is a formal way to describe classification networks and essentially defines the structure of domain knowledge. Ontology uses a formal language representation that is known as Web Ontology (OWL) (Bock et al., 2012). The OWL languages are characterized by formal semantics. It is designed for applications that need to process the content of the information instead of merely presenting the information. OWL is built based on World Wide Web Consortium (W3C) XML standard for objects called the Resource Description Framework (RDF). OWL facilitates greater machine interpretability of Web content that are supported by XML, RDF, and RDF Schema (RDFS) by providing additional vocabularies along with a formal semantics. OWL can also represent information about the objects themselves; the sort of information usually perceived as a data.

4.5.5 Ontology Development

The semantic representation of the plant data in the proposed ontology schema as illustrated previously in Figure 4.11 is then converted into ontological form using the ontology editor software, Protégé. Through this ontology which is in a machine-readable format specification, the computer is able to interpret the triple statements for data retrieval.

Using the wizard in Protégé, the ontology file is created with the internationalized resource identifier (IRI) as http://103.18.1.10:8080/plantdb/ontology/plantont, in owl file format named *poum.owl* and in RDF/XML serialization format for the triple statements, as shown in Figure 4.13.



Ontology IRI: http://103.18.1.10:8080/plantdb/ontology/plantont

Ontology Version IRI: e.g. http://103.18.1.10:8080/plantdb/ontology/plantont/1.0.0

rdfs:comment [type: xsd:string]

An ontology that describes various type of plants based on their characteristics

Metrics	
Axiom	12123
Logical axiom count	10135
Declaration axioms count	1957
Class count	14
Object property count	20
Data property count	44
Individual count	1877
Annotation Property count	2
DL expressivity	ALC(D)
Class axioms	
SubClassOf	11
EquivalentClasses	0
DisjointClasses	4
GCI count	0

Figure 4.13: Ontology IRI

The concepts and properties as described in Section 4.4.4.1 and Appendix A are then defined as formal structured vocabularies in the ontology. In Protégé 5.2, concept is known as class, while the object and datatype properties of each concept are defined as object property and data property, respectively. The following steps show the processes taken in adding the classes and properties into the ontology.

As shown in Figure 4.14, new class is added by clicking on the **Classes** tab, and clicking on **Add Subclass** button in the **Class Hierarchy** tab.

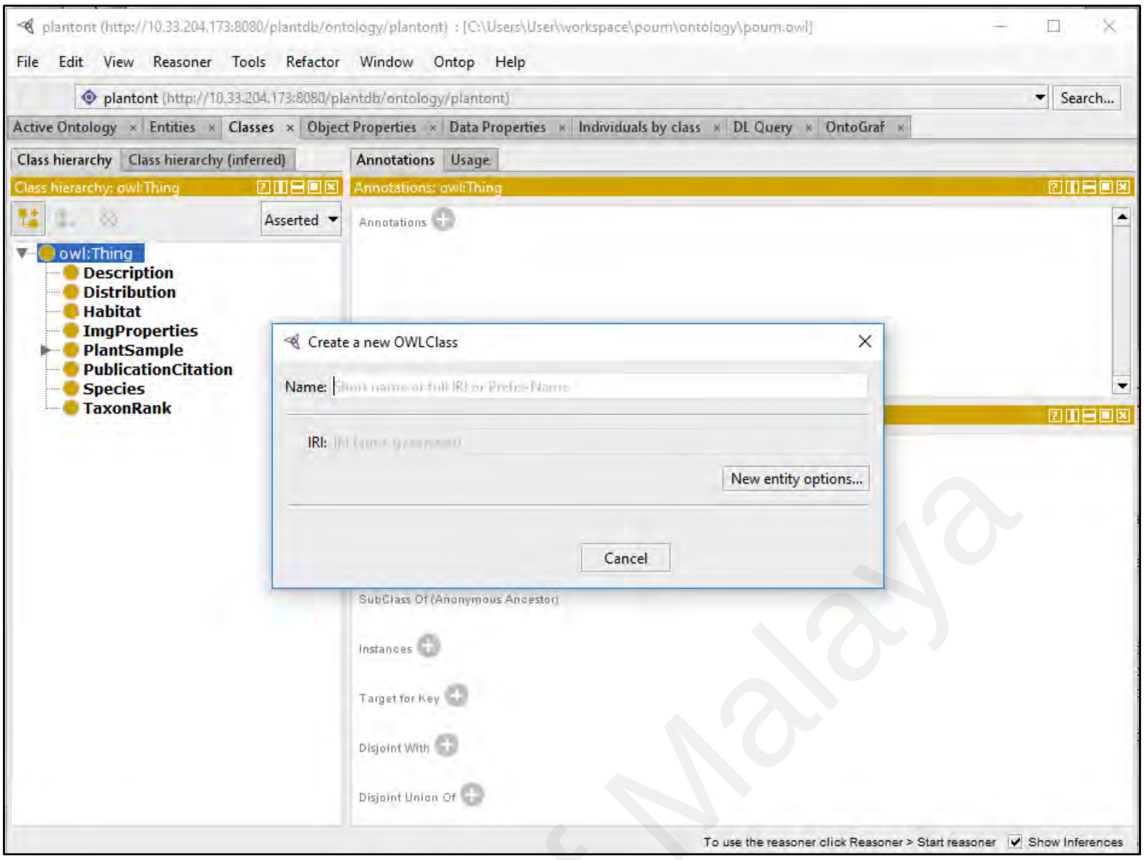

plantont (http://10.33.204.173:8080/plantdb/ontology/plantont) : [C:\Users\User\workspace\poum\ontology\poum.owl]
File Edit View Reasoner Tools Refactor Window Ontop Help
plantont (http://10.33.204.173:8080/plantdb/ontology/plantont)
Search...
Active Ontology Entities Classes Object Properties Data Properties Individuals by class DL Query OntoGraf
Class hierarchy Class hierarchy (inferred)
Class hierarchy: owl:Thing
Asserted
owl:Thing
Description
Distribution
Habitat
ImgProperties
PlantSample
PublicationCitation
Species
TaxonRank
Annotations Usage
Annotations: owl:Thing
Annotations
Create a new OWLClass
Name:
IRI:
New entity options...
Cancel
SubClass Of (Anonymous Ancestor)
Instances
Target for Key
Disjoint With
Disjoint Union Of
To use the reasoner click Reasoner > Start reasoner
☑ Show Inferences

Figure 4.14: Adding a new class into the ontology

Next is adding the properties into the ontology. To create an object property, under the **Object Properties** tab, click on the **Add Sub Property** button under the **Object Property Hierarchy** tab that is located on the left side of the interface as shown in Figure 4.15. Figure 4.16 shows how to create a data property; under the **Data Properties** tab, click on the **Add Sub Property** button under the **Data Property Hierarchy** tab that is located on the left side of the interface.

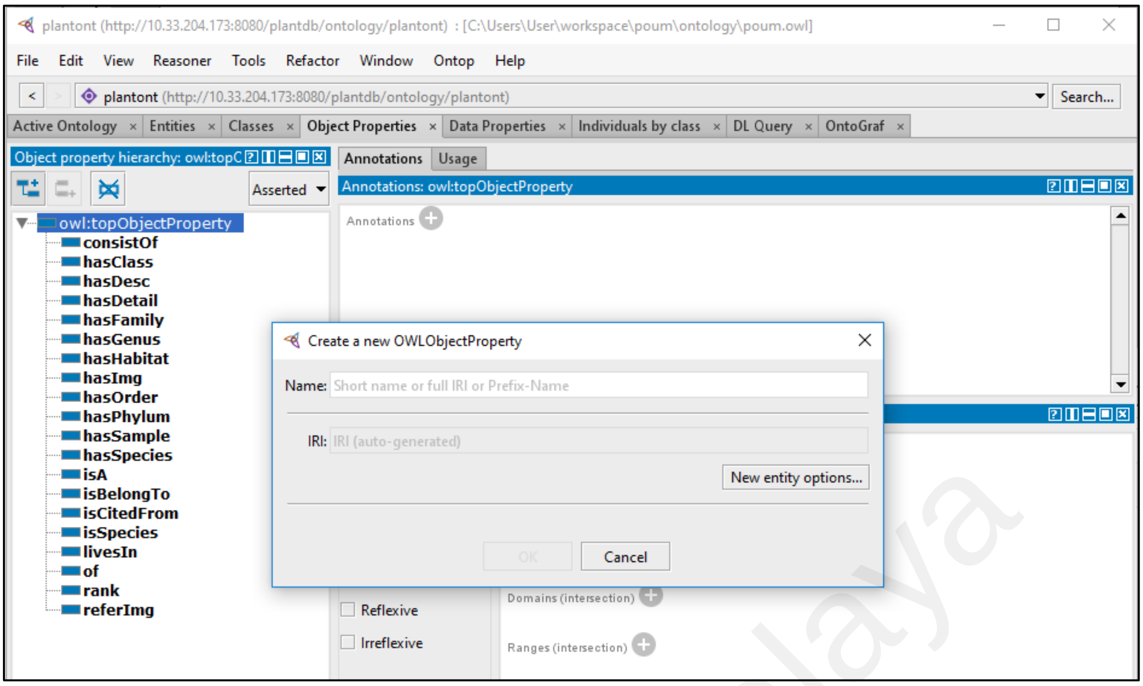

plantont (http://10.33.204.173:8080/plantdb/ontology/plantont) : [C:\Users\User\workspace\poum\ontology\poum.owl]
File Edit View Reasoner Tools Refactor Window Ontop Help
plantont (http://10.33.204.173:8080/plantdb/ontology/plantont)
Search...
Active Ontology × Entities × Classes × Object Properties × Data Properties × Individuals by class × DL Query × OntoGraf ×
Object property hierarchy: owl:topC
Annotations Usage
Annotations: owl:topObjectProperty
Asserted
Annotations
owl:topObjectProperty
consistOf
hasClass
hasDesc
hasDetail
hasFamily
hasGenus
hasHabitat
hasImg
hasOrder
hasPhylum
hasSample
hasSpecies
isA
isBelongTo
isCitedFrom
isSpecies
livesIn
of
rank
referImg
Create a new OWLObjectProperty
Name: Short name or full IRI or Prefix-Name
IRI: IRI (auto-generated)
New entity options...
OK
Cancel
Reflexive
Irreflexive
Domains (intersection)
Ranges (intersection)

Figure 4.15: Creating the object property

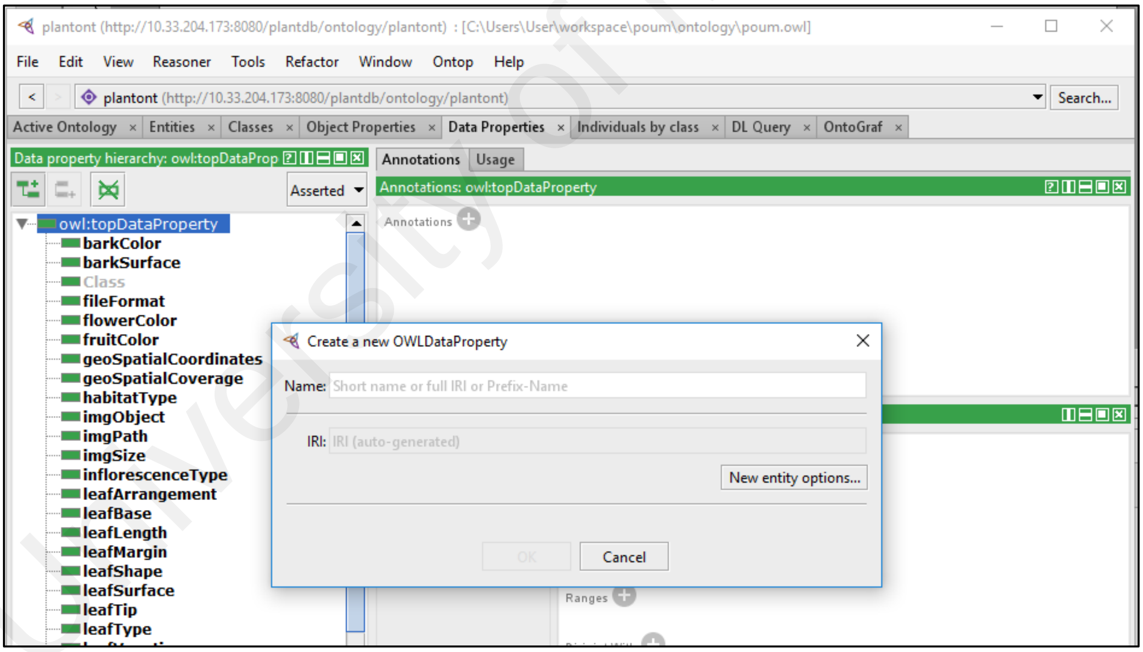

plantont (http://10.33.204.173:8080/plantdb/ontology/plantont) : [C:\Users\User\workspace\poum\ontology\poum.owl]
File Edit View Reasoner Tools Refactor Window Ontop Help
plantont (http://10.33.204.173:8080/plantdb/ontology/plantont)
Search...
Active Ontology × Entities × Classes × Object Properties × Data Properties × Individuals by class × DL Query × OntoGraf ×
Data property hierarchy: owl:topDataProp
Annotations Usage
Annotations: owl:topDataProperty
Asserted
Annotations
owl:topDataProperty
barkColor
barkSurface
Class
fileFormat
flowerColor
fruitColor
geoSpatialCoordinates
geoSpatialCoverage
habitatType
imgObject
imgPath
imgSize
inflorescenceType
leafArrangement
leafBase
leafLength
leafMargin
leafShape
leafSurface
leafTip
leafType
Create a new OWLDataProperty
Name: Short name or full IRI or Prefix-Name
IRI: IRI (auto-generated)
New entity options...
OK
Cancel
Ranges

Figure 4.16: Creating the data property

Description of all classes and properties are presented in Appendix B.

4.5.6 Data Annotation

Data collected during data acquisition activity are annotated in the form of instances. As mentioned before, consistency and standardizing the naming of instances are important to ensure there is no confusion in the future. A set of rules is defined to name all instances of plant data to help not only the developer of the ontology but also to enable users to add their own data. Moreover, by setting a set of rules in naming instances, it allows data sharing with other users in the community which encourage the extension of the data in the ontology. Table 4.5 presents the set of rules applied in naming instances of all concepts. Meanwhile, images obtained are annotated with the descriptions of the object in the images as well as observation taken during the data acquisition activity. Images are named according to the naming scheme as shown earlier in Table 4.4 to ensure uniformity and consistency which are useful for future work.

In the software Protégé 5.2, data of each concept are annotated in the form of instances. Instances are added and annotated with respective object properties and data properties. In general, adding new data involves creating new instance that belongs to the specific class, and annotating the instance with its properties. Table 4.6 shows an example of object and data properties for classes ‘PlantSample’, ‘Species’, ‘Distribution’, and ‘Habitat’ of plant species *Delonix regia*.

Table 4.5: Set of rules in naming the instance of each concept

Concept	Instance naming scheme		Example
	Instance	Description	
PlantSample	Sample	[“S”][first three alphabet of the genus and species name][numbering starts with 001]	SMurPan001
Description	Species’ Description	[first three alphabet of the genus and species name][underscore][”Desc”]	MurPan_Desc
Flower	Species’ Flower	[first three alphabet of the genus and species name][underscore][”Flower”]	MurPan_Flower
Leaf	Species’ Leaf	[first three alphabet of the genus and species name][underscore][”Leaf”]	MurPan_Leaf
Bark	Species’ Bark	[first three alphabet of the genus and species name][underscore][”Bark”]	MurPan_Bark
Fruit	Species’ Fruit	[first three alphabet of the genus and species name] underscore][”Fruit”]	MurPan_Fruit
Whole	Species’ Whole	[first three alphabet of the genus and species name][underscore][”Whole”]	MurPan_Whole
Species	Species	[first three alphabet of the genus and species name][underscore][“ScName”]	MurPan_ScName
Publication Citation	Data Source	[first three alphabet of the genus and species name][underscore][“Citation”]	MurPan_Citation
TaxonRank	Genus	[full name of the genus]	Murraya
	Family	[full name of the family]	Rutaceae
	Order	[full name of the order]	Sapindales
	Class	[full name of the class]	Magnoliopsida
	Phylum	[full name of the phylum]	Tracheophyta
	Kingdom	[full name of the kingdom]	Plantae

Table 4.5, continued

Concept	Instance naming scheme		Example
	Instance	Description	
Habitat	Habitat	[first three alphabet of the genus and species name] [numbering starts with 001][underscore][“Habitat”]	MurPan_Habitat
Distribution	Distribution	[“S”][first three alphabet of the genus and species name] [numbering starts with 001][underscore][“Distribution”]	SMurPan001_Distribution
Parts	Plant Sample’s Parts	[“S”][first three alphabet of the genus and species name] [numbering starts with 001][underscore][”Parts”]	SMurPan001_Parts
Parts	Plant Sample’s Parts	[“S”][first three alphabet of the genus and species name] [numbering starts with 001][underscore][”Parts”]	SMurPan001_Parts
Flower	Plant Sample’s Flower	[“S”][first three alphabet of the genus and species name] [numbering starts with 001][underscore][”Flower”]	SMurPan001_Flower
Leaf	Plant Sample’s Leaf	[“S”][first three alphabet of the genus and species name] [numbering starts with 001][underscore][”Leaf”]	SMurPan001_Leaf
Bark	Plant Sample’s Bark	[“S”][first three alphabet of the genus and species name] [numbering starts with 001][underscore][”Bark”]	SMurPan001_Bark
Fruit	Plant Sample’s Fruit	[“S”][first three alphabet of the genus and species name] [numbering starts with 001][underscore][”Fruit”]	SMurPan001_Fruit
Whole	Plant Sample’s Whole	[“S”][first three alphabet of the genus and species name] [numbering starts with 001][underscore][”Whole”]	SMurPan001_Whole
ImgProperties	All parts of plant	Follows the naming of the plant samples’ images	MURPANO-R001

Table 4.6: Example of object and data properties for plant species *Delonix regia*

Classes	Instances	Object Property	Data Property	Example
PlantSample	SDelReg001	isSpecies		DelReg_ScName
		livesIn		SDelReg001_Distribution
		consistOf		SDelReg001_Parts
		hasImg		DELREGO-L001-FF DELREGO-B001 DELREGO-T001 DELREGO-R001
			sampleId	SDelReg001
Species	DelReg_ScName	isCitedFrom		DelReg_Citation
		isBelongTo		Delonix
		hasDesc		DelReg_Desc
		hasHabitat		DelReg_Habitat
		hasSample		SDelReg001 SDelReg002 SDelReg003
			plantType	Tree
			plantName	Semarak Api
			scientificName	Delonix regia
			scientificNameAuthorship	(Bojer ex Hook) Raf.
			yearPublished	1837
Distribution	SDelReg001_Distribution		geoSpatialCoordinates	3.1184, 101.6587
			geoSpatialCoverage	Tasik Varsiti UM
Habitat	DelReg_Habitat		typeOfSoil	Loamy soil
			waterUsage	Low to moderate
			habitatType	Tropical, subtropical, monsoon forest

In Protégé, instances are known as individuals. To add new individuals, click on the **Individuals by Class** tab and choose the class under **Class Hierarchy** tab. Under the **Instances** tab, click on **Add Individual** button as shown in Figure 4.17. The processes involved for annotating the instances with their object and data properties are as shown in Figure 4.18 and Figure 4.19, respectively. To annotate an instance to an object property, click on the instance's name and click **Add** button beside **Object Property Assertions** list, then enter the object property name and individual name. To annotate an instance to a data property, click **Add** button beside **Data Property Assertion** list, choose the data property, enter the data value and set the type of data value and language.

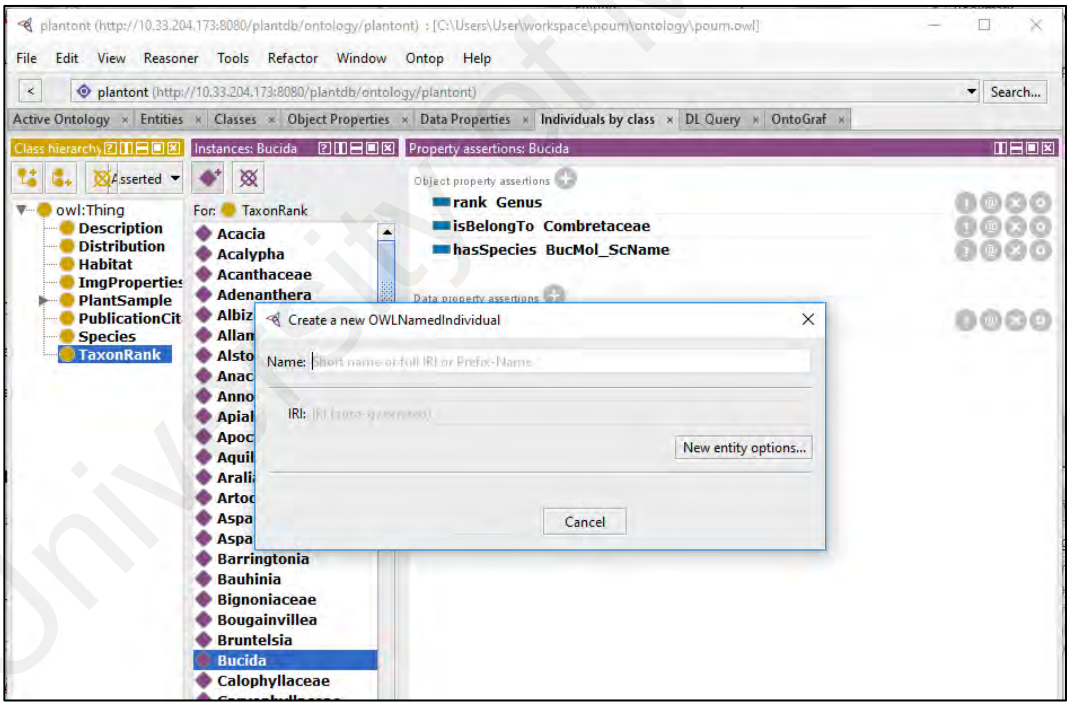
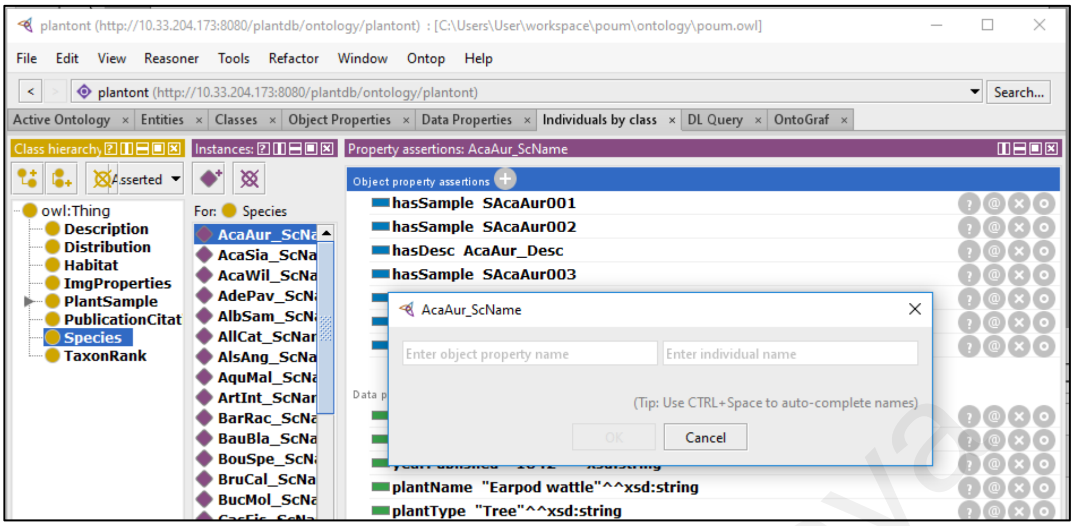
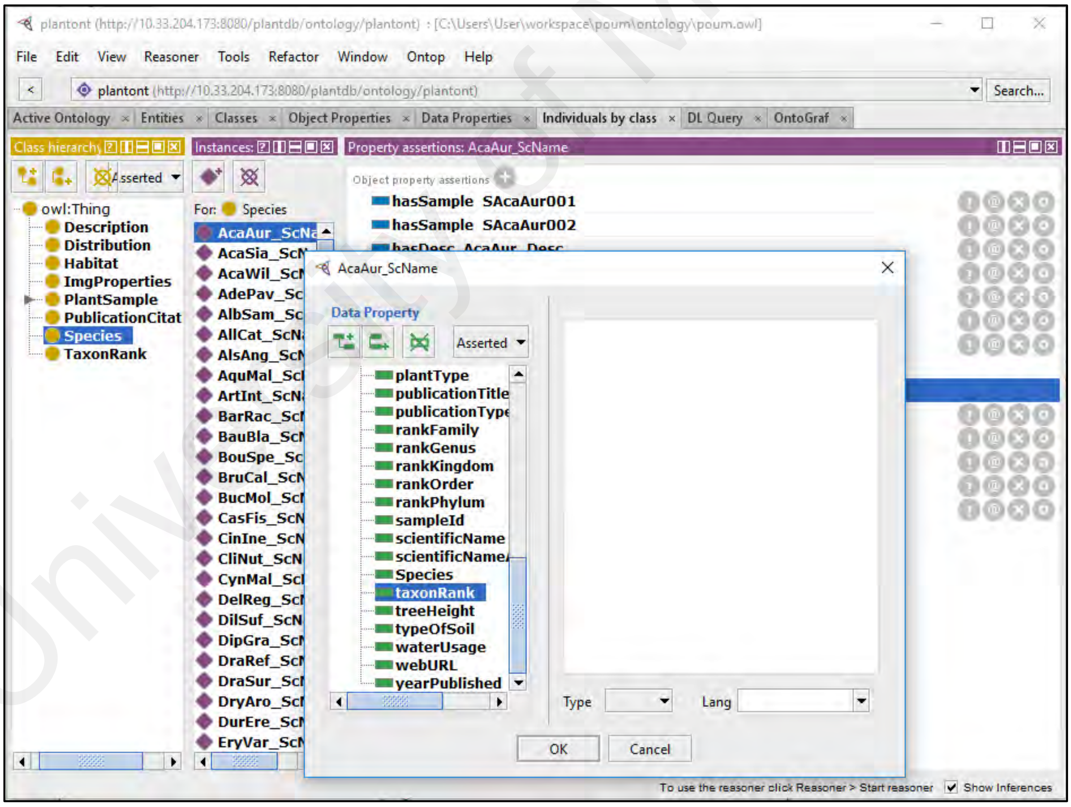

plantont (http://10.33.204.173:8080/plantdb/ontology/plantont) : [C:\Users\User\workspace\poum\ontology\poum.owl]
File Edit View Reasoner Tools Refactor Window Ontop Help
plantont (http://10.33.204.173:8080/plantdb/ontology/plantont)
Search...
Active Ontology × Entities × Classes × Object Properties × Data Properties × Individuals by class × DL Query × OntoGraf ×
Class hierarchy
Asserted
owl:Thing
Description
Distribution
Habitat
ImgProperties
PlantSample
PublicationCit
Species
TaxonRank
Instances: Bucida
For: TaxonRank
Acacia
Acalypha
Acanthaceae
Adenanthera
Albiz
Allan
Alsto
Anac
Anno
Apial
Apoc
Aquil
Arali
Artoc
Aspa
Aspa
Barringtonia
Bauhinia
Bignoniaceae
Bougainvillea
Bruntelsia
Bucida
Calophyllaceae
Property assertions: Bucida
Object property assertions
rank Genus
isBelongTo Combretaceae
hasSpecies BucMol_ScName
Data property assertions
Create a new OWLNamedIndividual
Name: Short name or full IRI or Prefix-Name
IRI: IRI (auto-generated)
New entity options...
Cancel

