DATA MANAGEMENT IN BIODIVERSITY

KHO SOON JYE

FACULTY OF SCIENCE

UNIVERSITI MALAYA

KUALA LUMPUR

2014

DATA MANAGEMENT IN BIODIVERSITY

KHO SOON JYE

(SGJ130004)

RESEARCH REPORT SUBMITTED TO

INSTITUTE OF BIOLOGICAL SCIENCES

FACULTY OF SCIENCE

UNIVERSITI MALAYA

IN PARTIAL FULFILMENT

OF THE REQUIREMENT FOR

THE DEGREE OF MASTER OF BIOINFORMATICS

2014

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: KHO SOON JYE (I.C No: 910803-10-6371)

Registration/Matric No: SGJ130004

Name of Degree: MASTER OF BIOINFORMATICS

Data Management in Biodiversity

Field of Study: BIOINFORMATICS

I do solemnly and sincerely declare that:
(1) I am the sole author/write 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 form, 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:

ABSTRACT

Biodiversity describes the variety of life found on Earth and it is a broad area to explore. In this field, a huge amount data has been collected by researchers. Thus, it is essential to manage those information with usage of computerized software, e.g. database. A database that stores biodiversity information is known as biodiversity database.

Malaysia is one of 17 mega-diversity countries. Its rich biodiversity has important economic, technological and social implications to the nation. In 16th April 1998, Ministry of Science, Technology and Environment declared Malaysia's national policy on biological diversity, to conserve Malaysia's biodiversity and to utilize it in a sustainable manner. The policy listed out 15 action plans and one of those plans encourages the expansion of scientific knowledge base.

A database framework is developed to store biodiversity information, along with the construction of information retrieval portal. The framework describes the overall architecture of the system, requirements of system, relational design of database, and user interface design of portal. Using the same framework, three databases were built due to different data schemes. Each database is to store biodiversity information of ginger, fern, and bat species in Malaysia respectively. The database development involves data cleaning and validation on three data files collected from biodiversity experts, followed by the construction of relational database system with the validated data.

An online information retrieval portal was also constructed using the PHP language. The portal is equipped with fundamental functions: information search function, administrative function and reporting function. The development of the database system is to help researchers systematically and conveniently manage the data, which helps in expansion of scientific base in Malaysia and thus, achieve the policy statement.

However, the system developed is introductory and there is still a lot of improvements work could be done. These improvement works will be discussed in this project report, hoping to further increase the effectiveness of system in achieving the objectives.

ABSTRAK

Biodiversiti ialah kepelbagaian spesies di Bumi dan merupakan satu bidang yang luas untuk diterokai. Terdapat banyak data yang telah dikumpulkan oleh para penyelidik di dalam bidang ini. Oleh itu aplikasi perisian komputer seperti pangkalan data adalah amat penting terutama di dalam bidang pengurusan maklumat. Pangkalan data yang dibina untuk menyimpan maklumat biodiversiti adalah dikenali sebagai pangkalan data biodiversiti.

Malaysia merupakan salah satu daripada 17 buah negara mega-diversiti. Kekayaan biodiveristi ini memberi impak yang signifikan kepada ekonomi, teknologi dan sosial negara. Pada 16hb April 1998, Kementerian Sains, Teknologi dan Alam Sekitar telah mengisytiharkan Dasar Kepelbagaian Biologi Kebangsaan, untuk pemuliharaan biodiversiti Malaysia dan memastikan bahawa komponennya digunakan secara lestari. Dasar ini menyenaraikan 15 pelan tindakan dan salah satu pelan ialah perkembangan pengetahuan saintifik.

Rangka kerja untuk pangkalan data ini telah dibina untuk menyimpan maklumat biodiversiti, bersama dengan pembinaan portal pencarian maklumat. Rangka kerja ini menerangkan struktur keseluruhan sistem, keperluan sistem, reka bentuk pangkalan data hubungan, dan reka bentuk antara muka pengguna portal. Kerana terdapat perbezaan pada skema data, tiga pangkalan data telah dibina dengan menggunakan rangka kerja yang sama. Setiap pangkalan data menyimpan maklumat biodiversiti untuk pelbagai spesies halia, paku-pakis, dan kelawar di Malaysia. Pembangunan pangkalan data ini melibatkan pembersihan dan pengesahan data terhadap tiga jenis fail data yang diperoleh dari pakar

biodiversiti, kemudian diikuti dengan pembinaan sistem pangkalan data hubungan berserta dengan data yang telahdisahkan.