Figure 4.17: Adding a new instance to the specific class


plantont (http://10.33.204.173:8080/plantdb/ontology/plantont) : [C:\Users\User\workspace\poum\ontology\poum.owl]
File Edit View Reasoner Tools Refactor Window Ontop Help
plantont (http://10.33.204.173:8080/plantdb/ontology/plantont)
Search...
Active Ontology Entities Classes Object Properties Data Properties Individuals by class DL Query OntoGraf
Class hierarchy
Asserted
owl:Thing
Description
Distribution
Habitat
ImgProperties
PlantSample
PublicationCitat
Species
TaxonRank
Instances:
For: Species
AcaAur_ScNa
AcaSia_ScNa
AcaWil_ScNa
AdePav_ScN
AlbSam_ScN
AllCat_ScNar
AlsAng_ScNa
AquMal_ScNa
ArtInt_ScNar
BarRac_ScNa
BauBla_ScNa
BouSpe_ScN
BruCal_ScNa
BucMol_ScNa
Property assertions: AcaAur_ScName
Object property assertions
Data p
AcaAur_ScName
Enter object property name
Enter individual name
(Tip: Use CTRL+Space to auto-complete names)
OK
Cancel

hasSample SAcaAur001
hasSample SAcaAur002
hasDesc AcaAur_Desc
hasSample SAcaAur003
plantName "Earpod wattle"^^xsd:string
plantType "Tree"^^xsd:string

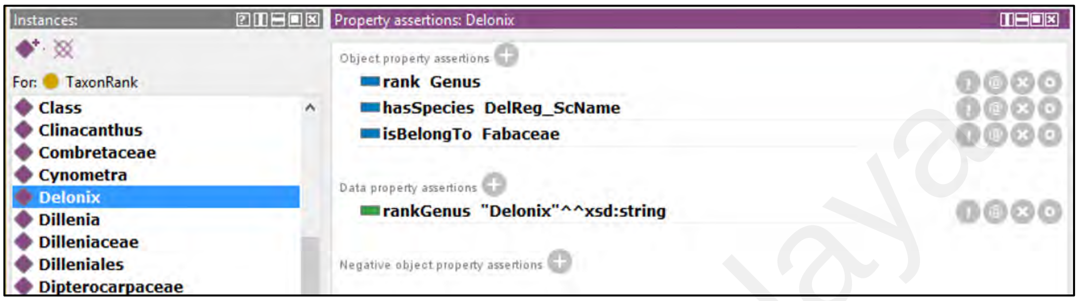
Figure 4.18: Annotating the object property to a specific instance



hasSample SAcaAur001
hasSample SAcaAur002
hasDesc AcaAur_Desc

Figure 4.19: Annotating the data property to a specific instance

Figure 4.20 shows an instance that is completely annotated with its object and data properties, based on the example mentioned earlier in Table 4.6. It shows that the **Delonix,** an instance of the class **TaxonRank**, annotated with object properties named **hasSpecies**, **rank**, and **isBelongTo**, and data property named **rankGenus**.



rank Genus
hasSpecies DelReg_ScName
isBelongTo Fabaceae
rankGenus "Delonix"^^xsd:string

Figure 4.20: An example, the Delonix, an instance of class TaxonRank is completely annotated with its object and data properties

4.5.7 Ontology Reasoning

Ontology is considered as one of the essential components of a system as it constructs the structure of the data. For this reason, it is important to design and maintain a solid ontology where it is meaningful to the domain and users and it has no unintentional synonym to minimize redundancy in the ontology (Chen, 2010). Reasoning process can prevent such matters from occurring. Ontology reasoning is a process of deducing facts not explicitly stated in the ontology (Koutsomitropoulos & Kalou, 2017). It infers logical consequences from axioms defined in ontology designing process. This process is done using a reasoning plug-in or commonly known as Reasoner which uses first-order predicate logic to perform reasoning.

4.6 Ontology-Based Data Retrieval and Visualisation Design and Development

This system design consists of data retrieval and visualisation. The data retrieval design is described in Section 4.6.1. Once the data are retrieved, the data are presented in a visualisation form and the processes design is defined in Section 4.6.2. Java programming language is chosen as the main language for the proposed system development.

4.6.1 Ontology-Based Data Retrieval Design

The data retrieval design in this research is using the classical Boolean search and SPARQL (W3C, 2013). Both object and datatype properties of the concepts from the proposed ontology (see Appendix A) are used as parameters to formulate the query and data retrieval for searching the patterns in the triple statements. This process is performed using the Java-based semantics framework that has built-in semantics libraries for creating the graph data model, querying the ontology, reasoning and inference.

4.6.2 Data Visualisation Design

The data visualisation design in this research is emphasized on the three types of relationships as described in Section 4.6.2.1. The selected type of data visualisation for this research, named network graph is described in Section 4.6.2.2 and Section 4.6.2.3 describes the tools used to build the network graph.

4.6.2.1 Types of plant data relationships

i) Relationships between one taxon to another taxon

A taxon is linked to another by their family name. As an example, *Delonix regia* and *Acacia auriculiformis* are linked to one another as both are in the same family of Fabaceae.

ii) Relationships between taxa and taxa's sample

Each taxon has three samples and each of the samples has a unique identifier. For instance, *Murraya paniculata* has three samples namely 'SMurPan001', 'SMurPan002', and 'SMurPan003'.

iii) Relationships between samples

Samples are related when they are from the same taxon or are obtained from the same location. For instance, samples of *Lagerstroemia indica*, *Manihot esculenta*, and *Terminalia catappa* are collected from the same location of 'DTC UM'.

4.6.2.2 Type of data visualisation

As described in Section 2.7, each types of data visualisation have different purposes whereby each type highlights a particular part of the data. In this research, network graph is chosen to visualize the relationships of plant data as mentioned above. Network graph is a type of graph that highlights the relationships between entities and consists of 'nodes' as entities and 'links' as lines to link between entities. In general, there are two types of network graph which are undirected network graph – illustrates the relationships between entities but not its direction; and directed network graph – shows directionality of relationships which is more meaningful (Teller, 2013).

Additionally, there are labels on the nodes and links as well as other additional information such as legends or expansion of the network graph. These depend on the specificity of the network graph.

This process is performed using the Java-based data visualisation tools which are supporting to create the node, link, label, legend and other interactive features such as to highlight the node's links, expand or shrink the group of nodes and thumbnails.

4.6.2.3 Tool for designing data visualisation

A good combination of functioning visualisation tool makes an effective data visualisation. A good visual library which consists of a set of programming languages helps in designing any kind of visualisation. A number of visualisation tools are reviewed beforehand based on the functionality offered by each tool. In this research, D3.js (Bostock, 2017) is chosen as the tool to develop the plant data visualisation. D3.js is a visual library based on JavaScript that can create dynamic and interactive data visualisation in web-based interface. It heavily utilizes CSS, HTML, and Scalable Vector Graphics (SVG) standards which provide controls to users over the final result.

JavaScript Object Notation (JSON) is a lightweight data format used in data-interchange. JSON works in attribute-value pairs, array data types, or in any form of serializable value. The flexibility of JSON allows the arrangement of the data in any manner and it does not restrict the type of data that can be used as JSON. JSON is independent from other programming language, yet its text format is almost similar to other programming languages such as C, C++, and Python (Boci et al., 2012). In this research, ontology data in graph data model are serialized into a JSON format before further developed to a network graph. In addition, another tool is also involved in the process of converting ontological data to the JSON format which is Jackson; a set of data processing tools for Java that is used in this process to serialize Java object to JSON (FasterXML, 2018).

4.6.3 Implementation of Data Retrieval and Visualisation

Figure 4.21 shows the flow of data retrieval and visualisation which involved a few processes as explained in the followings.

i) Query page

The interface for the system is designed to allow users to communicate with the system. The code that implements the interface is as presented in Appendix C(i). The user interface is developed using HTML, CSS, JS, and JSP. The query interface contains a drop-down menu with four options namely 'Scientific Name', 'Family Name', 'Location', and 'Water Usage', a textbox to enter a data query, a set of radio button with the given data values, and a submit button.

ii) Query processing

Query entered in the user interface will be then sent to query processing; parameters selected by users will be used in querying from the ontology. The code that implements for this step is as shown in Appendix C(ii).

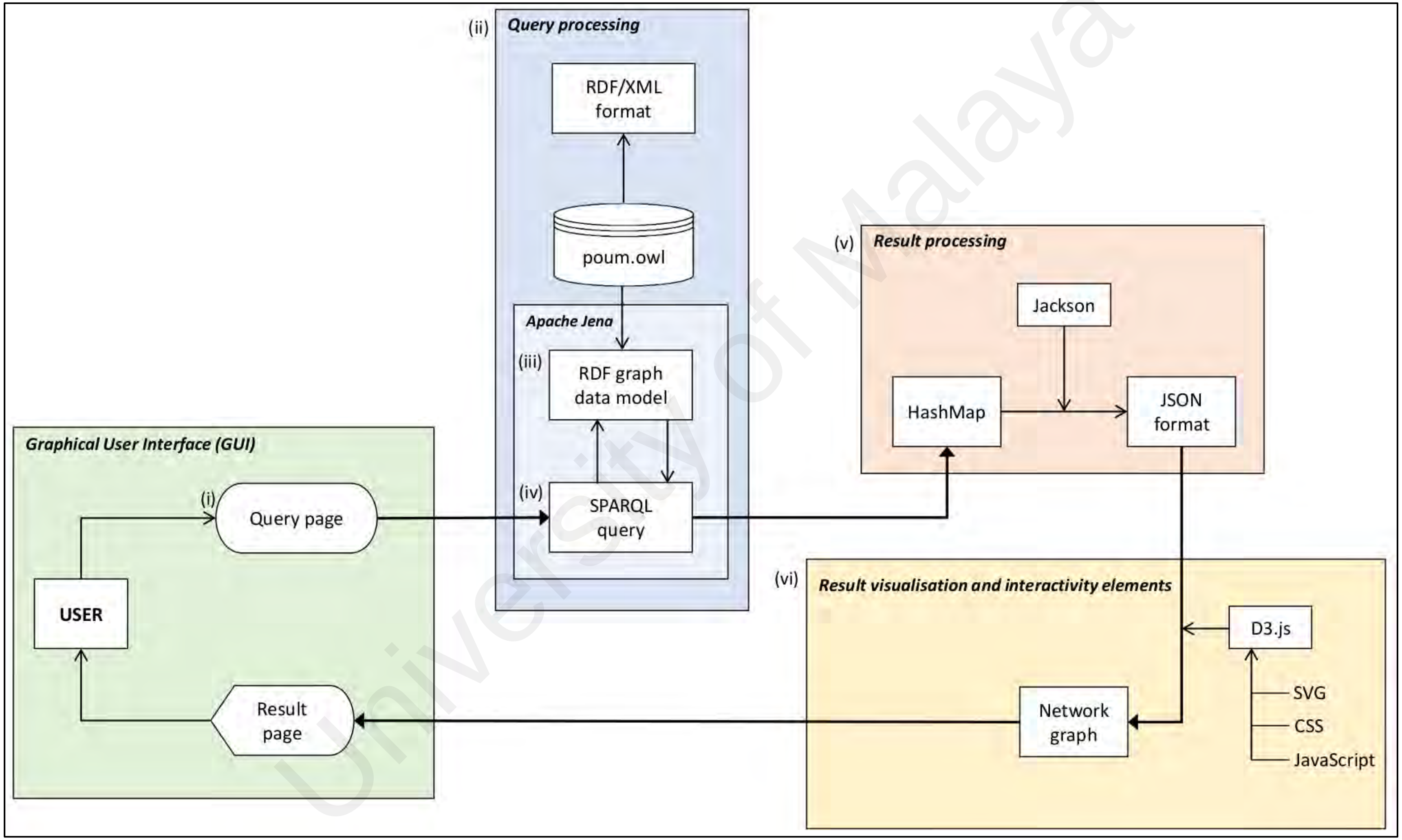


Figure 4.21: Process flow of data retrieval and visualisation

iii) Graph data

To query from the ontology, the ontological data which is in RDF/XML serialization format will be then converted into graph data model using *ModelFactory* class from Jena API. The graph data model is then stored and will be used for querying. Graph data model is stored temporarily in the computer memory, which means a new graph data model will be created every time querying process occurs. This is to ensure that the graph data model has the current ontological data. The code that implements the step is as shown in Appendix C(iii).

iv) SPARQL query

SPARQL is the query language in the Semantic Web where it can retrieve and manipulate data stored in RDF format. The advantage of SPARQL is that it allows users to write queries against data that follows the RDF specification. Once again Jena API is used where classes *QueryFactory* and *QueryExecution* are used to execute the query on the graph data. The code that implements this process is presented in Appendix C(iv).

v) Result processing for data retrieval

The result of SPARQL query is usually in textual form, which is passed as Java object and will be further processed to generate the graphical version of the result. The result is arranged into a set of 'HashMap' using Java *HashMap* class. A HashMap contains key-value pairs and can only have unique elements and not in orderly arrangement. Hence, it is suitable to be used as it is flexible to arrange the elements in any manner yet each element have their own identification for easy reference. The code that implements this process is as shown in Appendix C(v).

vi) Data visualisation

The structure of ‘HashMap’ is almost similar to the structure of JSON which makes it easier to deal with during the conversion to the JSON format. For this reason, Jackson is used to serialize ‘HashMap’ into JSON format and then it is passed to the D3.js library. The code that implements this process is as shown in Appendix C(vi).

D3.js utilizes different types of programming languages namely JavaScript, CSS and SVG in implementing the data visualisation. Besides that, it provides a wide range of data visualisation types and in this research, network graph is chosen to present the result to the users as it can illustrate the relationships between the data which is crucial as to highlight the data association in the knowledge domain.

4.6.4 Implementation of Interactive Features

In this process, data visualisation generated in the previous process is enhanced to provide a dynamic data visualisation to the users. Moreover, it encourages two-way communication between the users and data. Interactive features are added into the data visualisation. The code that implements these elements is as presented in Appendix C(vii).

4.7 Testing

In this research, two types of testing are performed; the ontology testing and the data visualisation system testing.

Ontology testing is done by the developer with the help of experts in botanical field. Meanwhile, the data visualisation system testing is done by 60 end users. The user evaluation is participated by 30 expert users with research background or experience in botanical field and may have little skill in IT field, and 30 novice users who have little or no research background or experience in botanical field but may have skills in IT field.

4.7.1 Ontology Testing

Ontology testing is performed to the querying process using SPARQL and the ontology evaluation. The purpose of testing the querying process is to ensure that the result of the query is accurate as SPARQL query is based on logic expression. Moreover, the purpose of ontology evaluation is to check the quality and practicality of the developed ontology. The ontology evaluation is adapted from (Gruber, 1993) evaluation's criteria which are *Clarity*, *Coherence*, and *Extendibility*; and (Gómez-Pérez, 1996) evaluation's criteria which are *Conciseness* and *Correctness*. The reason for choosing two different methodologies is to show that there is no biasedness and favouritism in selecting evaluation approach.

4.7.2 Data Visualisation System Testing

There is a variety of data visualisation tools available to help the developers to achieve the objective of data visualisation. Therefore, it is desirable to determine if the developed data visualisation system is successful in achieving users' needs.

Accordingly, the user evaluation is performed in which it involved two types of evaluation which are the usability heuristics evaluation, and query and visualisation evaluation. Example of the questionnaire given to the users is as shown in Appendix D.

a) Usability heuristics evaluation

Heuristic evaluation is adapted from Nielsen's 10 usability heuristics for UI design (Nielsen, 1992). These are the general principles for an interactive UI design. In this research, 10 usability heuristics are adapted to match with the developed data visualisation system. Users are given 5 minutes to explore the GUI before the evaluation. This step is performed to observe users' first impression on the visualisation system.

b) Query and visualisation evaluation

Query and visualisation evaluation assess the efficiency of the visualisation system in delivering the visualisation contents to the users (Amri et al., 2015; Hearst et al., 2016) to measure the performance of sending a query from GUI to the server and passing of the result to the visualisation. In addition, it is to observe how the users can interact with the data successfully using the interactive features in the visualisation.

Users are given guidelines and instructions on using PlantViz before performing the user evaluation. Query and visualisation are evaluated by rating of '1' to '5', in which '*1'-Poor, '2'-Fair, '3'-Average-, '4'-Good, '5'-Excellent*. Any comments from the users are taken into consideration to improve the developed visualisation system. There are four cases based on the search parameter and each is described in detail below. There are four cases based on the search parameter, which are *Scientific Name* (Case 1), *Family Name (*Case 2), *Location (*Case 3), and *Water Usage* (Case 4).

i) **Case 1: *Scientific Name***

In this case, 'Scientific Name' refers to a name used by scientists to identify an organism that consists of the genus and species. In this case, users were instructed to send query for scientific name *Albizia saman*, *Bruntelsia calycina*, and *Coloccasia esculenta*. Users were required to key in the scientific name.

ii) **Case 2: *Family Name***

In this case, 'Family Name' refers to the taxonomic rank after rank Genus. In this case, users were instructed to send queries for family Euphorbiaceae, Fabaceae, and Surianaceae. Users were required to key in the scientific name.

iii) **Case 3: *Location***

In this case, 'Location' refers to the location of the plant sample taken in UM. Users were instructed to send queries for locations 'DTC UM', 'Fakulti Kejuruteraan UM', 'Fakulti Perniagaan dan Perakaunan UM', 'Fakulti Sains UM', and 'Tasik Varsiti UM'. Users were required the location name from a dropdown list.

iv) **Case 4: *Water Usage***

In this case, 'Water Usage' refers to the amount of water in a plant's habitat. Users were instructed to send query by choosing from a dropdown list for 'Low', 'Low to moderate', 'Moderate to high', and 'High'.

Next, Fisher's Exact test and t-test are performed to analyse the outcome of the evaluation. Based on the usability heuristic evaluation, Fisher's Exact is conducted to check whether the GUI of the system is dependent on the users' knowledge in botanical and IT fields. Meanwhile, based on the query and visualisation evaluation, the t-test is conducted to check whether there is any significant difference between the evaluation done by expert users and novice users on the visualization system.

4.8 Summary

This chapter provides the overall system architecture of the proposed solution. In general, the proposed solution consists of database to store plant data in textual and image forms using ontology; the system for ontology-based data retrieval and visualisation; and the GUI for the users to communicate with the system. In addition, ontology testing and data visualisation testing are performed as the assessment of the system architecture.

CHAPTER 5: RESULT AND DISCUSSION

5.1 Introduction

Chapter 5 presents the results in implementing the proposed solution and its usability to the users. In this research, Plant database named **P**lant **O**ntology **U**niversiti **M**alaya (POUM) is developed whereby it consists of image database and ontological plant data. This POUM is used as dataset and integrated into data visualisation system named **Plant Vis**ualisation (PlantViz). The results are discussed further by evaluating their strengths and limitations, and compared with other existing systems. Besides that, results of testing on the ontology and data visualisation system are also presented and discussed further.

5.2 Plant Image Database

Plant image database consists of the images of tree and shrub samples that are collected from UM. There are 308 images of samples from 74 species which comprises the images of parts from plant such as bark, leaves, flower, fruit, and tree. The plant images are stored locally in the same directory of PlantViz system. Each image is named according to the naming scheme (see Table 4.4) and annotated accurately with the descriptions of the objects in the image as well as observation taken during the data acquisition activity using the structured vocabularies as defined in Appendix B. All images are 2D image and compressed into JPEG file format. Figure 5.1 shows a partial view of the images in the Plant image database.

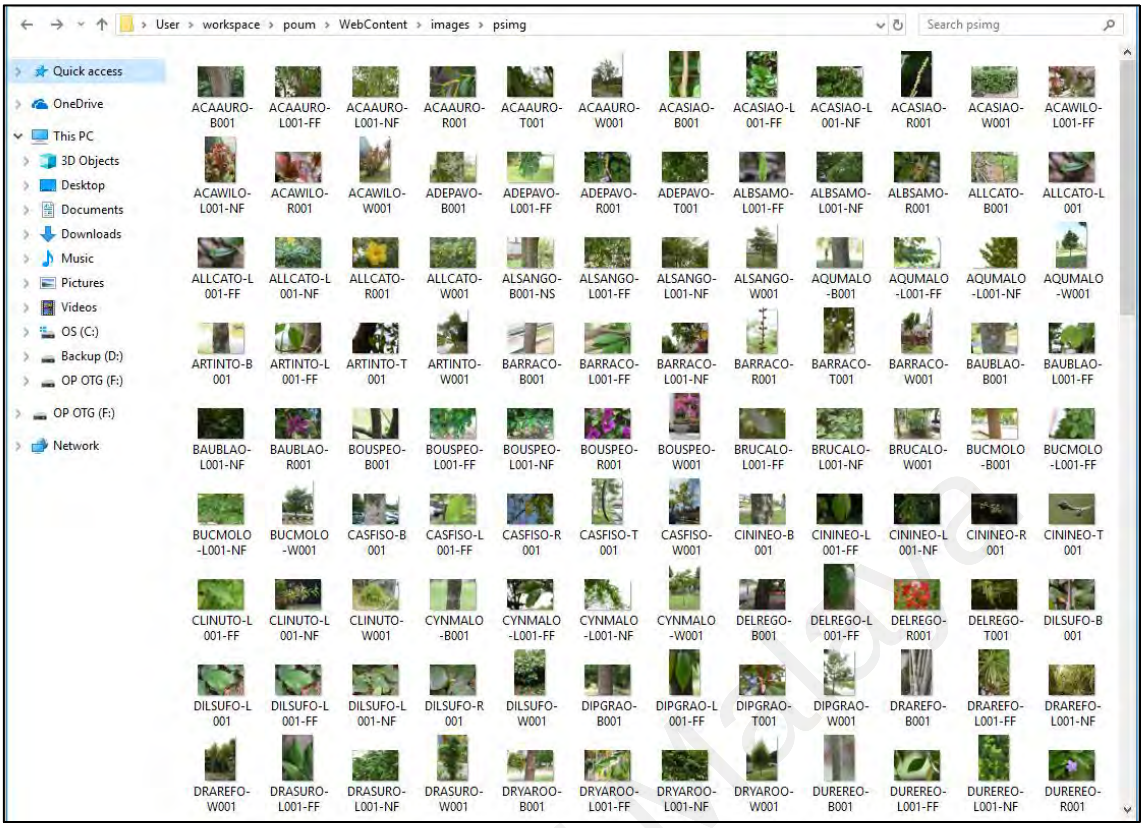


Figure 5.1: Plant image database

In the future, any additional images can be simply stored in the current existing directory and follow the standard naming scheme.

5.3 Plant Ontology Universiti Malaya – POUM

POUM consists of ontological plant data in textual form. Figure 5.2 shows a view of POUM using OntoGraph plug-in in Protégé 5.2. All collected plant data are annotated with vocabularies from the POUM, consisting of nine main classes, five subclasses, 22 object properties, and 39 data properties. The description of all classes and properties are as presented in Appendix B. This set of standardized vocabularies is designed to fit the requirement in this research, which is a combination of existing vocabularies from TDWG (TDWG, 2018) and newly defined vocabularies.

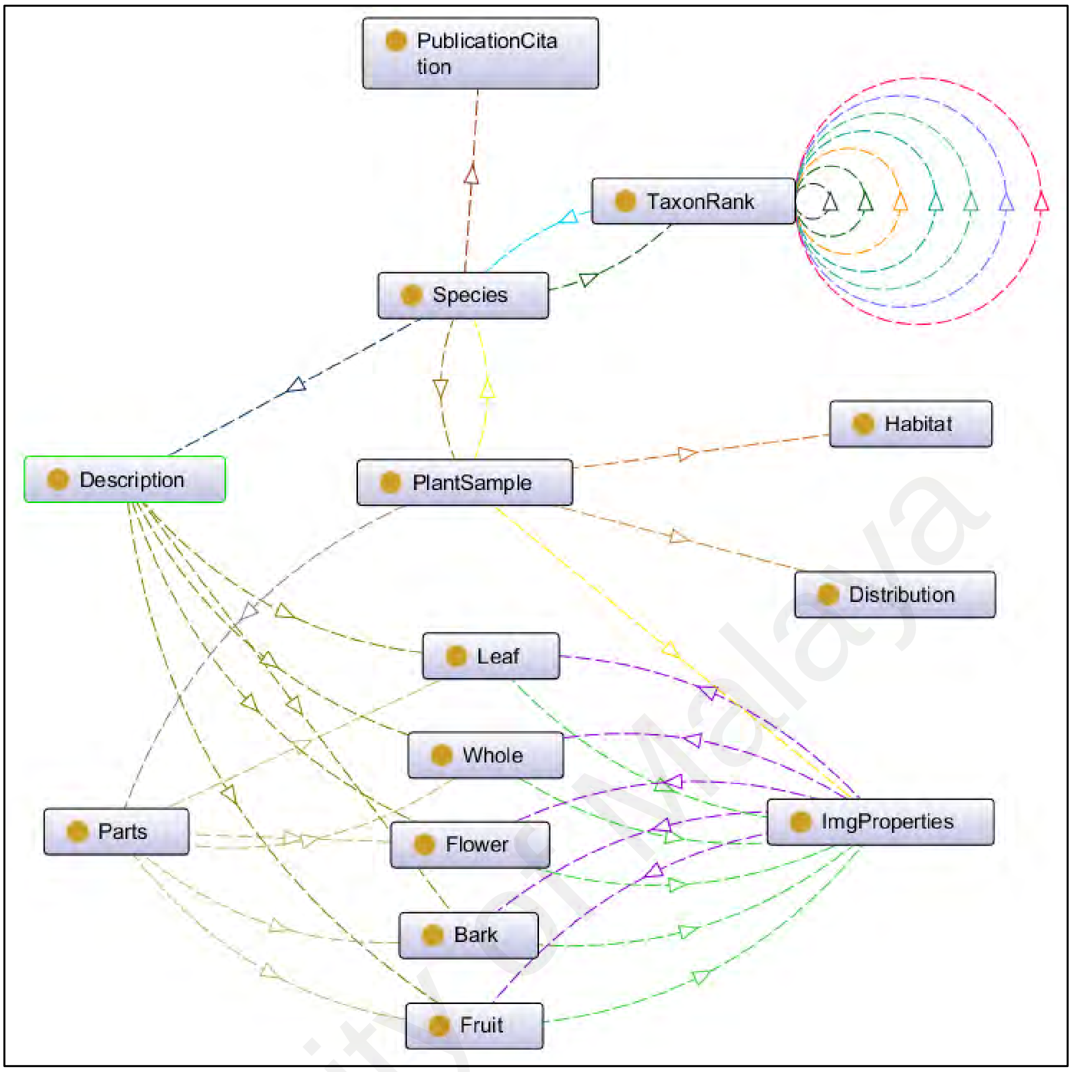


Figure 5.2: The top-level view of classes in POUM ontology

Class 'Species' and 'TaxonRank' contain data of taxonomical classification of the plant species; class 'Species' includes data such as scientific name, common name, authorship and type of plant while class 'TaxonRank' consists of taxonomical rank of plant species. Class 'PlantSample' comprises data of plant samples collected during data acquisition. Class 'Habitat' contains ecological attributes of the location where plant samples are collected while class 'Distribution' consists of geological distribution of plant samples. Data of both classes 'Habitat' and 'Distribution' are based on geological distribution where it refers to locations of plant samples. Besides that, class 'ImgProperties' contains images of each part of the plant samples where based on these images, morphological characteristics of plant samples are described in class 'Parts' where each part of plant sample is described in subclasses ‘Leaf’, ‘Whole’, ‘Bark’, ‘Flower’ and ‘Fruit’ respectively. In addition, class ‘Description’ contains morphological characteristic of plant species in which the data are obtained from multiple sources such as journal articles, books, and online public databases.

Currently, there are 43 species of 42 genera for trees and 31 species of 28 genera for shrubs as listed in Table 5.1 and Table 5.2 with a total of 222 samples in POUM.

The current amount of plant species covered in the POUM is insufficient to portray the whole Plantae kingdom. Additional data from other plant species especially of Malaysian based plant species are needed in future. More plant data can be added to POUM, for example other plant types including their descriptions, plant systematics, ecology, diversity and behaviour. Besides that, POUM as well can be linked to other existing plant-based ontologies (Hebeler et al., 2009; Smith et al., 2007) to provide more information for users. For example, Plant Ontology (PO) (Avraham et al., 2008) and Trait Ontology (Walls et al., 2012).

Comparing POUM schema to other plant ontology schemas, the advantage of POUM is that it describes the morphological characteristics of plant parts. While the common existing schemas, the PO describes more on anatomy and development of plants, and Trait Ontology describes on phenotypic traits in plants.

Table 5.1: List of selected tree species

Family	Genus	Species
Anacardiaceae	Mangifera	*Mangifera indica*
Annonaceae	Polyalthia	*Polyalthia longifolia*
Apocynaceae	Alstonia	*Alstonia angustiloba*
	Plumeria	*Plumeria rubra*
Bignoniaceae	Spathodea	*Spathodea campanulata*
	Tabebuia	*Tabebuia rosea*
Calophyllaceae	Mesua	*Mesua ferrea*
Combretaceae	Bucida	*Bucida molinetii*
	Terminalia	*Terminalia catappa*
Dipterocarpaceae	Dipterocarpus	*Dipterocarpus grandiflorus*
	Dryobalanops	*Dryobalanops aromatica*
	Hopea	*Hopea odorata*
Euphorbiaceae	Hura	*Hura crepitans*
Fabaceae	Acacia	*Acacia auriculiformis*
	Adenanthera	*Adenanthera pavonina*
	Albizia	*Albizia saman*
	Bauhinia	*Bauhinia blakaena*
	Cassia	*Cassia fistula*
	Cynometra	*Cynometra malaccensis*
	Delonix	*Delonix regia*
	Erythrina	*Erythrina variegata*
	Hymenaea	*Hymenaea courbaril*
	Pterocarpus	*Pterocarpus indicus*
	Saraca	*Saraca thaipingensis*
	Senna	*Senna surattensis*
Gentianaceae	Fagreae	*Fagraea fragrans*
Lauraceae	Cinnamomum	*Cinnamomum iners*
Lecythidaceae	Barringtonia	*Barringtonia racemosa*
Lythraceae	Lagerstroemia	*Lagerstroemia floribunda*
Malvaceae	Sterculia	*Sterculia foetida*
	Theobroma	*Theobroma cacao*
Meliaceae	Khaya	*Khaya senegalensis*
	Swietenia	*Swietenia macrophylla*
Moraceae	Artocarpus	*Artocarpus integer*
	Ficus	*Ficus microcarpa*

Table 5.2: List of selected shrub species

Family	**Genus**	**Species**
Myrtaceae	Eucalyptus	*Eucalyptus alba*
	Melaleuca	*Melaleuca cajuputi*
	Syzygium	*Syzygium aqueum*
		Syzygium campanulatum
	Tristaniopsis	*Tristaniopsis whiteana*
Sapindaceae	Filicium	*Filicium decipiens*
Sapotaceae	Mimusops	*Mimusops elengi*
Thymelaeaceae	Aquilaria	*Aquilaria malaccensis*
Acanthaceae	Clinacanthus	*Clinacanthus nutans*
	Graptophyllum	*Graptophyllum pictum*
	Strobilanthes	*Strobilanthes crispa*
Apocynaceae	Allamanda	*Allamanda cathartica*
	Tabernaemontana	*Tabernaemontana divaricata*
Araliaceae	Polyscias	*Polyscias balfouriana*
Asparagaceae	Dracaena	*Dracaena reflexa*
		Dracaena surculosa
Dilleniaceae	Dillenia	*Dillenia suffruticosa*
Euphorbiaceae	Acalypha	*Acalypha siamensis*
		Acalypha wilkesiana
	Excoecaria	*Excoecaria cochinchinensis*
	Manihot	*Manihot esculenta*
Hamamelidaceae	Loropetalum	*Loropetalum chinense*
Lythraceae	Lagerstroemia	*Lagerstroemia indica*
	Lawsonia	*Lawsonia inermis*
Magnoliaceae	Magnolia	*Magnolia figo*
Malvaceae	Hibiscus	*Hibiscus rosa-sinensis*
	Malvaviscus	*Malvaviscus arboreus*
Melastomataceae	Melastoma	*Melastoma malabathricum*
	Tibouchina	*Tibouchina urvilleana*
Nyctaginaceae	Bougainvillea	*Bougainvillea spectabilis*
Phyllanthaceae	Phyllanthus	*Phyllanthus myrtifolius*
	Sauropus	*Sauropus androgynus*
Rubiaceae	Ixora	*Ixora javanica*
	Mussaenda	*Mussaenda erythrophylla*
		Mussaenda philippica
Rutaceae	Murraya	*Murraya paniculata*
Solanaceae	Bruntelsia	*Brunfelsia calycina*
Verbenaceae	Duranta	*Duranta erecta*
	Lantana	*Lantana camara*

5.4 Ontology Evaluation

POUM ontology is evaluated using criteria-based approach (Burton-Jones et al., 2004; Duque-Ramos et al., 2013; Gruber, 1995) because of its clarity and lucidity in examining the developed ontology. In addition, it is straightforward on its purpose and uncomplicated, providing clear view for the developer in evaluating the ontology.

i) Clarity

Gruber (1995) states that definition of a term should be objective. In other words, the definition can only be interpreted in a specific way and should not be ambiguous. During the designing process, experts had analysed the choice of vocabulary used and refined the vocabularies based on their feedback until all vocabularies used in the final version are agreed upon. All definitions are documented with natural language to avoid any confusion by the users. Clarity of POUM is also inspected by running eight tests as listed below and all tests returned true.

1. No cardinality restriction on transitive properties
2. No classes or properties in enumerations
3. No import of system ontologies
4. No meta-class
5. No properties with class as range
6. No sub classes of RDF classes
7. No super or sub properties of annotation properties
8. Transitive properties cannot be functional

Examples of result for Test 1 and Test 8 are as shown in Figure 5.3. There is no transitive property applied to any object properties as biological data always evolve over time which means there is a possibility that new data may be added in the future. Moreover, all instances are related to more than one object property which means none of the object properties can be functional.

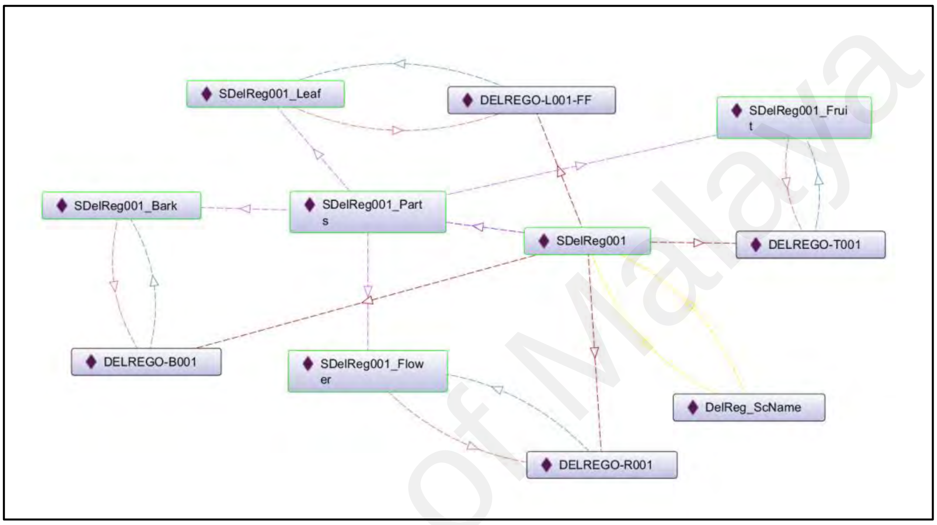
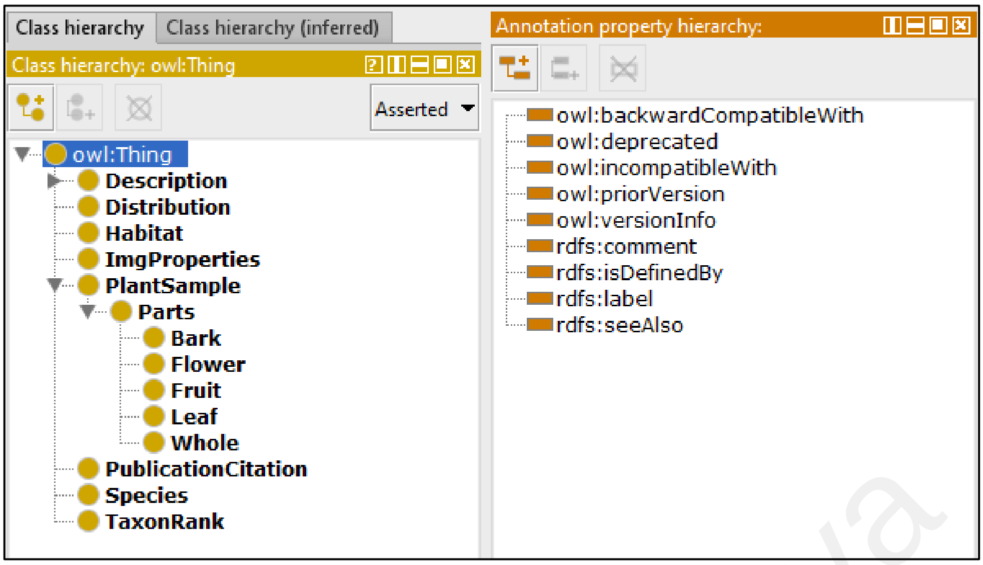


Figure 5.3: Result of Clarity test (Test 1 and Test 8)

Meanwhile, Figure 5.4 shows examples of result for Test 2, Test 3, and Test 7. It shows that there is imported system ontology into UM Plant Ontology, no classes in enumeration, as well as there is also no super or sub properties of Annotation properties. Moreover, result for Test 3 is justified by the fact that structured vocabulary used in this research consists of TDWG standard and newly defined vocabulary to suit the development of the ontology based on this research's requirement. Hence, there is no external ontology used in this research.



Class hierarchy
Class hierarchy (inferred)
Class hierarchy: owl:Thing
Asserted
Annotation property hierarchy:

- owl:Thing
  - Description
  - Distribution
  - Habitat
  - ImgProperties
  - PlantSample
    - Parts
      - Bark
      - Flower
      - Fruit
      - Leaf
      - Whole
  - PublicationCitation
  - Species
  - TaxonRank

- owl:backwardCompatibleWith
- owl:deprecated
- owl:incompatibleWith
- owl:priorVersion
- owl:versionInfo
- rdfs:comment
- rdfs:isDefinedBy
- rdfs:label
- rdfs:seeAlso

Figure 5.4: Result of Clarity test (Test 2, Test 3, and Test 7)

Results for tests 4, 5 and 6 are as illustrated in Figure 5.2 where there is no meta-class, properties with class as range and sub classes of RDF classes.

ii) Coherence

This criterion is described as the logical consistency of an ontology where there should be no contradictions in an ontology's definitions and axioms. The formality of this ontology is checked by running these eight consistency tests shown below and returned true.

1. Domain of a property should not be empty
2. Domain of a property should not contain redundant classes
3. Range of a property should not contain redundant classes
4. Inverse of a functional must be inverse functional
5. Inverse of inverse functional must be functional
6. Inverse of top level property must be top level property
7. Inverse property must have matching range and domain
8. Inverse of symmetric property must be symmetric property

Result for Test 1 returns true as domain for all properties in POUM is assigned (refer Table 5.3). Meanwhile, results for tests 2 to 7 return true where each object property is defined with the functional characteristics as well as matching domain and range as presented in Table 5.3 and do not have redundant classes.

Test 8 returned true as illustrated in Figure 5.5 where it shows a symmetric property 'isA'. Instance 'SDelReg001_Parts' is related to four other instances which are 'SDelReg001_Fruit', 'SDelReg001_Flower', 'SDelReg001_Bark', and 'SDelReg001_Leaf' by property 'isA'. Hence, we can deduce that the other four instances are also related to instance 'SDelReg001_Parts' by 'isA' property which are true. In other words, the symmetric property 'isA' is its own inverse property.

Table 5.3: The domain and range of object properties in POUM

Object Property	**Description**	**Domain**	**Range**
consistOf	Component of the plant sample	PlantSample	Parts
hasClass	Refer to the class level of the hierarchy	TaxonRank	TaxonRank
hasDesc	Describe the characteristics of parts of plant species	Species	Description
hasDetail	Description of the plant samples' images.	ImgProperties	Fruit Bark Flower Whole Leaf
hasFamily	Refer to the family level of the hierarchy.	TaxonRank	TaxonRank
hasGenus	Refer to the genus level of the hierarchy.	TaxonRank	TaxonRank
hasHabitat	Habitat of the plant sample.	PlantSample	Habitat
hasImg	Properties of the plant sample's images.	PlantSample	ImgProperties
hasOrder	Refer to the order level of the hierarchy.	TaxonRank	TaxonRank
hasPhylum	Refer to the phylum level of the hierarchy.	TaxonRank	TaxonRank
hasSample	Samples of plant species.	Species	PlantSample
hasSpecies	Refer to the species level of the hierarchy.	TaxonRank	Species
isA	Component of the parts of plant samples.	Parts	Fruit Bark Flower Whole Leaf
isBelongTo	Refer to the upper level of the hierarchy.	TaxonRank Species	TaxonRank
isCitedFrom	Source of the plant data obtained.	Species	PublicationCitation
isSpecies	Detail of the scientific name of the plant sample.	PlantSample	Species
livesIn	Location of the plant sample in UM.	PlantSample	Distribution
of	Component of the parts of plant species.	Description	Fruit Bark Flower Whole Leaf
hasRank	Details of the hierarchy of the taxonomical rank.	TaxonRank	TaxonRank
referImg	Refer to plant sample's images	Fruit Bark Flower Whole Leaf	ImgProperties

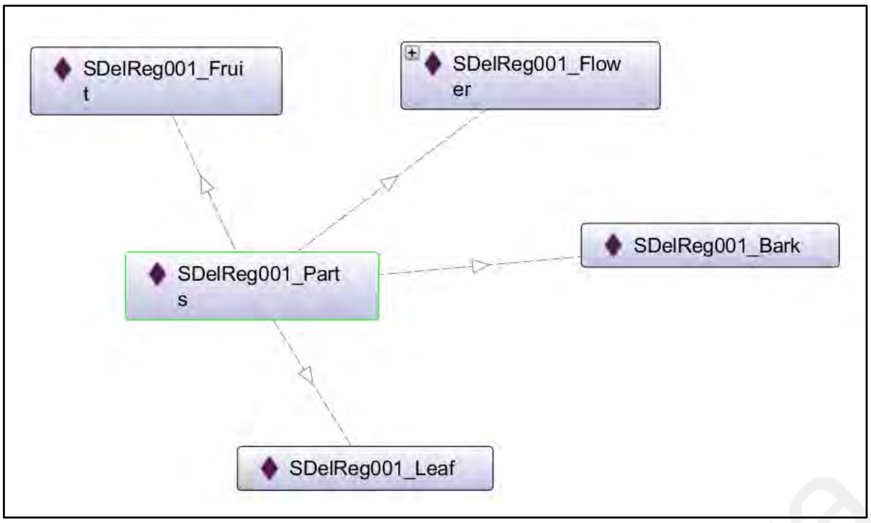
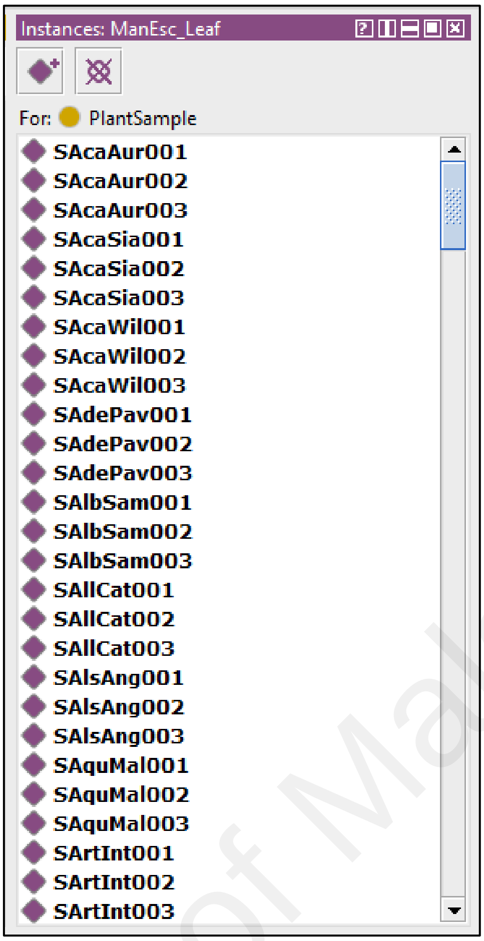


Figure 5.5: Result of Coherence test (Test 8)

iii) Extendibility

This criterion is defined as the ability to further describe the specific knowledge domains while at the same time does not alter the current definitions within the ontology (Gruber, 1995). It should be able to define new terms without the need to revise the existing definitions. Ontology extension is quite important for enabling current ontology to be further developed when new information or knowledge is achieved.

For instance, initially details of plant and its sample are to be set under the same class 'Plant' in the ontology. However, in future more data will be added to include more species and plant species, therefore class 'Plant' is divided into two separate classes which are 'Species' and 'PlantSample'. This is to avoid any confusion about the specification of both vocabularies. Besides that, the naming scheme of individuals for some classes is also emphasized such as 'ImgProperties' and 'PlantSample'. For instance, as shown in Figure 5.6, each individual of class 'PlantSample' is numbered starting from 001 to allow further additional up to 999 of plant samples.



SAcaAur001
SAcaAur002
SAcaAur003
SAcaSia001
SAcaSia002
SAcaSia003
SAcaWil001
SAcaWil002
SAcaWil003
SAdePav001
SAdePav002
SAdePav003
SAlbSam001
SAlbSam002
SAlbSam003
SAllCat001
SAllCat002
SAllCat003
SAlsAng001
SAlsAng002
SAlsAng003
SAquMal001
SAquMal002
SAquMal003
SArtInt001
SArtInt002
SArtInt003

Figure 5.6: Individuals in class 'PlantSample'

iv) Conciseness

This criterion states that an ontology should not have any redundancy in term of definitions used. An ontology is said to be concise if there is no useless and redundant definition and it cannot be inferred from its definitions and axioms. This is where definitions of every terms used in the ontology is compiled in a document to ensure that there is no repetition or ambiguous meanings. Classes ‘Description’ and ‘Parts’ had same subclasses name. Despite that, subclasses from both classes do not have the same definitions. In natural language, class ‘Description’ refers to ‘description of each parts of the plant species’ while class ‘Parts’ refers to ‘information of parts found on plant sample taken’ where only parts of plant sample found at the time during sampling process are described in detail.

v) Correctness

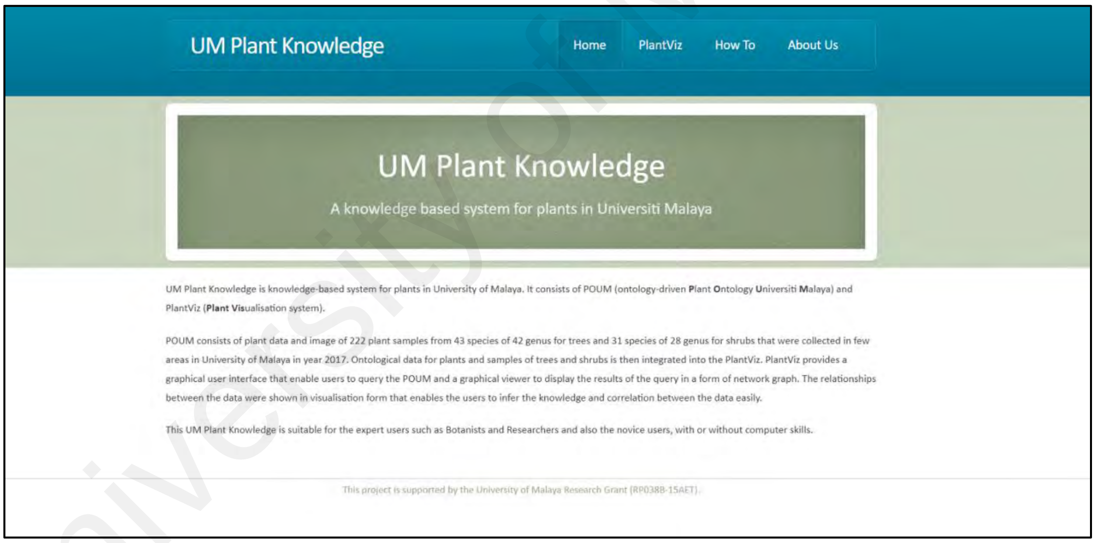
This criterion means that the representation of the knowledge in the ontology follows the real-world concepts. In pursuance of fulfilling this criterion, feedbacks and opinions from domain experts as well as references to other trusted information sources are used in refining the process of ontology structure. Example mentioned previously in Conciseness criterion test shows that the correct meanings of terms used by botanists in real world are applied in this ontology.

Based on this ontology evaluation, it is proven that POUM is reliable and extendable for future advancement.

As described previously, ontology can be represented in many serialization formats such as RDF, XML and RDF/XML and system developers have many options of format that they can choose to suit their system's needs. Therefore, data retrieval using ontology is more efficient compared to others. In this research, the developed visualisation system utilizes ontology-based data retrieval where it uses SPARQL query, which based on logical description. SPARQL uses expression that is closer to humans' mental description of the domain compared to SQL (Staab & Studer, 2013) as it uses triple statement. In addition, data in the ontology is stored and retrieved in RDF format. Thus, the conceptual data model can be fully explored through easily adjustable SPARQL query.

5.5 Plant Visualisation – PlantViz

PlantViz, a prototype of the web-based plant data visualisation system consists of a query tool and graphical viewer. The PlantViz as shown in Figure 5.7 is freely accessible at http://103.18.1.10:8080/plantviz/. The query tool provides four query parameters which are 'Scientific Name', 'Family Name', 'Location', and 'Water Usage' to perform a text-based query. The graphical viewer displays the retrieved data in visualisation form along with interactive features to allow users to communicate directly with the data. The detailed manual on how to use the PlantViz is provided at http://103.18.1.10:8080/plantviz/howto.html. The examples for all query parameters and the visualised data are described more as the followings.



UM Plant Knowledge is knowledge-based system for plants in University of Malaya. It consists of POUM (ontology-driven **P**lant **O**ntology **U**niversiti **M**alaya) and PlantViz (**Plant Vis**ualisation system).

POUM consists of plant data and image of 222 plant samples from 43 species of 42 genus for trees and 31 species of 28 genus for shrubs that were collected in few areas in University of Malaya in year 2017. Ontological data for plants and samples of trees and shrubs is then integrated into the PlantViz. PlantViz provides a graphical user interface that enable users to query the POUM and a graphical viewer to display the results of the query in a form of network graph. The relationships between the data were shown in visualisation form that enables the users to infer the knowledge and correlation between the data easily.

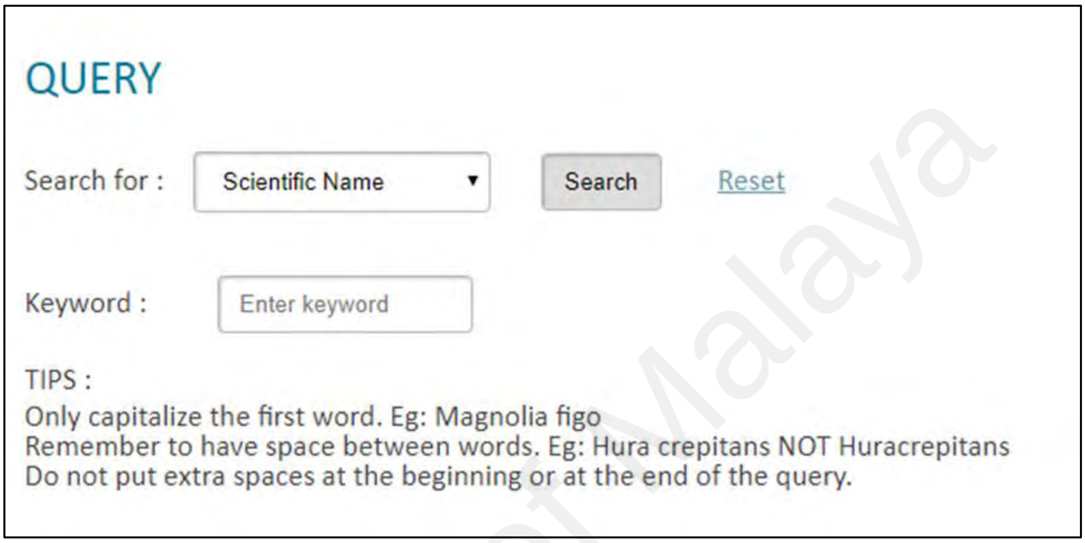
This UM Plant Knowledge is suitable for the expert users such as Botanists and Researchers and also the novice users, with or without computer skills.

This project is supported by the University of Malaya Research Grant (RP038B-15AET).

Figure 5.7: PlantViz system that is freely accessible at http://103.18.1.10:8080/plantviz/

i) Query parameter - 'Scientific Name'

Parameter of 'Scientific Name' as shown Figure 5.8 is commonly used as search parameter in many public databases (Evergreen, 2017; MyBIS, 2017; Natural Resources Conservation Service, 2017). Tips on search keywords are provided to assist users in performing the query. For instance, Figure 5.9 shows the retrieved result for 'Ixora javanica'. In the graphical result, it is shown that node containing the scientific name ‘Ixora javanica’ and of ‘Sample’ is highlighted in purple colour to emphasize the relation between these nodes. Further detail of nodes such as ‘General Information’, ‘Taxon Rank’, ‘SIxoJav001’, ‘SIxoJav002’. and ‘SIxoJav003’ can be obtained by clicking on the nodes.



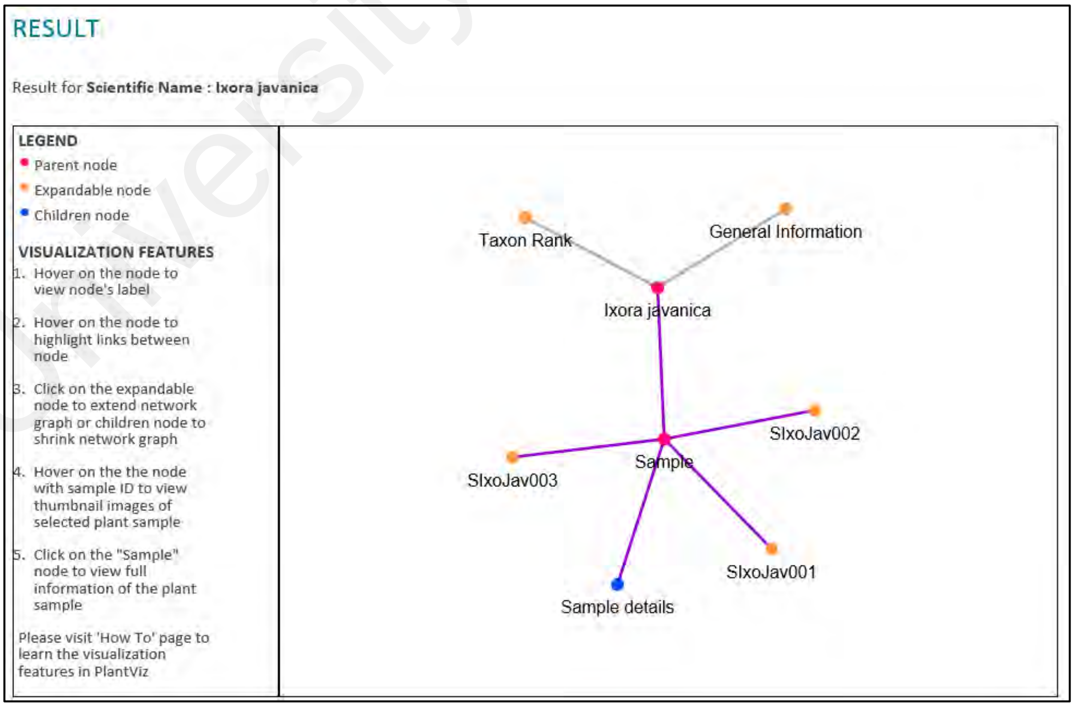
QUERY

Search for : Scientific Name ▾

Keyword : Enter keyword

TIPS :
Only capitalize the first word. Eg: Magnolia figo
Remember to have space between words. Eg: Hura crepitans NOT Huracrepitans
Do not put extra spaces at the beginning or at the end of the query.

Figure 5.8: Search parameter 'Scientific Name' in PlantViz



RESULT

Result for **Scientific Name : Ixora javanica**

LEGEND
- ● Parent node
- ● Expandable node
- ● Children node

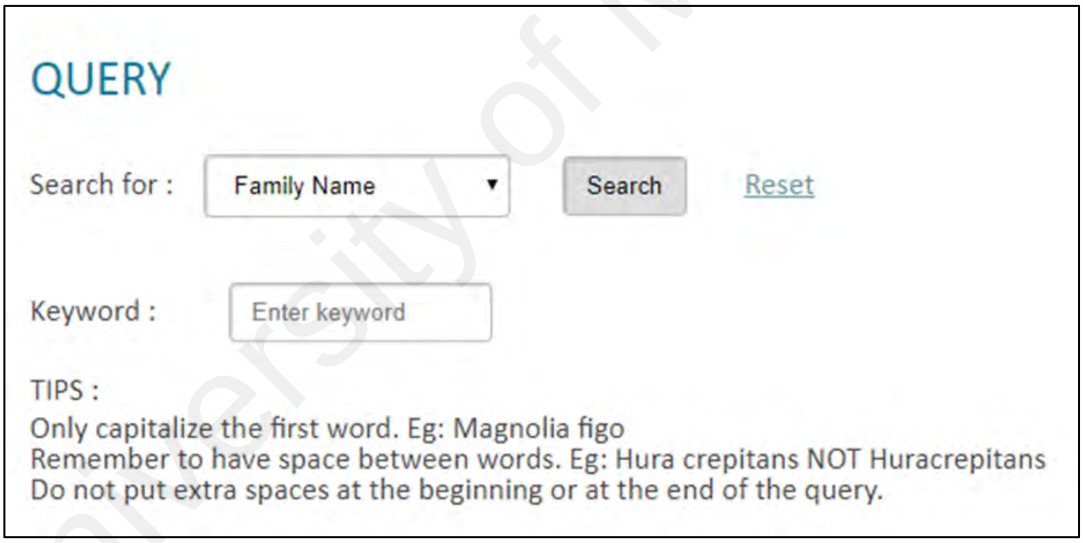
VISUALIZATION FEATURES
1. Hover on the node to view node's label
2. Hover on the node to highlight links between node
3. Click on the expandable node to extend network graph or children node to shrink network graph
4. Hover on the the node with sample ID to view thumbnail images of selected plant sample
5. Click on the "Sample" node to view full information of the plant sample

Please visit 'How To' page to learn the visualization features in PlantViz

Figure 5.9: Result for query parameter 'Scientific Name'

ii) Query parameter - ‘Family Name’

Parameter ‘Family Name’ is another common search parameter used in public databases. This textual-based query parameter allows users to search for family name where users need to enter a textual keyword in the form provided. Figure 5.10 shows that tips on search keywords are provided to assist users in performing the query. Figure 5.11 shows the result for the query parameter where keyword ‘Acanthaceae’ is queried. The graphical result shows genus that are belonged to the Acanthaceae which are Strobilanthes, Graptophyllum, and Clinacanthus. More detail of each genus can be obtained by clicking on the nodes containing the genus name.

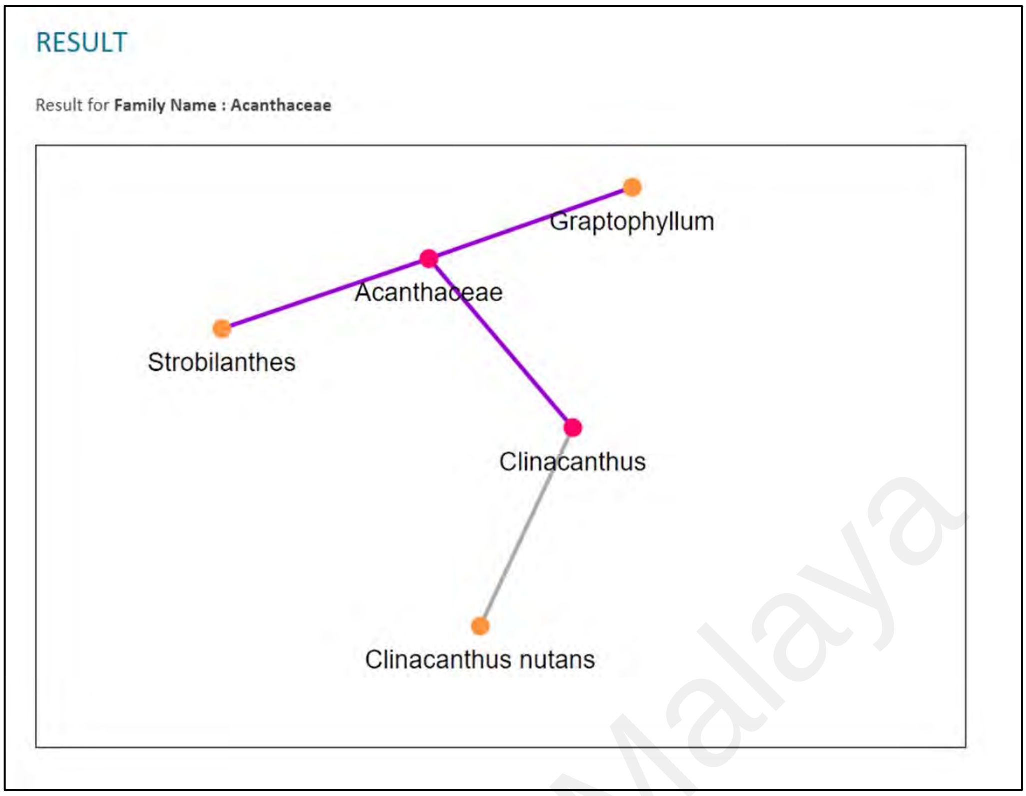


Search for : Family Name

Keyword :

TIPS :
Only capitalize the first word. Eg: Magnolia figo
Remember to have space between words. Eg: Hura crepitans NOT Huracrepitans
Do not put extra spaces at the beginning or at the end of the query.

Figure 5.10: Search parameter 'Family Name' in PlantViz


RESULT

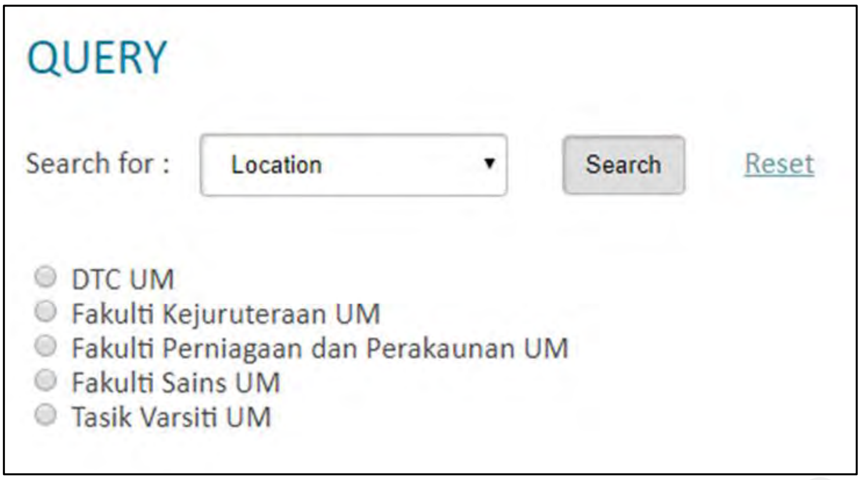
Result for **Family Name : Acanthaceae**

Figure 5.11: Result for query parameter 'Family Name'

iii) Query parameter - 'Location'

Parameter of 'Location' is for location of where the plant samples are collected. It is chosen as one of query parameter because plant samples are mainly collected in various areas of University Malaya. In addition, PlantViz's target users are members of the university who are familiar with the locations in University of Malaya.

As shown in Figure 5.12, when users choose this parameter, a list of available locations are displayed which are 'DTC UM', 'Fakulti Kejuruteraan UM', 'Fakulti Perniagaan and Perakaunan UM', 'Fakulti Sains UM', and 'Tasik Varsiti UM'. Users are required to choose one of the locations to avoid users from entering a location that do not yet exist in the database.



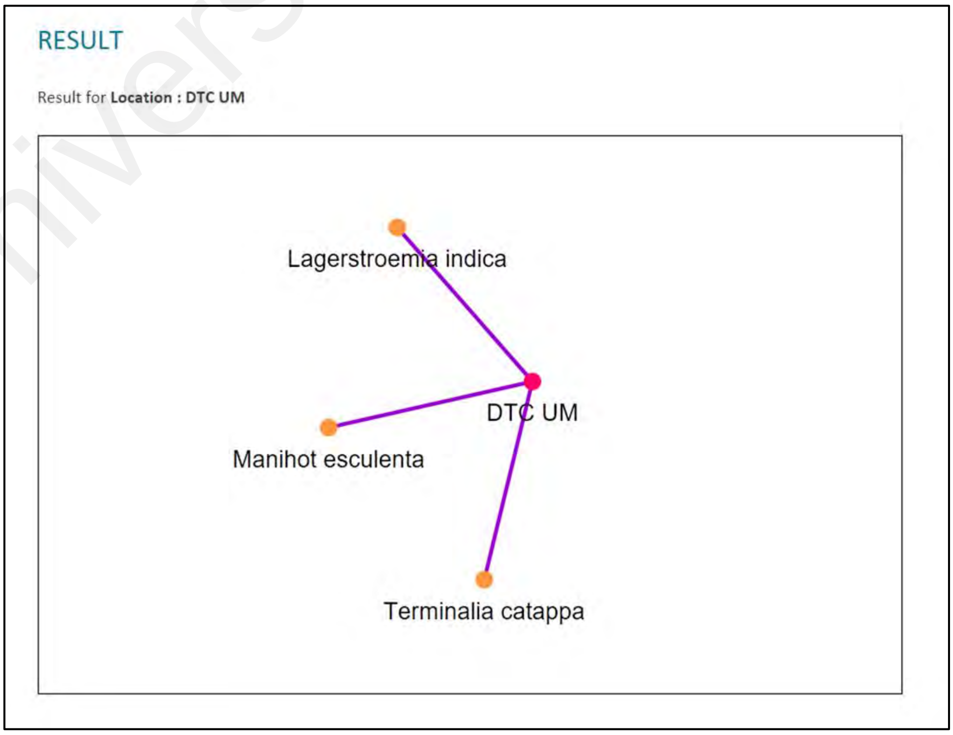
QUERY

Search for : Location ▾ Search Reset

○ DTC UM
○ Fakulti Kejuruteraan UM
○ Fakulti Perniagaan dan Perakaunan UM
○ Fakulti Sains UM
○ Tasik Varsiti UM

Figure 5.12: Search parameter 'Location' in PlantViz

Figure 5.13 shows the result for query parameter ‘Location’ where users queried for ‘DTC UM’. Plant species in which samples are collected at the same location are linked together. In this case, nodes of plant species *Lagerstroemia indica*, *Manihot esculenta*, and *Terminalia catappa* are linked to node ‘DTC UM’. More detail of each plant species can be obtained by clicking on the nodes containing the scientific name.



RESULT

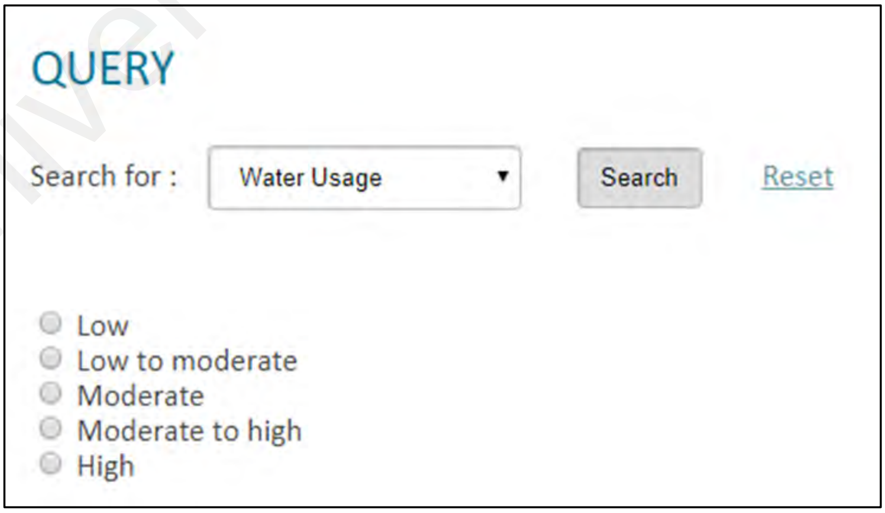
Result for **Location : DTC UM**

Figure 5.13: Result for query parameter 'Location'

iv) Query parameter - ‘Water Usage’

Parameter of ‘Water Usage’ is defined as the level of water used by a plant species for its growth. Since other parameters used are to represent the taxonomical and geographical information of a plant species, the parameter of ‘Water Usage’ is chosen to represent the morphology attribute of a plant. Figure 5.14 shows a list of water usage level that are displayed when choosing this parameter which are ‘Low’, ‘Low to moderate’, ‘Moderate’, ‘Moderate to high’, and ‘High’.

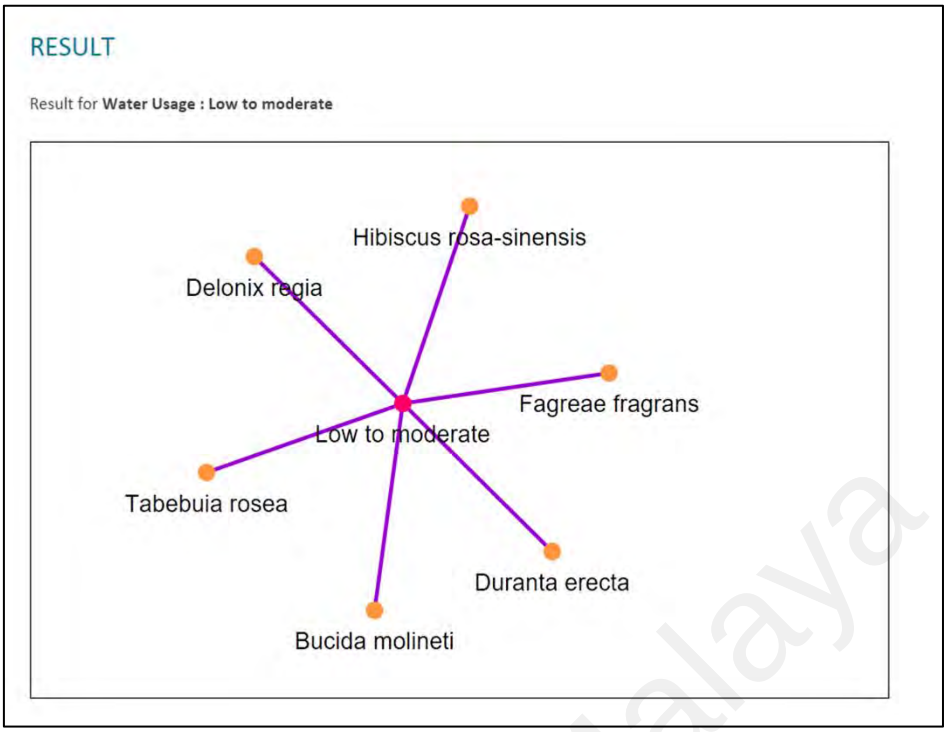
As for the result, Figure 5.15 shows the result for query parameter ‘Water Usage’ where users queried for ‘Low to moderate’. The graphical result shows that nodes of plant species *Delonix regia, Hibiscus rosa-sinensis*, *Fagreae fragrans*, *Duranta erecta*, *Bucida molinetii*, and *Tabebuia rosea* are linked to node ‘Low to moderate’ where further information of each plant species can be obtained by clicking on nodes containing the scientific name.



QUERY

Search for : Water Usage ▾

Figure 5.14: Search parameter 'Water Usage' in PlantViz


RESULT

Result for **Water Usage : Low to moderate**

Figure 5.15: Result for query parameter 'Water Usage'

5.6 Interactive Features in the Graphical Viewer

The graphical viewer in PlantViz is developed to visualize the query result as it helps in eliminating the problem of lengthy texts retrieved as part of the result. In addition, relationships between data are accurately visualized; hence it helps in emphasizing crucial points of the data to users. Moreover, the graphical viewer consists of a number of interactive features which allow users to explore the result (Lohmann et al., 2015).

i) View label of node

Users can view label of each node when hovering the cursor over a node as shown in Figure 5.16. The label is only visible when users hover the cursor over a node to avoid clutters in the graphical viewer.

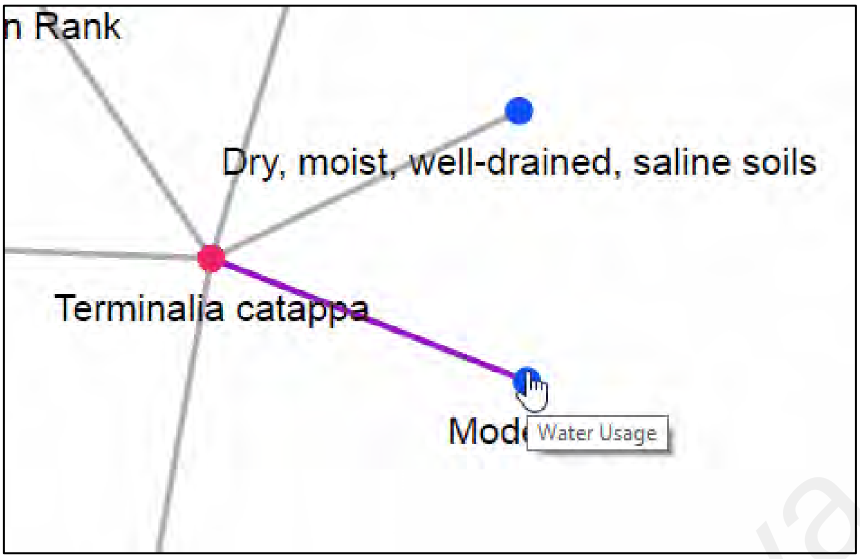


Figure 5.16: Viewing the label of a selected node

ii) Highlighting the links between the nodes

Another interactive feature is highlighting the links whereby the links of related nodes will be highlighted in different colours when the user is hovering the cursor over the selected node as shown in Figure 5.17. The purpose of this feature is to allow user to see the relationships between the data, thus encouraging them to identify new patterns and glean new insight from the visualisation. For example, based on Figure 5.17, users can determine the relationship between *Manihot esculanta* and its samples where there are 3 samples for this species which are 'SManEsc001', 'SManEsc002', and 'SManEsc003'. Further exploration of the visualisation can show the relationships between samples by clicking on 'Sample details'.

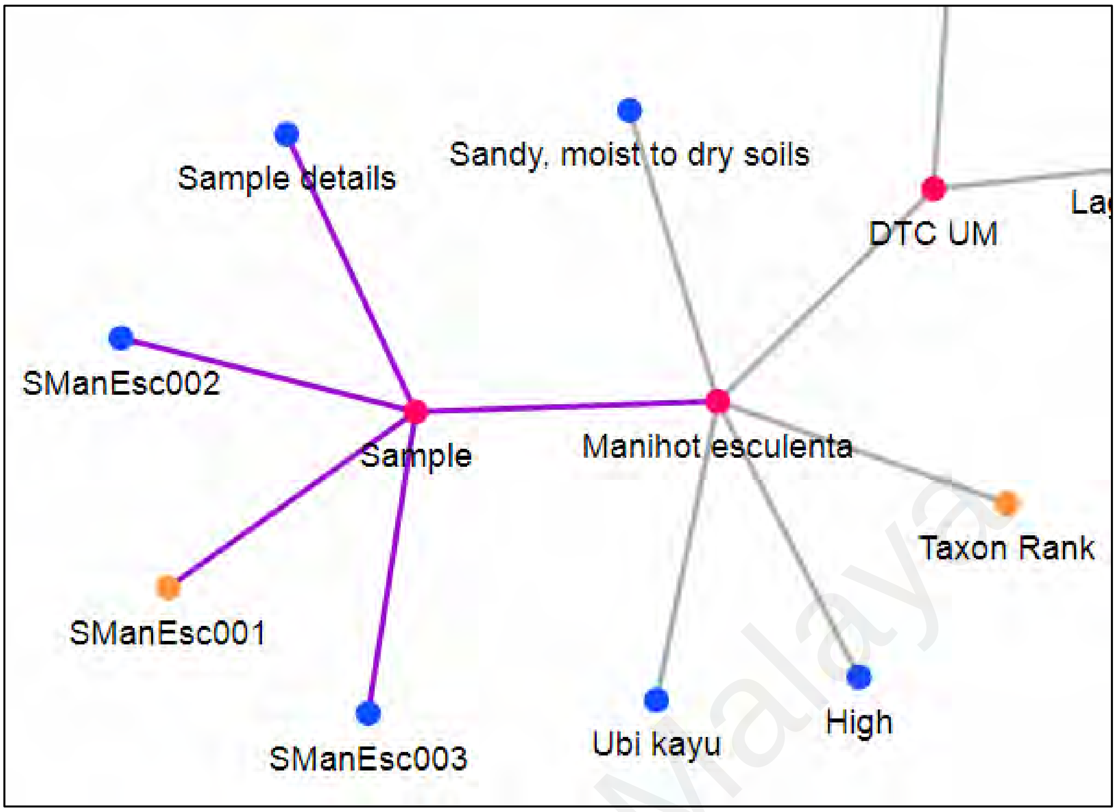


Figure 5.17: Highlighting the links of the nodes

iii) Expand or shrink a group of nodes

From the graphical viewer, users also can expand a node by clicking on ‘parent’ node with orange colour; or shrink a group of nodes by clicking on ‘children’ node with pink colour as shown in Figure 5.18. This is to avoid cluttered look in the graphical viewer as some of the search query may generate plenty of nodes.

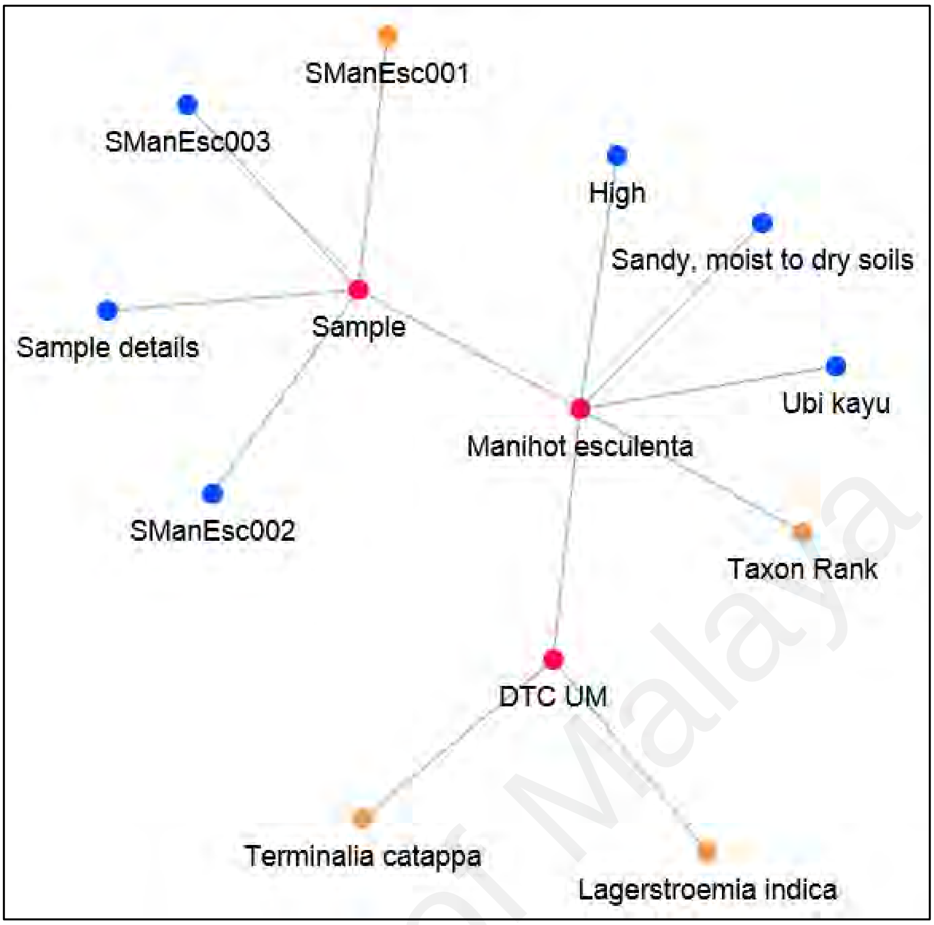


Figure 5.18: Expanding or shrinking a group of nodes

iv) **View thumbnail sized images**

Besides that, Figure 5.19 shows another feature where users are able to view the thumbnail sized images of plant samples. These images are only visible when users hovered on nodes with label 'Sample ID'.

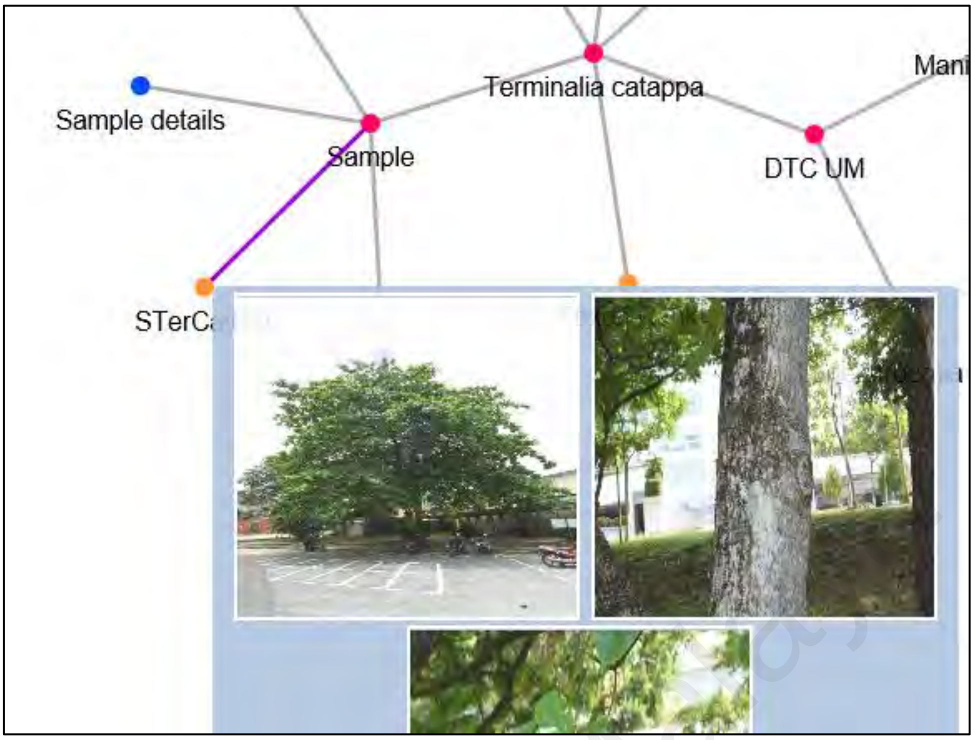
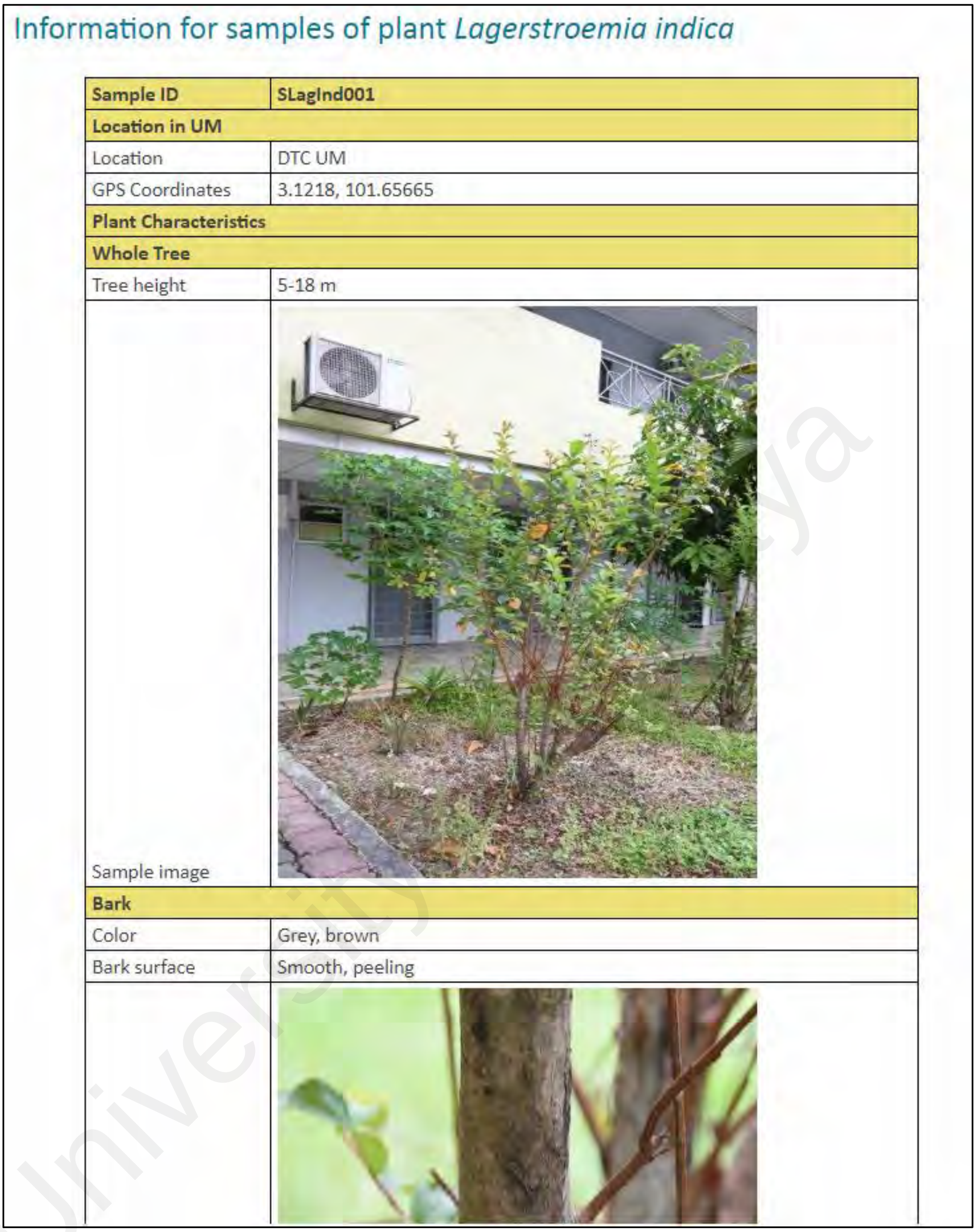


Figure 5.19: Viewing the thumbnail images of the plant sample

v) Redirecting to webpage of plant sample's information

As there are some users who might prefer information of the sample of the plant species, there is another interactive feature that redirects the users to another webpage of plant sample's information. Users can access this webpage by clicking on the node with label 'More detail' and a new webpage will pop up. This webpage contains the GPS coordinate and location of the plant sample in UM, image of the plant sample as well as the description of the plant sample. Example of the plant sample's information is as shown in Figure 5.20.



Information for samples of plant *Lagerstroemia indica*

Sample ID	SLagInd001
Location in UM	
Location	DTC UM
GPS Coordinates	3.1218, 101.65665
Plant Characteristics	
Whole Tree	
Tree height	5-18 m
Sample image	
Bark	
Color	Grey, brown
Bark surface	Smooth, peeling

Figure 5.20: A new page displaying the plant sample's information

5.7 User Evaluation

User evaluation is performed on the querying and visualisation of PlantViz. 60 users are involved in which 30 are expert users and 30 are novice users. Expert users consist of botanists, researchers and postgraduate in biodiversity fields with little knowledge in IT while novice users are students of undergraduate programme in the University of Malaya.

5.7.1 Usability Heuristics Evaluation

Figure 5.21 presents an analysis of the usability heuristics evaluation by both the expert and novice users. This shows that the majority of users rated ‘Yes’ for most of the features. As shown in Figure 5.21(a), *E6* has the highest number for the rating ‘Yes’ (all 30 expert users voted ‘Yes’), while for novice users, as shown in Figure 5.21(b), *E2*, *E6*, and *E9* have the highest number for the rating ‘Yes’ (all 30 novice users voted ‘Yes’). Meanwhile, for both types of users, *E1* has the highest number for rating ‘No’ (16 out of 30 expert users and 14 out of 30 novice users voted ‘No’). This is consistent with the prototype development. The guidelines for using the system are available with no status for the system being shown.

Fisher’s Exact test checks whether the user experience in using PlantViz is dependent on their expertise level. The null hypothesis H_0 is that there is no difference between the usability heuristics evaluation performed between expert users and novice users. The two-tailed probability (*p*) value of Fisher’s Exact test on usability heuristics evaluation is 0.312, ($p > 0.05$), which means that there is no significant difference in the usability heuristic evaluation between expert users and novice users. This also indicates that PlantViz’s user interface is adequate for all types of users regardless of their IT knowledge.

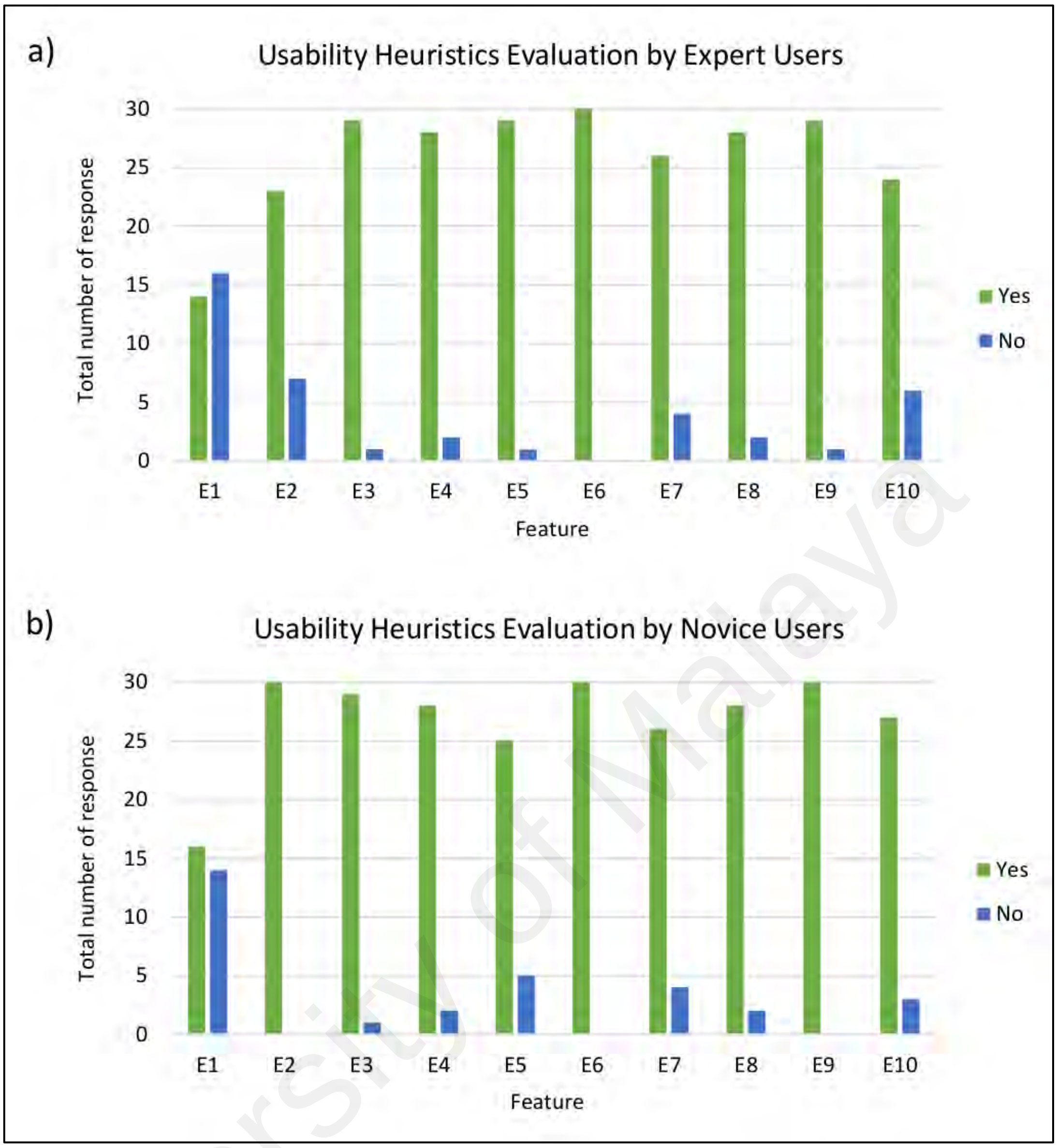


Figure 5.21: Analyses of usability heuristic evaluation by expert and novice users

5.7.2 Query and Visualisation Evaluation

Complete analysis of query and visualisation evaluation for all four cases by both expert and novice users are shown in Appendix E.

Figure 5.22 shows the analysis of query evaluation by expert and novice users where the total number of responses for each case is plotted against evaluation rating. The query evaluation by both types of users is shown in Figure 5.22(a) and Figure 5.22(b). These show similar results whereby the majority of expert and novice users rated '4' and '5' for most the query criteria. There is one response that gave a rating of '1' (one expert user) and a total of nine responses that gave a rating of '2' (six expert and three novice users). A low rating is given for *Q4* which is '*Limitation in number of keyword at a time*' as shown in Appendix E(i).

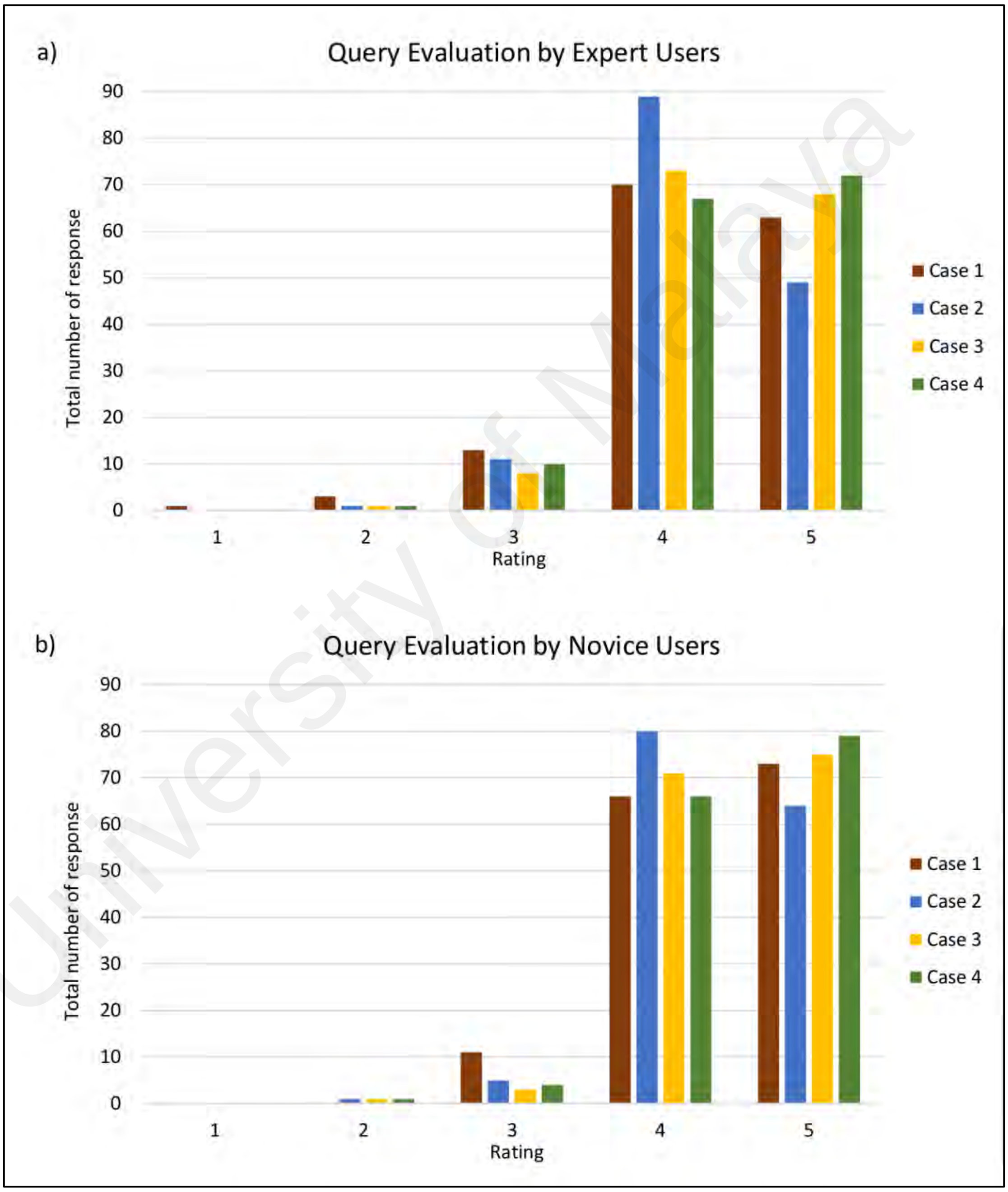


Figure 5.22: Analyses of query evaluation by expert and novice users

Figure 5.23 shows the analysis of visualisation evaluation by expert and novice users where the total number of responses for each case is plotted against evaluation rating as well. Figures 5.23(a) and 5.23(b) show that both types of users rated ‘4’ for most the visualisation criteria. There is one response with a rating of ‘1’ and 35 responses with a rating of ‘2’ (see Appendix E(ii)). *V7* is given a rating of ‘1’ (one expert user rated ‘1’) and had the highest number for a rating of ‘2’ (four expert and nine novice users rated ‘2’). Besides that, *V6* had seven responses with a rating of ‘2’ (two expert and five novice users rated ‘2’).

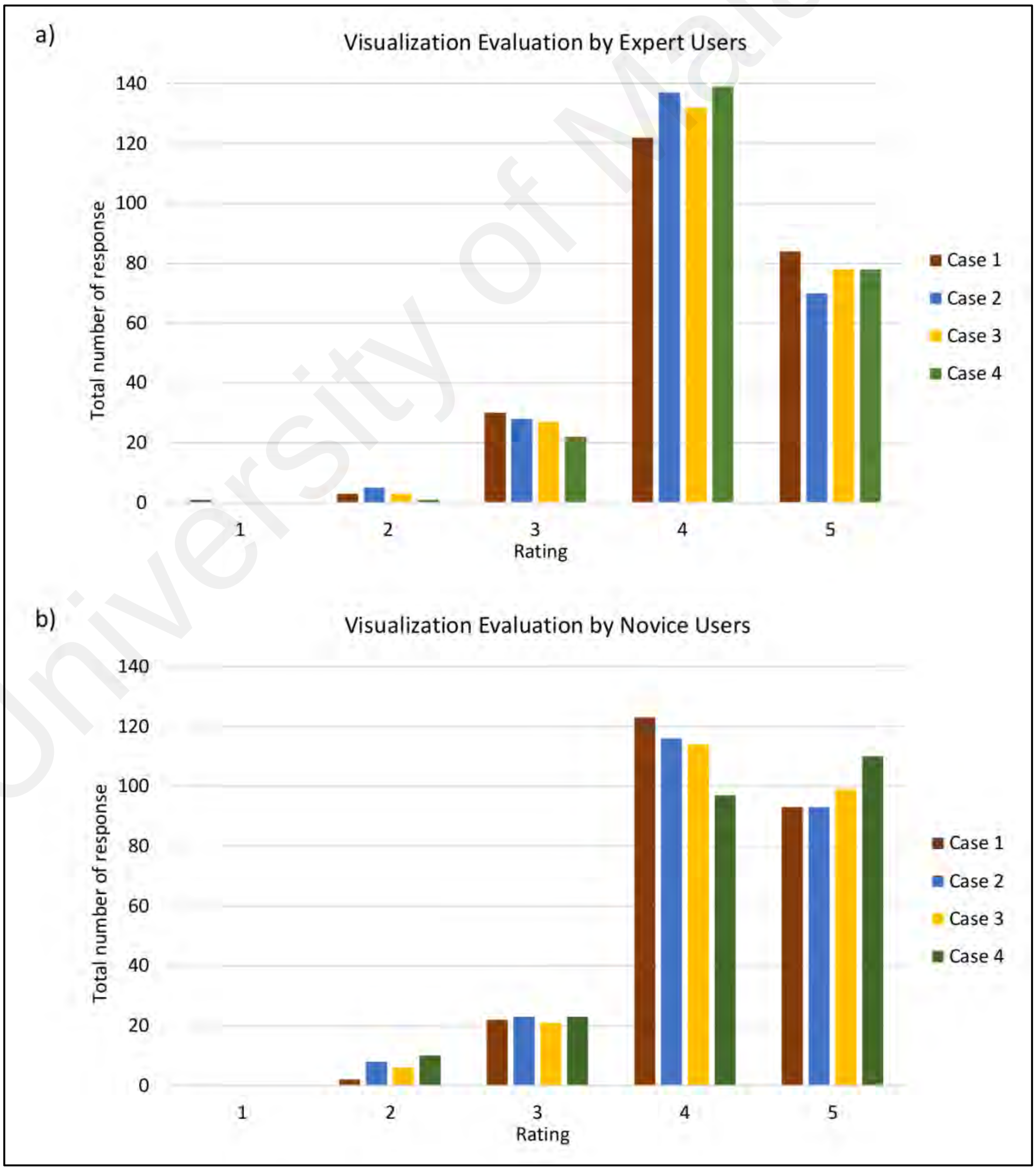


Figure 5.23: Analyses of visualisation evaluation by expert and novice users

Independent sample t-tests test the significance of the evaluation done by both types of users. The null hypothesis H_0 declares that there is no difference between the evaluation of expert users and novice users. T-tests are performed on each case for both evaluations with results as shown in Table 5.4. The *p*-values for all cases are higher than the significance level of 0.05, thus there is a lack of evidence to reject H_0. Hence, there is no statistically significant difference between the query and visualisation evaluation performed by expert and novice users.

From the observation, *p*-values of the query and visualisation evaluations share the same pattern, where Case 1 has the lowest *p*-value. It can be presumed that in Case 1, users are not yet familiar with the graphical viewer of the PlantViz system as the graphical viewer is not a common tool in many public plant-based databases such as The Plant List and NParks Flora & Fauna. However, all other cases namely cases 2, 3, and 4, for both evaluations have *p*-values higher than Case 1. Hence, the previous assumption is valid, as users had just started to become acquainted with PlantViz. Moreover, the GUI design of both query tool and graphical viewer in PlantViz are simple yet still appropriate for both types of users. As for Case 4, both evaluations had the highest *p*-values. Therefore, it can be concluded that users are easily accustomed to the PlantViz. This also verifies that the GUI for PlantViz is consistent throughout all cases, where each case used different search parameters.

Table 5.4: Results of independent sample t-test for query and visualisation evaluations

Cases	***p*-value**	
	Query evaluation	**Visualisation evaluation**
Case 1	0.082	0.133
Case 2	0.105	0.165
Case 3	0.177	0.172
Case 4	0.225	0.409

Even so, as mentioned earlier, the GUI's design of the PlantViz is simple yet appropriate. The query method used in retrieving the data is a typical Boolean search. To improve the querying process, a more robust and efficient query method can be achieved by implementing other query methods such as ranking algorithm (Tran et al., 2009; Zhiguo & Zhengjie, 2010), precision models (Cox, 2005; Kwak et al., 2013) and natural language query processing (Damljanovic et al., 2010; Paredes-Valverde et al., 2015). Improvement to the query tool can also be made where filtering feature can be added to the graphical viewer to provide users the ability to filter search results. For instance, the 'Location' parameter is used as the search parameter for Case 3 and this can generate cluttered network graph, as many plant samples are collected from the same location. Thus, users can filter the results (Cline et al., 2007) by only selecting certain parameters such as family name, number of plant samples collected, or type of plant which can help in retrieving better graphical result.

Additionally, current user interface only allows users to select from four search parameters and only enter one data query at a time, therefore it limits users' selection to perform the query. This can be enhanced where other relevant search parameters can be included in the query tool. For example, common name, type of plant, and trait (Kattge et al., 2011). It is vital for users to query the system without attaching to a fixed one, and to select pre-defined parameters. Based on the low-rated Q4 (refer Appendix E), users are dissatisfied with the number of search parameters that can be used to query at a time. This limitation will be enhanced in the future by allowing users to add more than one search parameter to narrow down the retrieved results, so that more relevant results can be retrieved.

For the graphical viewer, some improvements can be made to increase its efficiency and usability to users. The graphical viewer in PlantViz is developed with the purpose to display retrieved data in form of visualisation. Therefore, it helps in eliminating the problem of lengthy texts retrieved as part of the result. In addition, relationships between data are accurately visualized, which assist in emphasizing the crucial point of data to users. However, more additional interactive features can help users to have better experience in exploring the result (Lohmann et al., 2015). For example, filtering tool can be added to the graphical viewer where users are able to select certain nodes to be shown in the graphical viewer (Kerren et al., 2017). In a more advanced interactive feature, similarity measurement tool can be added where from the graphical viewer, users are able to calculate the similarity between nodes (Wang et al., 2013). Other feature such as to allow users to choose different types of visualisation to be generated using circular layout (Krzywinski et al., 2009), hyperbolic tree (Bingham & Sudarsanam, 2000; Manning et al., 2002) or phylogram (Huson et al., 2007; Okonechnikov et al., 2012) can be included as well. Users can choose the colour scheme too, to differentiate the relationships between data (Ciccarelli et al., 2006) too.

5.8 Summary

This chapter provides the outcomes of the proposed solution. The POUM and PlantViz performed well and follow the listed user requirements as stated previously in Section 3.5. Thus, the objectives of this research are successfully achieved.

CHAPTER 6: CONCLUSION AND FUTURE WORK

6.1 Introduction

In this chapter, the summary of the research is discussed in detail whereby the conclusion of this research is presented and the future works for system enhancement are proposed.

6.2 Proposed Data Visualisation System

In biological field, online database is one of the main sources to obtain biological data. It is the fastest way to retrieve data other than from journals and books. As discussed earlier in Chapter 2, many databases only provide textual data in a plain view. While there are users who are accustomed to this conventional method of data presentation, it does not help them in data analysis and new knowledge finding. Moreover, the advancement in research methodology and technology is very rapid, hence it is crucial that the data obtained from experiments and researches can be obtained and analysed in shorter time. Besides that, it only makes sense that the advancement of both research methodology and data analysis occur together.

The highlight of this research is to use an alternative method to improve the way data is presented to users which is by applying the data visualisation approach. A data visualisation system is designed as illustrated in Figure 6.1. It is developed based on the proposed three-tier system architecture as shown in Figure 4.2. Plant data and samples that are collected in UM are obtained during data acquisition activity. Then, it is organized into POUM that contains textual data and images of plant samples.

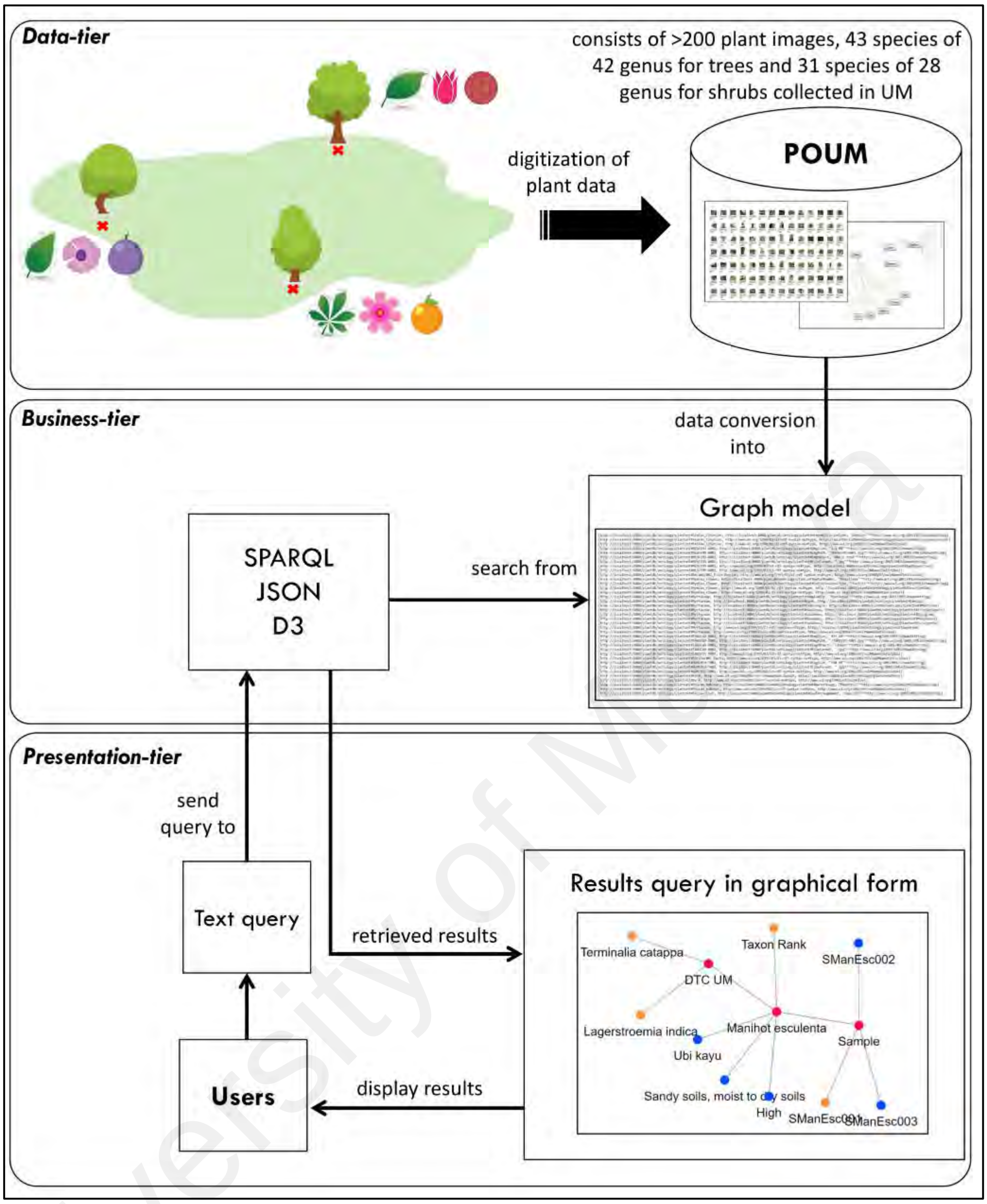


Figure 6.1: The proposed system architecture consists of POUM and PlantViz

Users' query is processed in SPARQL query and querying process is done to the graph data of the ontology. As the result of SPARQL query which is in form of texts, conversion of the textual result to graphical form is required. Detailed process flow of this conversion is as shown in Figure 4.21.

The proposed system architecture also includes user interface in presentation-tier. There are two types of user interface which are query and result interfaces. Query interface eases the search for data, so that they do not need to deal directly with the querying process in the server. Query interface collects data query from the users and results of the query are displayed in the result interface in interactive graphical form. Interactive features in the visualisation of the result enable users to communicate with the data and provide cognitive support for data analysis and encourage new insight.

Furthermore, the method used in this work is relevant even with the increasing amount and complexity of the data with the help of external dataset and additional methods to organise the visualisation. For instance, many of visualisation tools are linked directly with public databases (Saraiya et al., 2005; Pavlopoulos et al., 2008) which allow users to interpret their own data based on previous knowledge. Apart from that, the amount of clutter in the visual where visual entities are crowded and disordered can be reduced by applying clutter reduction technique (Carpendale et al., 1996; Huey et al., 1999). Thus, the integrity of the data and information content are still intact without losing any crucial information. For example, the dimension of the data in the visualisation can be reordered as different dimension of the data can reveal different aspects of the data and thus affect the perceived clutter and structure of the visualisation (Peng et al., 2004). Ellis and Dix (2007) provides an in-depth analysis of clutter reduction techniques that focus on different aspect of the visualisation such as type of data, features, and algorithms used. Thus, it is shown that the inevitable issues of large and complex dataset can be tackled by adding extra processes to ensure the visualisation is always fitting to users' needs.

6.3 Conclusion

The field of biology generates thousands of data every day with the advancements in modern tools and technologies. Biological data, especially plant data is easily accessible from many data sources. However, it is commonly available in lengthy descriptions and texts. Likewise, many plant-based databases do not emphasize on the relationships between plant data which devalues the information. Thus, it is important to implement a proper methodology for retrieving data and to make the data accessible to users in an effective way. This research focuses on presenting retrieved data from the ontology to users in a visualisation form.

Consequently, in this research a visual-based representation system of plant data, PlantViz, is proposed and developed. Data from POUM ontology are used as the dataset in which it is converted as a graph data model and queried using SPARQL. Then, query results are structured in JSON format before being transformed into a visualisation form and presented to users in GUI. PlantViz consists of a query tool and graphical viewer that comes with interactive elements features which allow interaction between the users and data. Query tool consists of four search parameters that are commonly used as search parameters in many public databases except ‘Water Usage’ parameter which is chosen to represent the morphology attribute of the plant. Graphical viewer makes PlantViz a unique plant-based data visualisation system as it is not a common tool in many public plant-based databases particularly in Malaysia compared to MyCHM (FRIM, 2017), MyBIS (MyBIS, 2017) and Malaysia Botanical Garden (Malaysia Botanical Garden, 2018). Furthermore, visualisation of plant data using PlantViz successfully emphasized on relationships between plant data. Unlike other databases that has static graphical viewer (Dash et al., 2012; Jaiswal et al., 2005; USDA, 2017), interactive features in the PlantViz’s graphical viewer allow users to explore the result and encourage two-way communication. The usability of PlantViz is measured by carrying out user evaluation which is in two parts: (i) usability heuristics evaluation, and (ii) query and visualisation evaluation. From the analysis of user evaluation, it shows that PlantViz can be used by users with different background, either experts from the botanical field, students, or laymen with interests in the botanical area, with or without the IT skill.

Besides that, PlantViz uses POUM which is an ontology that consists of data for plant species and images of 222 plant samples. Ontology modelling used in representing the plant data in semantic manner has successfully described the plant data accurately while reducing the ambiguity of the vocabularies used. Ontology successfully accentuated on the relationships between data by declaring properties of each instance in each concept. Properties of a concept are classified into object property and datatype property which are important in defining the meaning of each concept accurately. Besides that, ontological data are retrieved in natural language in logical expression of the domain, hence it is easier to form a query. This is because ontological data is converted into a graph data and is retrieved using SPARQL that uses the triple statement; "*subject-predicate-object*" structure to form the query. The result of the data retrieval is transformed into a form of visualisation which provides cognitive support to users in making data analysis to provide more meaningful data to users whereby the relationships between data are highlighted.

In conclusion, the aim of this research is achieved and the main contributions of this research are: (i) the alternative approach in presenting plant data to users using visualisation approach, (ii), provide knowledge enrichment through inference of relationships between plant data, (iii) implementation of semantic representation of plant data in visualisation system, (iv) integration of ontology-based data retrieval into a data visualisation system, and (v) the plant ontology that describes on the morphological characteristics in plants.

6.4 Future Work

Content-based image retrieval (CBIR) is a method that allows digital images to be organized based on their visual features such as shape, colour, and texture (Jadhav & Patil, 2012; Liu & Yang, 2013; Yue et al., 2011). The query using CBIR analyses for the contents of the image and not the metadata of the image. It is a method that is increasingly popular in many different fields (Huang et al., 2018; Long et al., 2009; Mallik et al., 2010). Therefore, for system improvement in the future, PlantViz can perform automatic identification of organism at species or genus using CBIR in which users can perform query using images of plants and the result of the query can be presented to users in graphical form. Moreover, in a more advanced system, searching algorithms can be applied in PlantViz CBIR system such as ranking algorithms (Tran et al., 2009; Zhiguo & Zhengjie, 2010), precision models (Cox, 2005; Kwak et al., 2013), and graph-theory algorithms (Dogrusoz et al., 2009; Sojoudi et al., 2014). This future work provides better query results to users with high accuracy as well as decreasing the semantic gaps between images' features and its semantic representations (Hu & Gao, 2009; Ying et al., 2005).

Besides that, the usability of PlantViz can be enhanced by allowing users to upload their own ontology as the dataset for PlantViz. This is to show that PlantViz is not restricted to predefined data and is able to visualize other data that is not plant-based. It is an achievable work as there are other systems that imply similar idea (Gilson et al., 2008; Lohmann et al., 2015).

Meanwhile, data modelling using ontology brings many advantages to developers as discussed in Section 2.3. For instance, reusability, flexibility, and ability to perform knowledge analysis which is a crucial process when users need to reuse or extend the existing ontology (Lu & Jin, 2002). Furthermore, it promotes knowledge enrichment which is available when the ontology is combined with different ontologies where semantic annotations can associate information with specific entities from different ontologies, yet still within the domain of interest (Valarakos et al., 2004). Therefore, in the future, POUM can be modified by extending the ontology, such as merging other existing ontology such as Plant Ontology (Avraham et al., 2008) and Trait Ontology (Walls et al., 2012) or adding more data into POUM. This can improve the performance of PlantViz and POUM as well as to optimize the usage of the ontology. Moreover, it enables the developers to create an organized and centralized knowledge without the fuss of managing a huge amount of data at one time.

REFERENCES

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., . . . Terauchi, R. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nature Biotechnology, 30*, 174.

Abrol, D. P. (2011). Pollination - basic concepts. In *Pollination biology: Biodiversity conservation and agricultural production* (pp. 37-53). New York: Springer.

Agarwal, M., Shrivastava, N., & Padh, H. (2008). Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports, 27*(4), 617-631.

Ahmad, M. N. (2012). *Ontology-based applications for enterprise systems and knowledge management*. Pennsylvania: IGI Global.

Agafonkin, V. (2017). Leaflet - a JavaScript library for interactive maps. Retrieved 7th January 2018, from https://leafletjs.com/

Ai, J., Smith, B., & Wong, D. T. (2010). Saliva ontology: An ontology-based framework for a salivaomics knowledge base. *BMC Bioinformatics, 11*(1), 302.

Alba, R., Fei, Z., Payton, P., Liu, Y., Moore, S. L., Debbie, P., . . . Giovannoni, J. (2004). ESTs, cDNA microarrays, and gene expression profiling: Tools for dissecting plant physiology and development. *The Plant Journal, 39*(5), 697-714.

Aldinucci, M., Coppo, M., Damiani, F., Drocco, M., Torquati, M., & Troina, A. (2011). On designing multicore-aware simulators for biological systems. In *International Euromicro Conference on Parallel, Distributed and Network-Based Processing* (pp. 318-325). Ayia Napa, Cyprus: IEEE.

Amri, S., Ltifi, H., & Ayed, M. B. (2015). Towards an intelligent evaluation method of medical data visualisations. In *International Conference on Intelligent Systems Design and Applications (ISDA)* (pp. 673-678). Morocco: IEEE.

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology, 11*(10), 106.

Andorf, C. M., Cannon, E. K., Portwood, I. I. J. L., Gardiner, J. M., Harper, L. C., Schaeffer, M. L., . . . Lawrence-Dill, C. J. (2016). MaizeGDB update: New tools, data and interface for the maize model organism database. *Nucleic Acids Research, 44*(1), 195-1201.

Angles, R., & Gutierrez, C. (2008). Survey of graph database models. *ACM Computing Surveys, 40*(1), 1-39.

Apache Jena (2018). Apache Jena - Reasoners and rule engines: Jena inference support. Retrieved 5th April 2018, from https://jena.apache.org/documentation/inference/index.html

ArangoDB. (2014). ArangoDB - Highly available multi-model NoSQL database. Retrieved 2nd January 2018, from https://www.arangodb.com/

Arivazhagan, S., Shebiah, R. N., Ananthi, S., & Varthini, S. V. (2013). Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. *Agricultural Engineering International: CIGR Journal, 15*(1), 211-217.

Armstead, I., Huang, L., Ravagnani, A., Robson, P., & Ougham, H. (2009). Bioinformatics in the orphan crops. *Briefings in Bioinformatics, 10*(6), 645-653.

Arteca, R. N. (2013). *Plant growth substance: Principles and applications*. New York, US: Springer Science & Business Media.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., . . . Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics, 25*(1), 25-29.

Aurum, A., & Wohlin, C. (2005). *Engineering and managing software requirements*. New York, US: Springer.

Avraham, S., Tung, C.-W., Ilic, K., Jaiswal, P., Kellogg, E. A., McCouch, S., . . . Ware, D. (2008). The Plant Ontology database: A community resource for plant structure and developmental stages controlled vocabulary and annotations. *Nucleic Acids Research, 36*, 449-454.

Barrett, R., & Tay, E. P. (2016). *Perth plants: A field guide to the bushland and coastal flora of Kings Park and Bold Park* (2nd ed.). Australia: Csiro Publishing.

Barrett, T., Troup, D. B., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., . . . Soboleva, A. (2011). NCBI GEO: Archive for functional genomics data sets - 10 years on. *Nucleic Acids Research, 39*, 1005-1010.

Bateman, J., Magnini, B., Rinaldi, F., & Henschel, R. (1994). The generalized Italian, German, English upper model. In *Proceedings of the ECAI-94 Workshop on Implemented Ontologies* (pp. 1-12). Amsterdam.

Bates, P. D., Jewell, J. B., & Browse, J. (2013). Rapid separation of developing Arabidopsis seeds from siliques for RNA or metabolite analysis. *Plant Methods, 9*(1), 9.

Belleau, F., Nolin, M.-A., Tourigny, N., Rigault, P., & Morissette, J. (2008). Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. *Journal of Biomedical Informatics, 41*(5), 706-716.

Bhardwaj, A. A. (2017). Effects of invasive plant leaf litter on a lake ecosystem. *Columbia Undergraduate Science Journal, 11*, 24-33.

Bingham, J., & Sudarsanam, S. (2000). Visualizing large hierarchical clusters in hyperbolic space. *Bioinformatics, 16*(7), 660-661.

Bisby, F. A., Roskov, Y. R., Orrell, T. M., Nicolson, D., Paglinawan, L. E., Bailly, N., . . . Baillargeon, G. (2010). Species 2000 & ITIS catalogue of life: 2010 annual checklist. Retrieved 5th Feb 2018, from http://www.catalogueoflife.org/annual-checklist/2010

Blasi, C., Marignani, M., Copiz, R., Fipaldini, M., Bonacquisti, S., Del Vico, E., . . . Zavattero, L. (2011). Important plant areas in Italy: From data to mapping. *Biological Conservation, 144*(1), 220-226.

Boci, L., Yan, C., Xu, C., & Yingying, Y. (2012). Comparison between JSON and XML in applications based on AJAX. In *International Conference on Computer Science and Service System (CSSS)* (pp. 1174-1177). Nanjing, China: IEEE.

Bock, C., Fokoue, A., Haase, P., Hoekstra, R., Horrocks, I., Ruttenberg, A., . . . Smith, M. (2012*).* OWL 2 web ontology language : Structural specification and functional-style syntax. Retrieved 10th Jan 2018, from https://www.w3.org/TR/owl2-syntax/

Bollacker, K., Evans, C., Paritosh, P., Sturge, T., & Taylor, J. (2008). Freebase: A collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data* (pp. 1247-1250). New York, US: ACM.

Boo, C. M., Chew, S. Y. J., & Yong, J. W. H. (2014). *Plants in tropical cities*. Singapore: Uvaria Tide.

Boratyn, G. M., Camacho, C., Cooper, P. S., Coulouris, G., Fong, A., Ma, N., . . . Zaretskaya, I. (2013). BLAST: A more efficient report with usability improvements. *Nucleic Acids Research, 41*(1), 29-33.

Bostock, M. (2017). D3.js - Data-driven documents. Retrieved 12th October 2017 from https://d3js.org/

Botanic Gardens Conservation International. (2017). PlantSearch database. Retrieved 28th May 2017 from https://www.bgci.org/plant_search.php

Bourgeois, D. T. (2014). Information systems for business and beyond. Retrieved 6th Jan 2018, from http://lib.hpu.edu.vn/handle/123456789/21478

Brewster, C., Alani, H., Dasmahapatra, S., & Wilks, Y. (2004). Data driven ontology evaluation. In *International Conference on Language Resources and Evaluation (LREC)* (pp. 24-30). Lisbon, Portugal

Brittain, J. (2005). *Horticulture - plant names explained: Botanical terms and their meaning* (S. Gordon & M. Cady, Eds.). Winchester, Hampshire: David & Charles.

Brohée, S., Faust, K., Lima-Mendez, G., Sand, O., Janky, R. S., Vanderstocken, G., . . . van Helden, J. (2008). NeAT: A toolbox for the analysis of biological networks, clusters, classes and pathways. *Nucleic Acids Research, 36*, 444-451.

Burton-Jones, A., Storey, V. C., Sugumaran, V., & Ahluwalia, P. (2004). A semiotic metrics suite for assessing the quality of ontologies. *Data & Knowledge Engineering, 55*(1), 84-102.

Campbell, N. A., Reece, J. B., Mitchell, L. G., & Taylor, M. R. (2003). *Biology: Concepts & connections* (4th ed.). New York, US: Pearson Prentice Hall.

Carbon, S., Ireland, A., Mungall, C. J., Shu, S., Marshall, B., & Lewis, S. (2009). AmiGO: Online access to ontology and annotation data. *Bioinformatics, 25*(2), 288-289.

Carpendale, M. S. T., Cowperthwaite, D. J., & Fracchia, F. D. (1996). Distortion viewing techniques for 3-dimensional data. In *Proceedings IEEE Symposium on Information Visualization '96* (pp. 46-53). California, USA: IEEE.

Carranza-Rojas, J., Goeau, H., Bonnet, P., Mata-Montero, E., & Joly, A. (2017). Going deeper in the automated identification of herbarium specimens. *BMC Evolutionary Biology, 17*(1), 181.

Catalogue of Life. (2018). Catalogue of life - 2010 annual checklist : Species details. Retrieved 20th Jan 2018, from http://www.catalogueoflife.org/annual-checklist/2010/details/species/id/7229164

Chen, C. M., Lai, K. J., Pai, T. W., & Chang, H. T. (2014). Transcriptome data visualisation in pathways with application to zebrafish embryo datasets. In *Conference on Complex, Intelligent and Software Intensive Systems (CISIS)* (pp. 515-518). Birmingham: IEEE.

Chen, W. (2008). Nanoparticle fluorescence based technology for biological applications. *Journal of Nanoscience and Nanotechnology, 8*(3), 1019-1051.

Chen, Y. J. (2010). Development of a method for ontology-based empirical knowledge representation and reasoning. *Decision Support Systems, 50*(1), 20.

Chen, Y. L., Kao, H. P., & Ko, M. T. (2004). Mining DAG Patterns from DAG databases. In *International Conference on Web-Age Information Management* (pp. 579-588). Berlin, Heidelberg: Springer.

Chung, L., Nixon, B. A., Yu, E., & Mylopoulos, J. (2012). *Non-functional requirements in software engineering*. New York, US: Springer Science & Business Media.

Ciccarelli, F. D., Doerks, T., Mering, C. v., Creevey, C. J., Snel, B., & Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science, 311*(5765), 1283-1287.

Cline, M. S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., . . . Bader, G. D. (2007). Integration of biological networks and gene expression data using Cytoscape. *Nature Protocols, 2*(10), 2366-2382.

Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM, 13*(6), 377-387.

Conesa, A., & Mortazavi, A. (2014). The common ground of genomics and systems biology. *BMC Systems Biology, 8*, 1.

Cooper, L., Meier, A., Laporte, M.-A., Elser, J. L., Mungall, C., Sinn, B. T., . . . Jaiswal, P. (2018). The Planteome database: An integrated resource for reference ontologies, plant genomics and phenomics. *Nucleic Acids Research, 46*(1), 1168-1180.

Cope, J. S., Corney, D., Clark, J. Y., Remagnino, P., & Wilkin, P. (2012). Plant species identification using digital morphometrics: A review. *Expert Systems with Applications, 39*(8), 7562-7573.

National Research Council (1989). Plant biology and agriculture. In *Opportunities in biology* (pp. 365-402). Washington, DC: The National Academies Press.

County ITS, J. (2015). Colorado plant database, Colorado State University extension. Retrieved 12th January 2018, from https://coloradoplants.jeffco.us/intro.jsp

Cox, E. (2005). Chapter 8 - Fuzzy rule induction. In *Fuzzy modeling and genetic algorithms for data mining and exploration* (pp. 265-339). San Francisco, California: Morgan Kaufmann.

Cukier, K. (2010). Data data everywhere: A special report on managing information. Retrieved 13th December 2017, from http://www.economist.com/node/15557443

Cullina, W. (2002). *Native trees, shrubs, & vines: A guide to using, growing, and propagating North American woody plants*. Boston, Massachusetts: Houghton Mifflin.

Czauderna, T., & Schreiber, F. (2017). Information visualisation for biological data. In J. M. Keith (Ed.), *Bioinformatics - volume II: Structure, function, and applications* (pp. 403-415). New York, US: Springer.

Damljanovic, D., Agatonovic, M., & Cunningham, H. (2010). Natural language interfaces to ontologies: Combining syntactic analysis and ontology-based lookup through the user interaction. In *International Conference on The Semantic Web: Research and Applications* (pp. 106-120). Berlin, Heidelberg: Springer.

Dang, T. N., Murray, P., & Forbes, A. G. (2015). PathwayMatrix: Visualizing binary relationships between proteins in biological pathways. *BMC Proceedings, 9*(6), 3.

Dash, S., van Hemert, J., Hong, L., Wise, R. P., & Dickerson, J. A. (2012). PLEXdb: Gene expression resources for plants and plant pathogens. *Nucleic Acids Research, 40*(1), 1194-1201.

Day-Richter, J., Harris, M. A., Haendel, M., Gene Ontology OBO-Edit Working Group & Lewis, S. (2007). OBO-Edit - An ontology editor for biologists. *Bioinformatics, 23*(16), 2198-2200.

Degtyarenko, K., de Matos, P., Ennis, M., Hastings, J., Zbinden, M., McNaught, A., . . . Ashburner, M. (2008). ChEBI: A database and ontology for chemical entities of biological interest. *Nucleic Acids Research, 36*, 344-350.

Dias, D., Jones, O., Beale, D., Boughton, B., Benheim, D., Kouremenos, K., . . . Wishart, D. (2016). Current and future perspectives on the structural identification of small molecules in biological systems. *Metabolites, 6*(4), 46.

Dietrich, S. W., & Urban, S. D. (2011). *Fundamentals of object databases: Object-oriented and object-relational design.* California: Morgan & Claypool.

Division of Agriculture. (2018). Arkansas plant disease database. Retrieved 3rd Mar 2018, from https://www.uaex.edu/yard-garden/resource-library/diseases/

Dogrusoz, U., Cetintas, A., Demir, E., & Babur, O. (2009). Algorithms for effective querying of compound graph-based pathway databases. *BMC Bioinformatics, 10*(1), 376.

Domingue, J., Fensel, D., & Hendler, J. A. (2011). *Handbook of semantic web technologies*. New York, US: Springer.

Dong, Q., Schlueter, S. D., & Brendel, V. (2004). PlantGDB, plant genome database and analysis tools. *Nucleic Acids Research, 32*, 354-359.

Dreibelbis, A., Hechler, E., Milman, I., Oberhofer, M., Run, P. v., & Wolfson, D. (2008). *Enterprise master data management: An SOA approach to managing core information.* Massachusetts, US: IBM Press.

Duque-Ramos, A., Fernández-Breis, J. T., Iniesta, M., Dumontier, M., Egaña Aranguren, M., Schulz, S., . . . Stevens, R. (2013). Evaluation of the OQuaRE framework for ontology quality. *Expert Systems with Applications, 40*(7), 2696-2703.

Easlon, H. M., & Bloom, A. J. (2014). Easy leaf area: Automated digital image analysis for rapid and accurate measurement of leaf area. *Applications in Plant Sciences, 2*(7), 1-14.

Eckhoff, S. R., Paulsen, M. R., & Yang, S. C. (2003). Maize. In L. Trugo & P. M. Finglas (Eds.), *Encyclopedia of food sciences and nutrition* (pp. 3647-3653). Oxford, UK: Academic Press.

El Ghosh, M., Naja, H., Abdulrab, H., & Khalil, M. (2016). Towards a middle-out approach for building legal domain reference ontology. *International Journal of Knowledge Engineering, 2*(3), 109-114.

Eldridge, J. (2006). Data visualisation tools - a perspective from the pharmaceutical industry. *World Patent Information, 28*(1), 43-49.

Ellis, G., & Dix, A. (2007). A taxonomy of clutter reduction for information visualisation. *IEEE Transactions on Visualization and Computer Graphics, 13*(6), 1216-1223.

Elmasri, R., & Navathe, S. B. (2016). *Fundamentals of database systems* (7th ed.). New York, US: Pearson Education.

Eren, A. M., Morrison, H. G., Lescault, P. J., Reveillaud, J., Vineis, J. H., & Sogin, M. L. (2015). Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *The ISME Journal, 9*(4), 968-979.

Euzenat, J. (2007). Semantic precision and recall for ontology alignment evaluation. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence* (pp. 348-353). San Francisco, CA: Morgan Kaufmann.

Evergreen. (2017). Evergreen native plant database. Retrieved 28th May 2017, from https://nativeplants.evergreen.ca/.

Faiez, G. (2010). *Ontology theory, management and design: Advanced tools and models.* Pennsylvania: Information Science Reference.

FasterXML. (2018). FasterXML/jackson: Main portal page for the Jackson project. Retrieved 28th Jan 2018, from https://github.com/FasterXML/jackson.

Fernández-López, M. (1999). Overview of methodologies for building ontologies. In *Workshop on Ontologies and Problem-Solving Methods: Lessons Learned and Future Trends: International Joint Conference for Artificial Intelligence (IJCAI'99)* (pp. 34-47). Stockholm, Sweden.

Fernández-López, M., & Gómez-Pérez, A. (2002). Overview and analysis of methodologies for building ontologies. *The Knowledge Engineering Review, 17*(2), 129-156.

Few, S. (2004). Eenie, meenie, minie, moe: Selecting the right graph for your message. *Intelligent Enterprise*, *1*(5), 8.

Fiorelli, M., Pazienza, M. T., Petruzza, S., Stellato, A., & Turbati, A. (2010). Computer-aided ontology development: An integrated environment. In *Workshop of New Challenges For NLP Frameworks: International Conference of Language Resources and Evaluation 2010* (pp. 33-40). Valletta, Malta.

Fowler, A. (2015). *NoSQL for dummies.* New Jersey: John Wiley & Sons.

Francesconi, E., Montemagni, S., Peters, W., & Tiscornia, D. (2010). Integrating a bottom-up and top-down methodology for building semantic resources for the multilingual legal domain. In *Semantic processing of legal texts* (pp. 95-121). Germany: Springer-Verlag.

FRIM. (2017). Malaysian biological diversity Clearing House Mechanism (CHM). Retrieved 8th May 2017, from http://www.chm.frim.gov.my/Home.aspx.

Galperin, M. Y., Fernández-Suárez, X. M., & Rigden, D. J. (2017). The 24th annual Nucleic Acids Research database issue: A look back and upcoming changes. *Nucleic Acids Research, 45*(1), 1-11.

Gangopadhyay, A., Molek, M., Yesha, Y., Brady, M., & Yesha, Y. (2012). A methodology for ontology evaluation using topic models. In *4th International Conference on Intelligent Networking and Collaborative Systems* (pp. 390-395). Bucharest, Romania: IEEE.

Garden, B. M. (2018). Tropics name - Binkgo biloba L. Retrieved 23rd Mar 2018, from http://www.tropicos.org/Name/14100001?tab=distribution

Gardner, S., Sitthisunthō̜n, P., & Lai, E. M. (2011). *Heritage trees of Penang.* Pulau Pinang: Areca Books.

Gasteiger, E., Jung, E., & Bairoch, A. M. (2001). SWISS-PROT: Connecting biomolecular knowledge via a protein database. *Current Issues in Molecular Biology, 3*(3), 47-55.

Giesbertz, P., Evelo, C., Hanspers, K., Slenter, D., Kutmon, M., Olia, A., . . . Gaj, S. (2018). Arylamine metabolism (Homo sapiens) - wikipathways. Retrieved 23rd Jan 2018, from https://www.wikipathways.org/index.php/Pathway:WP694.

GO Consortium. (2017). Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Research, 45*(1), 331-338.

Gómez-Pérez, A. (1996). Towards a framework to verify knowledge sharing technology. *Expert Systems with Applications, 11*(4), 519-529.

Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (2007). *Ontological engineering: With examples from the areas of knowledge management, e-commerce and the semantic web*. New Jersey: Springer-Verlag.

Gómez, J., García, L. J., Salazar, G. A., Villaveces, J., Gore, S., García, A., . . . Jiménez, R. C. (2013). BioJS: An open source JavaScript framework for biological data visualisation. *Bioinformatics, 29*(8), 1103-1104.

Grewe, J., Wachtler, T., & Benda, J. (2011). A bottom-up approach to data annotation in neurophysiology. *Frontiers in Neuroinformatics, 5*(16), 1-56.

Grierson, C. S., Barnes, S. R., Chase, M. W., Clarke, M., Grierson, D., Edwards, K. J., . . . Bastow, R. (2011). One hundred important questions facing plant science research. *New Phytologist, 192*(1), 6-12.

Groenendyk, M. (2013). Emerging data visualisation technologies for map and geography libraries: 3D printing, holographic imaging, 3D city models, and 3D model-based animations. *Journal of Map & Geography Libraries: Advances in Geospatial Information, Collections & Archives, 9*(3), 220-238.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition, 5*(2), 199-220.

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies, 43*(5), 907-928.

Grubov, V. I. (2007). *Plants of central Asia - plant collection from China and Mongolia Vol. 14A: Compositae (Anthemideae)* (V. I. Grubov, Ed.). Florida, US: CRC Press.

Grüninger, M., & Fox, M. S. (1995). Methodology for the design and evaluation of ontologies. In *International Joint Conference on Artificial Intelligence* (pp. 1-10). Quebec, Canada.

Gupta, S. B., & Mittal, A. (2009). *Introduction to database management system*. New Delhi, India: University Science Press.

Haase, P., Lewen, H., Studer, R., Tran, D. T., Erdmann, M., d'Aquin, M., & Motta, E. (2008). The NeOn ontology engineering toolkit. In *WWW 2008 Developers Track* (pp. 1-3). Beijing, China.

Hamrick, J. L., Godt, M. J. W., & Sherman-Broyles, S. L. (1992). Factors influencing levels of genetic diversity in woody plant species. In W. T. Adams, S. H. Strauss, D. L. Copes & A. R. Griffin (Eds.), *Population genetics of forest trees* (pp. 95-124). Dordrecht, Netherlands: Springer.

Hanum, F., & Hamzah, N. (1999). The use of medicinal plant species by the Temuan tribe of Ayer Hitam forest, Selangor, Penisular Malaysia. *Pertanika Journal Tropical Agriculture Science, 22*(2), 85-94.

Hare, J. S., Sinclair, P. A. S., Lewis, P. H., Martinez, K., Enser, P. G. B., & Sandom, C. J. (2006). Bridging the semantic gap in multimedia information retrieval: Top-down and bottom-up approaches. In P. Bouquet, R. Brunelli, J. P. Chanod, C. Niederée & H. Stoermer (Eds.), *European Semantic Web Conference, Budva, Montenegro* (pp. 45-60).

Harper, L., Gardiner, J., Andorf, C., & Lawrence, C. J. (2016). MaizeGDB: The maize genetics and genomics database. In D. Edwards (Ed.), *Plant bioinformatics: Methods and protocols* (pp. 187-202). New York, US: Springer.

Harris, R. W., Clark, J. R., & Matheny, N. P. (2003). *Arboriculture: Integrated management of landscape trees, shrubs, and vines* (4th ed.). New Jersey, US: Pearson Education.

Hearst, M. A., Laskowski, P., & Silva, L. (2016). Evaluating information visualisation via the interplay of heuristic evaluation and question-based scoring. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5028-5033). San Jose, California: ACM.

Hebeler, J., Fisher, M., Blace, R., & Perez-Lopez, A. (2009). *Semantic web programming*. Indianapolis, Indiana: Wiley Publishing.

Hepp, M. (2008). GoodRelations: An ontology for describing products and services offers on the web. In *International Conference on Knowledge Engineering and Knowledge Management* (pp. 329-346). Berlin, Heidelberg: Springer.

Hill, S. M., Heiser, L. M., Cokelaer, T., Unger, M., Nesser, N. K., Carlin, D. E., . . . Mukherjee, S. (2016). Inferring causal molecular networks: Empirical assessment through a community-based effort. *Nature Methods, 13*(4), 310-322.

Hobbs, C., & Foster, S. (2002). *A Peterson field guide to western medicinal plants and herbs.* New York, US: Houghton Mifflin.

Holsapple, C. W., & Joshi, K. D. (2002). A collaborative approach to ontology design. *Communications of the ACM, 45*(2), 42-47.

Hong, D. Y., & Blackmore, S. (2015). *Plants of China: A companion to the flora of China*. Cambridge, UK: Cambridge University Press.

Hu, G., & Gao, Q. (2009). An interactive image feature visualisation system for supporting CBIR study. In *International Conference Image Analysis and Recognition (ICIAR) 2009* (pp. 239-247). Berlin, Heidelberg: Springer.

Huala, E., Dickerman, A. W., Garcia-Hernandez, M., Weems, D., Reiser, L., LaFond, F., . . . Rhee, S. Y. (2001). The Arabidopsis information resource (TAIR): A comprehensive database and web-based information retrieval, analysis, and visualisation system for a model plant. *Nucleic Acids Research, 29*(1), 102-105.

Huang, J., Dang, J., Borchert, G. M., Eilbeck, K., Zhang, H., Xiong, M., . . . Tan, M. (2014). OMIT: Dynamic, semi-automated ontology development for the microRNA domain. *PLoS One, 9*(7).

Huang, T., Yu, Z., Lin, X., Jiang, L., & Zhao, D. (2018). A distributed CBIR system based on improved SURF on Apache Spark. In *IT Convergence and Security 2017* (pp. 147-155). Singapore: Springer.

Huerta-Cepas, J., Serra, F., & Bork, P. (2016). ETE 3: Reconstruction, analysis, and visualisation of phylogenomic data. *Molecular Biology and Evolution, 33*(6), 1635-1638.

Huey, F. Y., Ward, M. O., & Rundensteiner, E. A. (1999). Hierarchical parallel coordinates for exploration of large datasets. In *Proceedings Visualization '99* (pp. 43-508). California, USA: IEEE.

Hughes, S. (2006). Opinion piece: Genomics and crop plant science in Europe. *Plant Biotechnology Journal, 4*(1), 3-5.

Huson, D. H., Richter, D. C., Rausch, C., Dezulian, T., Franz, M., & Rupp, R. (2007). Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics, 8*(1), 460.

Hussain, A. G., Mohd Noor, N., & Hussin, K. (2015). *Nature's medicine: A collection of medicinal plants from Malaysia's rainforest.* Malaysia: Landskap Malaysia.

Hutmacher, D. W. (2010). Biomaterials offer cancer research the third dimension. *Nature Materials, 9*(2), 90-93.

Hyvönen, E., Styrman, A., & Saarela, S. (2002). Ontology-based image retrieval. In *Proceedings of XML Finland 2002 Conference* (pp. 15-27). Helsinki, Finland: HIIT Publications.

Iglesias-Martinez, L. F., Kolch, W., & Santra, T. (2016). BGRMI: A method for inferring gene regulatory networks from time-course gene expression data and its application in breast cancer research. *Scientific Reports, 6*(1), 1-12.

Ilic, K., Kellogg, E. A., Jaiswal, P., Zapata, F., Stevens, P. F., Vincent, L. P., . . . Rhee, S. Y. (2007). The plant structure ontology, a unified vocabulary of anatomy and morphology of a flowering plant. *Plant Physiol, 143*(2), 587.

Jaakkola, H., & Thalheim, B. (2011). Architecture-driven modelling methodologies. In *Proceedings of the 2011 Conference on Information Modelling and Knowledge Bases XXII* (pp. 97-116). Amsterdam: IOS Press.

Jadhav, S. M., & Patil, V. (2012). An effective content based image retrieval (CBIR) system based on evolutionary programming (EP). *International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)* (pp. 310-315). Ramanathapuram, India: IEEE.

Jaiswal, P., Avraham, S., Ilic, K., Kellogg, E. A., McCouch, S., Pujar, A., . . . Zapata, F. (2005). Plant Ontology (PO): A controlled vocabulary of plant structures and growth stages. *Comparative and Functional Genomics, 6*(7), 388-397.

Jensen, P. A., & Papin, J. A. (2014). MetDraw: Automated visualisation of genome-scale metabolic network reconstructions and high-throughput data. *Bioinformatics, 30*(9), 1327–1328.

Jin, T., Hou, X., Li, P., & Zhou, F. (2015). A novel method of automatic plant species identification using sparse representation of leaf tooth features. *PLoS One, 10*(10).

Joachimiak, M. P., Weisman, J. L., & May, B. C. (2006). JColorGrid: Software for the visualisation of biological measurements. *BMC Bioinformatics, 7*(1), 225.

Jones, R. L., & Wofford, B. E. (2013). *Woody plants of Kentucky and Tennessee: The complete winter guide to their identification and use*. Lexington, Kentucky: The University Press of Kentucky.

JSTOR, G. P. (2018). Global plants on JSTOR. Retrieved 24th Mar 2018, from https://plants.jstor.org/

Junker, B. H., Klukas, C., & Schreiber, F. (2006). VANTED: A system for advanced data analysis and visualisation in the context of biological networks. *BMC Bioinformatics, 7*(1), 109.

Kasprzyk, A., Keefe, D., Smedley, D., London, D., Spooner, W., Melsopp, C., . . . Birney, E. (2004). EnsMart: A generic system for fast and flexible access to biological data. *Genome Research, 14*(1), 160-169.

Kattge, J., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Bönisch, G., . . . Wirth, C. (2011). TRY – a global database of plant traits. *Global Change Biology, 17*(9), 2905– 2935.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., . . . Drummond, A. (2012). Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics, 28*(12), 1647-1649.

Keim, D. A. (2002). Information visualisation and visual data mining. *IEEE Transactions on Visualisation and Computer Graphics, 8*(1), 1-8.

Kelder, T., Pico, A. R., Hanspers, K., van Iersel, M. P., Evelo, C., & Conklin, B. R. (2009). Mining biological pathways using WikiPathways web services. *PLoS One, 4*(7).

Kerren, A., Kucher, K., Li, Y.-F., & Schreiber, F. (2017). BioVis explorer: A visual guide for biological data visualisation techniques. *PLoS One, 12*(11), 1-14.

Khare, C. P. (2007). *Indian medicinal plants: An illustrated dictionary.* New York, US: Springer-Verlag.

Khoshbakht, K., & Hammer, K. (2008). How many plant species are cultivated? *Genetic Resources and Crop Evolution, 55*(7), 925-928.

Kim, W. (1990). *Introduction to object-oriented databases.* Massachusetts, US: MIT Press.

Kim, W., Banerjee, J., Chou, H.-T., & Garza, J. F. (1990). Object-oriented database support for CAD. *Computer-Aided Design, 22*(8), 469-479.

King, Z. A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N. E., & Palsson, B. O. (2015). Escher: A web application for building, sharing, and embedding data-rich visualisations of biological pathways. *PLoS Computational Biology, 11*(8).

Kirk, A., Timms, S., Rininsland, A., & Teller, S. (2016). *Data visualisation: Representing information on modern web*. Birmingham, UK: Packt Publishing.

Kleiberg, E., Wetering, H. v. d., & Wijk, J. J. v. (2001). Botanical visualisation of huge hierarchies. In *Symposium on Information Visualisation (INFOVIS) 2001* (pp. 87-94). San Diego, California: IEEE.

Kourmpetli, S., & Drea, S. (2014). The fruit, the whole fruit, and everything about the fruit. *Journal of Experimental Botany, 65*(16), 4491-4503.

Koutsomitropoulos, D. A., & Kalou, A. K. (2017). A standards-based ontology and support for big data analytics in the insurance industry. *ICT Express, 3*(2), 57-61.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., . . . Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research, 19*(9), 1639-1645.

Kumar, N., Belhumeur, P. N., Bisis, A., Jacobs, D. W., Kress, W. J., Lopez, I. C., & Soares, J. V. B. (2012). Leafsnap: A computer vision system for automatic plant species identification. In *European Conference on Computer Vision (ECCV)* (pp. 502-516). Berlin, Heidelberg: Springer.

Kunii, H. S. (1990). Graph data model. In *Graph data model: And its data language* (pp. 7-20). Tokyo, Japan: Springer Japan.

Kwak, M., Leroy, G., Martinez, J. D., & Harwell, J. (2013). Development and evaluation of a biomedical search engine using a predicate-based vector space model. *Journal of Biomedical Informatics, 46*(5), 929-939.

Lady Bird Johnson Wildflower Center. (2018). Lady Bird Johnson Wildflower Center. Retrieved 23rd Mar 2018, from https://www.wildflower.org/plants/.

Lee, T.-H., Kim, Y.-K., Pham, T. T. M., Song, S. I., Kim, J.-K., Kang, K. Y., . . . Nahm, B. H. (2009). RiceArrayNet: A database for correlating gene expression from transcriptome profiling, and its application to the analysis of coexpressed genes in rice. *Plant Physiol, 151*(1), 16-33.

Lei, W., Shen, H., Sheng, H., Jiaen, L., & Bo, X. (2008). An effective and efficient method for query by humming system based on multi-similarity measurement fusion. In *International Conference on Audio, Language and Image Processing* (pp. 471-475). Shanghai, China: IEEE.

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., . . . Rombauts, S. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research, 30*(1), 325-327.

Lewen, H., Tran, T., & Haase, P. (2007). On the role and application of ontologies in information systems. In *International Conference on Research, Innovation and Vision for the Future* (pp. 14-21). Hanoi, Vietnam: IEEE.

Liang, J., Chang, T. Y. P., & Chan, C. M. (1998). An object-oriented database management system for computer-aided design of tall buildings. *Engineering with Computers, 14*(4), 275-286.

Limited, I. E. S. (2010). *Introduction to database systems.* India: Pearson Education.

Liu, G.-H., & Yang, J.-Y. (2013). Content-based image retrieval using color difference histogram. *Pattern Recognition, 46*(1), 188-198.

Lohmann, S., Link, V., Marbach, E., & Negru, S. (2015). WebVOWL: Web-based visualisation of ontologies. In *International Conference on Knowledge Engineering and Knowledge Management* (pp. 154-158). Cham, Switzerland: Springer.

Long, L. R., Antani, S., Deserno, T. M., & Thoma, G. R. (2009). Content-based image retrieval in medicine: Retrospective assessment, state of the art, and future directions. *International Journal of Healthcare Information Systems and Informatics, 4*(1), 1-16.

Lu, R., & Jin, Z. (2002). Formal ontology: Foundation of domain knowledge sharing and reusing. *Journal of Computer Science and Technology, 17*(5), 535-548.

Lučić, V., Förster, F., & Baumeister, W. (2005). Structural studies by electron tomography: From cells to molecules. *Annual Review of Biochemistry, 74*(1), 833-865.

Maedche, A., & Staab, S. (2002). Measuring similarity between ontologies. In *International Conference on Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web* (pp. 251-263). Siguenza, Spain: Springer.

Malaysia Botanical Garden. (2018). Malaysia Botanical Garden - plant info at your fingertip. Retrieved 5th July 2018, from http://mybotanicalgarden.my/

Mallik, J., Samal, A., & Gardner, S. L. (2010). A content based image retrieval system for a biological specimen collection. *Computer Vision and Image Understanding, 114*(7), 745-757.

Manning, G., Plowman, G. D., Hunter, T., & Sudarsanam, S. (2002). Evolution of protein kinase signaling from yeast to man. *Trends in Biochemical Sciences, 27*(10), 514-520.

MarkLogic. (2017). Best database for integrating data from silos. Retrieved 29th December 2017, from http://www.marklogic.com/.

Martinez-Cruz, C., Blanco, I. J., & Vila, M. A. (2012). Ontologies versus relational databases: Are they so different? A comparison. *Artificial Intelligence Review, 38*(4), 271–290.

Marx, V. (2013). The big challenges of big data. *Nature, 498*(7453), 255-260.

McCaleb, M. R. (1999). A conceptual data model of datum systems. *Journal of Research (NIST JRES), 104*(4), 349.

Metscher, B. D. (2009). MicroCT for comparative morphology: Simple staining methods allow high-contrast 3D imaging of diverse non-mineralized animal tissues. *BMC Physiology, 9*(1), 11.

Mew, T. W., Hibino, H., Savary, S., Vera Cruz, C. M., Opulencia, R., & Hettel, G. P. (2018). Rice diseases online resource. Retrieved 4th Mar 2018, from http://rice-diseases.irri.org/.

Mewes, H. W., Frishman, D., Gruber, C., Geier, B., Haase, D., Kaps, A., . . . Weil, B. (2000). MIPS: A database for genomes and protein sequences. *Nucleic Acids Research, 28*(1), 37-40.

Mishra, S. R. (2004). *Morphology of plants*. New Delhi, India: Discovery Publishing House.

Mukherjee, S., & Speed, T. P. (2008). Network inference using informative priors. *Proceedings of the National Academy of Sciences, 105*(38), 14313-14318.

Müller, H.-M., Kenny, E. E., & Sternberg, P. W. (2004). Textpresso: An ontology-based information retrieval and extraction system for biological literature. *PLoS Biology, 2*(11).

Murray, P., McGee, F., & Forbes, A. G. (2017). A taxonomy of visualisation tasks for the analysis of biological pathway data. *BMC Bioinformatics, 18*, 21.

MyBIS. (2017). Malaysia Biodiversity Information System (MyBIS). Retrieved 11th October 2017, from http://www.mybis.gov.my/one/analysis.php

NARO Genebank Project (2018). NARO Genebank - database of plant diseases in Japan. Retrieved 3rd Mar 2018, from https://www.gene.affrc.go.jp/databases-micro_pl_diseases_en.php

Natale, D. A., Arighi, C. N., Barker, W. C., Blake, J., Chang, T.-C., Hu, Z., . . . Wu, C. H. (2007). Framework for a protein ontology. *BMC Bioinformatics, 8*, 1.

National Research Council Committee, (2005). On the nature of biological data. In J. C. Wooley & H. S. Lin (Eds.), *Catalyzing inquiry at the interface of computing and biology* (pp. 35-56). Washington, DC: The National Academies Press.

National Research Council Committee, (1992). Why plant-biology research today? In *Plant biology research and training for the 21st century* (pp. 15-18). Washington, DC: National Academies Press.

Natural Resources Conservation Service (2017). The PLANTS Database. Retrieved 18th October 2017, from http://plants.usda.gov

Nielsen, C. B., Cantor, M., Dubchak, I., Gordon, D., & Wang, T. (2010). Visualizing genomes: Techniques and challenges. *Nature Methods, 7*, 5-15.

Nielsen, J. (1992). Finding usability problems through heuristic evaluation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 373-380). New York, US: ACM.

Nitoslawski, S., & Duinker, P. (2016). Managing tree diversity: A comparison of suburban development in two Canadian cities. *Forests, 7*(6), 119.

Normah, M. N., Chin, H. F., & Reed, B. M. (2013). *Conservation of tropical plant species.* New York, US: Springer-Verlag.

Noy, N. F., & McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology. Retrieved 1st November 2017, from http://protege.stanford.edu/publications/ontology_development/ontology101.pdf.

NParks Flora & Fauna Web. (2013). Flora fauna web - home. Retrieved 7th March 2018, from https://florafaunaweb.nparks.gov.sg/Home.aspx.

NParks Flora & Fauna Web. (2018). Flora fauna web - plant detail - Allamanda cathartica. Retrieved 15th Mar 2018, from https://florafaunaweb.nparks.gov.sg/special-pages/plant-detail.aspx?id=1303.

O'Donoghue, S. I., Gavin, A.-C., Gehlenborg, N., Goodsell, D. S., Hériché, J.K., Nielsen, C. B., . . . Wong, B. (2010). Visualizing biological data - now and in the future. *Nature Methods, 7*, 2-4.

Okonechnikov, K., Golosova, O., & Fursov, M. (2012). Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics, 28*(8), 1166–1167.

Padilla, E. (2016). *Substation automation systems: Design and implementation.* Hoboken, NJ: John Wiley & Sons.

Paredes-Valverde, M. A., Rodríguez-García, M. Á., Ruiz-Martínez, A., Valencia-García, R., & Alor-Hernández, G. (2015). ONLI: An ontology-based system for querying DBpedia using natural language paradigm. *Expert Systems with Applications, 42*(12), 5163-5176.

Pavlopoulos, G. A., Wegener, A.L., & Schneider, R. (2008). A survey of visualization tools for biological network analysis. *BioData Mining, 1*(1), 12.

Pellet. (2017). Pellet: An open source OWL DL reasoner for Java. Retrieved 20th April 2017, from https://github.com/stardog-union/pellet

Petryszak, R., Keays, M., Tang, Y. A., Fonseca, N. A., Barrera, E., Burdett, T., . . . Brazma, A. (2016). Expression Atlas update - an integrated database of gene and protein expression in humans, animals and plants. *Nucleic Acids Research, 44*(1), 746-752.

PFAF. (2017). PFAF. Retrieved 19th October 2017, from http://www.pfaf.org/user/Default.aspx.

Ponniah, P. (2007). *Data modeling fundamentals: A practical guide for IT professionals*. Hoboken, New Jersey: John Wiley & Sons.

Porzel, R., & Malaka, R. (2004). A task-based approach for ontology evaluation. In *Proceedings of European Conference on Artificial Intelligence (ECAI) Workshop on Ontology Learning and Population* (pp. 1-6). Valencia, Spain.

Prabhu, C. S. R. (2011). *Object-oriented database systems : Approaches and architectures* (3rd ed.). New Delhi, India: PHI Learning Private Limited.

Prieto-Diaz, R. (2003). A faceted approach to building ontologies. In *International Conference on Information Reuse and Integration* (pp. 458-465). Las Vegas: IEEE.

Protégé. (2017). Protégé. Retrieved 3rd February 2017, from https://protege.stanford.edu/

Quinn, G. P., & Keough, M. J. (2002). E*xperimental design and data analysis for biologists.* Cambridge, UK: Cambridge University Press.

Ray, C. (2009). Mobile databases and object-oriented DBMS. In *Distributed database systems* (pp. 239-236). India: Pearson Education.

Reece, J. B., Urry, L. A., Cain, M. L., Isserman, S. A., Minorsky, P. V., & Jackson, R. B. (2013). *Campbell biology* (10th ed.). New Jersey, US: Prentice Hall.

Robinson, I., Webber, J., & Eifrem, E. (2015). *Graph databases: New opportunities for connected data* (2nd ed.). California, US: O'Reilly Media Inc.

Roussey, C., Pinet, F., Kang, M. A., & Corcho, O. (2011). An introduction to ontologies and ontology engineering. In G. Falquet, C. Métral, J. Teller & C. Tweed (Eds.), *Ontologies in urban development projects* (pp. 9-38). London, UK: Springer.

Rowley, J. E., & Hartley, R. J. (2008). *Organizing knowledge: An introduction to managing access to information* (4th ed.). Hampshire, England: Ashgate Publishing Ltd.

Ruan, W., Hou, H., & Hu, Z. (2017). Detecting dynamics of hot topics with alluvial diagrams: A timeline visualisation. *Journal of Data and Information Science, 2*(3), 37-48.

Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., & Lorensen, W. E. (1991). *Object-oriented modeling and design.* New Jersey, US: Prentice Hall.

Rumpf, T., Mahlein, A. K., Steiner, U., Oerke, E. C., Dehne, H. W., & Plümer, L. (2010). Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. *Computers and Electronics in Agriculture, 74*(1), 91-99.

Sage, A. P., & Rouse, W. B. (2009). *Handbook of systems engineering and management* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons.

Said, I., Omar, Z., & Cheng, L. F. (2001). *Plant material booklet: Wayside trees of Malaysia* (Vol. 2). Johor, Malaysia: Penerbit UTM.

Sakr, S., & Gaber, M. M. (2014). *Large scale and big data: Processing and management.* Florida, US: CRC Press.

Saraiya, P., North, C., & Duca, K. (2005). Visualizing biological pathways: Requirements analysis, systems evaluation and research agenda. *Information Visualization, 4*(3), 191-205.

Şaykol, E., Güdükbay, U., & Ulusoy, Ö. (2005). A histogram-based approach for object-based query-by-shape-and-color in image and video databases. *Image and Vision Computing, 23*(13), 1170-1180.

Schmid, B., Schindelin, J., Cardona, A., Longair, M., & Heisenberg, M. (2010). A high-level 3D visualisation API for Java and ImageJ. *BMC Bioinformatics, 11*(1), 274.

Schweppe, D. K., Huttlin, E. L., Harper, J. W., & Gygi, S. P. (2018). BioPlex display: An interactive suite for large-scale AP-MS protein-protein interaction data. *Journal of Proteome Research, 17*(1), 722-726.

Secrier, M., Pavlopoulos, G. A., Aerts, J., & Schneider, R. (2012). Arena3D: Visualizing time-driven phenotypic differences in biological systems. *BMC Bioinformatics, 13*(1), 45.

Sedova, M., Jaroszewski, L., & Godzik, A. (2015). Protael: Protein data visualisation library for the web. *Bioinformatics, 32*(4), 602-604.

Segaran, T., Evans, C., & Taylor, J. (2009). *Programming the semantic web: Build flexible applications with graph data.* California, US: O'Reilly Media Inc.

Sen, T. Z., Harper, L. C., Schaeffer, M. L., Andorf, C. M., Seigfried, T. E., Campbell, D. A., & Lawrence, C. J. (2010). Choosing a genome browser for a model organism database: Surveying the maize community. *Database, 2010*, 1-32.

Senturia, S. D., Harris, R. M., Johnson, B. P., Kim, S., Nabors, K., Shulman, M. A., & White, J. K. (1992). A computer-aided design system for microelectromechanical systems (MEMCAD). *Journal of Microelectromechanical Systems, 1*(1), 3-13.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualisations. In *Symposium on Visual Languages* (pp. 336-343). Boulder, Colorado: IEEE.

Sikos, L. (2015). *Mastering structured data on the semantic web: From HTML5 microdata to linked open data.* New York, US: Apress.

Singh, N. P., & Gupta, C. S. (2014). *Relational database management systems*. New Delhi, India: Abhishek Publications.

Singh, V., & Misra, A. K. (2017). Detection of plant leaf diseases using image segmentation and soft computing techniques. *Information Processing in Agriculture, 4*(1), 41-49.

Siricharoen, W. V. (2008). *A software engineering approach to comparing ontology modeling with object modeling.* In *International Symposium on Computer Science and its Applications* (pp. 320-325). Hobart, Australia: IEEE.

Slenter, D. N., Kutmon, M., Hanspers, K., Riutta, A., Windsor, J., Nunes, N., . . . Willighagen, E. L. (2018). WikiPathways: A multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Research, 46*(1), 661-667.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., . . . Lewis, S. (2007). The OBO foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology, 25*(11), 1251-1255.

Sojoudi, S., Madani, R., Fazelnia, G., & Lavaei, J. (2014). Graph-theoretic algorithms for polynomial optimization problems. In *Conference on Decision and Control* (pp. 2257-2271). Los Angeles, California: IEEE.

Song, M. H., Lim, S. Y., Park, S. B., Kang, D. J., & Lee, S. J. (2006). Ontology-based automatic classification of web pages. In A. Abraham, B. d. Baets, M. Köppen & B. Nickolay (Eds.), *Applied soft computing technologies: The challenge of complexity* (pp. 483-493). Berlin, Heidelberg: Springer.

Spannagl, M., Noubibou, O., Haase, D., Yang, L., Gundlach, H., Hindemitt, T., . . . Mayer, K. F. X. (2007). MIPSPlantsDB - plant database resource for integrative and comparative plant genome research. *Nucleic Acids Research, 35*, 834-840.

Sreetheran, M., Adnan, M. R., & Khairil Azuar, A. K. (2011). Street tree inventory and tree risk assessment of selected major roads in Kuala Lumpur, Malaysia. *Arboriculture & Urban Forestry, 37*(5), 226-235.

Styrman, A. (2005). *Ontology-based image annotation and retrieval.* (Master's thesis, University of Helsinki). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.98.6759&rep=rep1&type=pdf.

Suárez-Figueroa, M., Gómez-Pérez, A., & Villazón-Terrazas, B. (2009). How to write and use the ontology requirements specification document. In R. Meersman, T. Dillon & P. Herrero (Eds.), *Lecture notes in computer science* (pp. 966-982). Berlin, Heidelberg: Springer.

Sumathi, S., & Esakkirajan, S. (2007). *Fundamentals of relational database management systems*. Secaucus, New Jersey: Springer-Verlag.

Sure, Y., Erdmann, M., Angele, J., Staab, S., Studer, R., & Wenke, D. (2002). OntoEdit: Collaborative ontology development for the semantic web. In *International Semantic Web Conference (ISWC) 2002* (pp. 221-235). Berlin, Heidelberg: Springer.

Sure, Y., Staab, S., & Studer, R. (2004). On-to-knowledge methodology (OTKM). In S. Staab & R. Studer (Eds.), *Handbook on ontologies* (pp. 117-132). Berlin, Heidelberg: Springer.

Swartout, B., Patil, R., Knight, K., & Russ, T. (1996). Toward distributed use of large-scale ontologies. In *Proceedings of the Tenth Knowledge Acquisition for Knowledge-Based Systems Workshop* (pp. 138-148). Alberta, Canada: AAAI Press.

TDWG, B. I. S. (2018). Biodiversity information standards (TDWG). Retrieved 25th March 2018, from https://github.com/tdwg.

Tegarden, D. P. (1999). Business information visualisation. *Communications of the AIS, 1*(1), 1-38.

Telea, A. C. (2014). *Data visualisation: Principles and practice* (2nd ed.). Wellesley, Massachusetts: AK Peters Ltd.

Teller, S. (2013). *Data visualisation with d3.js*. Birmingham, UK: Packt Publishing Ltd.

Tesoriero, C. (2013). *Getting started with OrientDB*. Birmingham, UK: Packt Publishing Ltd.

The Plant List. (2013). Version 1.1. Retrieved 7th February 2018, from http://www.theplantlist.org/.

Topalis, P., Mitraka, E., Bujila, I., Deligianni, E., Dialynas, E., Siden-Kiamos, I., . . . Louis, C. (2010). IDOMAL: An ontology for malaria. *Malaria Journal, 9*(1), 230.

TopQuadrant, I. (2018). TopBraid composer standard edition | TopQuadrant Inc. Retrieved 20th April 2018, from https://www.topquadrant.com/tools/modeling-topbraid-composer-standard-edition/.

Tory, M., & Moller, T. (2004). Human factors in visualisation research. *IEEE Transactions on Visualisation and Computer Graphics, 10*(1), 72-84.

Tran, V. X., Tsuji, H., & Masuda, R. (2009). A new QoS ontology and its QoS-based ranking algorithm for web services. *Simulation Modelling Practice and Theory, 17*(8), 1378–1398.

Tsai, S. S., Chen, D., Takacs, G., Chandrasekhar, V., Vedantham, R., Grzeszczuk, R., & Girod, B. (2010). Fast geometric re-ranking for image-based retrieval. In *International Conference on Image Processing* (pp. 1029-1032). Hong Kong: IEEE.

UCONN. (2017). Plant database. Retrieved 15th October 2017, from http://hort.uconn.edu/index.php.

Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *Knowledge Engineering Review, 11*(2), 93-136.

USDA. (2017). USDA plants database. Retrieved 16th October 2017, from https://plants.usda.gov/java/.

USM. (2018). Plant database. Retrieved 7th July 2018, from http://www.amdi.usm.my/index.php/31-plantdatabase.

Valarakos, A. G., Paliouras, G., Karkaletsis, V., & Vouros, G. (2004). Enhancing ontological knowledge through ontology population and enrichment. In *International Conference on Knowledge Engineering and Knowledge Management* (pp. 144-156). Berlin, Heidelberg: Springer.

van der Vet, P. E., & Mars, N. J. I. (1998). Bottom-up construction of ontologies. *IEEE Transactions on Knowledge and Data Engineering, 10*(4), 513-526.

van Stan, J. T., Jarvis, M. T., & Levia, D. F. (2010). An automated instrument for the measurement of bark microrelief. *IEEE Transactions on Instrumentation and Measurement, 59*(2), 491-493.

Velasco-Garcia, M. N., & Mottram, T. (2003). Biosensor technology addressing agricultural problems. *Biosystems Engineering, 84*(1), 1-12.

W3C (2013). SPARQL 1.1 overview. Retrieved 9th April 2018, from https://www.w3.org/TR/sparql11-overview/.

Walls, R. L., Athreya, B., Cooper, L., Elser, J., Gandolfo, M. A., Jaiswal, P., . . . Stevenson, D. W. (2012). Ontologies as integrative tools for plant science. *Am J Bot, 99*(8), 1263-1275.

Wang, C., Reese, J. P., Zhang, H., Tao, J., Gu, Y., Ma, J., & Nemiroff, R. J. (2015). Similarity-based visualisation of large image collections. *Information Visualisation*, *14*(3), 183-203.

Wang, J., Duncan, D., Shi, Z., & Zhang, B. (2013). Web-based gene set analysis toolkit (WebGestalt): Update 2013. *Nucleic Acids Research, 41*(1), 77-83.

Wang, S., Pandis, I., Wu, C., He, S., Johnson, D., Emam, I., . . . Guo, Y. (2014). High dimensional biological data retrieval optimization with NoSQL technology. *BMC Genomics, 15*(8), 3.

Ware, C. (2012). *Information visualisation: Perception for design* (3rd ed.). Waltham, Massachusetts: Morgan Kaufmann Publishers Inc.

Isson, C. S. (2015). *System engineering analysis, design, and development: Concepts, principles, and practices*. New Jersey, US: John Wiley & Sons.

Webb, R. (1998). Urban forestry in Kuala Lumpur, Malaysia. *Arboricultural Journal, 22*(3), 287-296.

Wei, P., Ward, M. O., & Rundensteiner, E. A. (2004). Clutter reduction in multi-dimensional data visualization using dimension reordering. In *Symposium on Information Visualization* (pp. 89-96). Texas, USA: IEEE.

Whetzel, P. L., Noy, N. F., Shah, N. H., Alexander, P. R., Nyulas, C., Tudorache, T., & Musen, M. A. (2011). BioPortal: Enhanced functionality via new web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Research, 39*, 541-545.

Williams, A. J., Harland, L., Groth, P., Pettifer, S., Chichester, C., Willighagen, E. L., . . . Mons, B. (2012). Open PHACTS: Semantic interoperability for drug discovery. *Drug Discovery Today, 17*(21), 1188-1198.

Willis, K. J. (2017). *State of the world's plants 2017*. Kew, Richmond: Royal Botanic Gardens.

Wu, H.H., & Bello, J. P. (2010). Audio-based music visualisation for music structure analysis. In *Proceedings of Sound and Music Computing Conference (SMC)* (pp. 11). Barcelona, Spain: Springer.

Yazici, A., & George, R. (2013). *Fuzzy database modeling* (Vol. 26). New York, US: Springer-Verlag.

Ying, L., Dengsheng, Z., Guojun, L., & Wei-Ying, M. (2005). Region-based image retrieval with high-level semantic color names. In *International Multimedia Modelling Conference* (pp. 180-187). Melbourne, Australia: IEEE.

Yue, J., Li, Z., Liu, L., & Fu, Z. (2011). Content-based image retrieval using color and texture fused features. *Mathematical and Computer Modelling, 54*(3), 1121-1127.

Yuk, M., & Diamond, S. (2014). *Data visualisation for dummies*. Hoboken, New Jersey: John Wiley & Sons Inc.

Zhiguo, D., & Zhengjie, D. (2010). Improved ontology ranking algorithm based on semantic web. In *International Conference on Ubi-Media Computing (U-Media)* (pp. 103-107). Jinhua, China: IEEE.

Zhou, L. (2007). Ontology learning: State of the art and open issues. *Information Technology and Management, 8*(3), 241-252.

Zhou, M., Liu, J., & Zheng, Y. (2018). Web pages ranking with domain ontology. In Park J., Loia V., Yi G., Sung Y. (Eds.), *Advances in computer science and ubiquitous computing: International Conference on Computer Science and its Applications* (pp. 516-521). Singapore: Springer.

Zhou, X., Wu, Z., Yin, A., Wu, L., Fan, W., & Zhang, R. (2004). Ontology development for unified traditional Chinese medical language system. *Artificial Intelligence in Medicine, 32*(1), 15-27.

Zoss, A. (2018). Visualisation types - data visualisation. Retrieved 21st Mar 2018, from https://guides.library.duke.edu/datavis/vis_types.

Zuo, Y., Yu, G., & Ressom, H. W. (2015). Integrating prior biological knowledge and graphical LASSO for network inference. In *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1543-1547). Washington, DC: IEEE.

LIST OF PUBLICATIONS AND PAPERS PRESENTED

PUBLICATION

1. **Mohamad-Matrol, A. A.**, Chang, S. W., & Abu, A. (2018). Plant data visualisation using network graphs. *PeerJ*, *6*, e5579

PAPERS PRESENTED

2. **Mohamad-Matrol A. A.**, Chang, S. W., & Abu A. (2017). *Plant Data Annotation and Retrieval using Semantic*. Paper presented at the 12th International Symposium in Science and Technology, 14-16th August 2017, Universiti Sains Malaysia, Pulau Pinang, Malaysia.

3. **Mohamad-Matrol A. A.**, Chang, S. W., & Abu A. (2017). *Presentation of Plant Data in Graphical Form.* Paper presented at the 22nd Biological Sciences Graduate Congress, 19-21st December 2017, National University of Singapore, Singapore.