Portal pencarian maklumat secara talian juga telah dibina berdasarkan bahasa pengaturcaraan PHP. Portal ini dilengkapi dengan beberapa fungsi asas: fungsi carian maklumat, fungsi pentadbiran dan fungsi pelaporan. Tujuan pembangunan sistem ini adalah untuk membantu penyelidik menguruskan data secara sistematik dan mudah. Ini akan membantu pengembangan pengetahuan saintifik di Malaysia dan mempercepatkan kecapaian kenyataan dasar.

Namun demikian, sistem yang dibangunkan ini masih berada dalam peringkat awal dan memerlukan banyak kerja penambahbaikan. Kerja-kerja penambahbaikan akan dibincangkan dalam laporan projek ini, supaya dapat meningkatkan keberkesanan sistem dalam mencapai objektif projek ini.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude and appreciation to my supervisor, Dr Sarinder Kaur, for her endless patience, guidance and valuable advices in this project. I would also like to thank my fellow lab mates, Najib, Bee Guan, Haris, Elham, Lee Kien and Aqilah who always share their knowledge and give their support throughout the project. Finally, I am grateful to my family for their moral support, love, and sacrifices. Thank you very much.

TABLE OF CONTENTS

LIST OF FIGURES

List of Figures, continued.

LIST OF TABLES

LIST OF SYMBOL AND ABBREVIATIONS

CPU	Central Processing Unit
EOL	Encyclopedia of Life
ERD	Entity Relationship Diagram
GB	Gigabyte
GBIF	Global Biodiversity Information Facility
GIS	Geographic Information System
GPS	Global Positioning System
ITIS	Integrated Taxonomy Information System
PDF	Portable Document Format
PHP	Hypertext Preprocessor
SDLC	System Development Life Cycle
SQL	Structured Query Language
UBD	University Brunei Darussalam
UM	University Malaya
UNIMAS	University Malaysia Sarawak

CHAPTER 1

INTRODUCTION

1.1 Background

Biological database is a large, organized library of life sciences information, collected from scientific experiments, published literature, high-throughput experiment technology, and computational analyses. It is associated with usage of computerized software to store, update, query, and retrieve information within the system (Kumar, 2005). Database provides a convenient method to store enormous amount of data and allows biologists or researchers to collect data, organize them systematically, analyze the data, and share information with other researchers efficiently.

Biological data is often comprised of complicated data. There are different types of biological data, e.g. DNA sequence, protein sequence, protein structure, taxonomy, and others. Development of a database should acknowledge the complexity of biological data and be capable in managing such complex data.

1.2 Taxonomy

Taxonomy is the practice and science of classification. In biology, taxonomy defines group of biological organisms on the basis of shared characteristics, giving names for those groups and organize those groups hierarchically. Taxonomy has been called “the world’s oldest profession”. It has been likely taking place as far as 1500 BC. In year 1735, a

Swedish botanist Carl Linnaeus (1707-1778) revolutionized modern taxonomy. His works implemented a standardized binomial naming system for animal and plant species. The system creates a more organized solution to the chaotic taxonomic literature at that time and resulted in the born of Linnaean system we have been using until now (Knapp, 2010).

A biologist who practices taxonomy is known as "taxonomist". Biological classification uses taxonomic ranks (in order of the highest to the lowest): Domain, Kingdom, Phylum, Class, Order, Family, Genus, and Species. The groups in the classification are known as taxa.

Domain is the highest rank in biological taxonomy. It was introduced in the 1960s, making the kingdom is no longer the highest rank since then. It was only two domains until Woese, Kandler and Wheelis (1990) proposed three domains system: Archaea, Bacteria, and Eucarya. Species is the lowest rank in biological taxonomy. It is defined as the largest group of organisms capable of interbreeding and producing fertile offspring. However, the definition is often inadequate and other measures are often used, such as similarity of DNA, morphology, or ecological niche (De Queiroz, 2007).

As such, the work of classification for a group of organisms often involves vast amount of data, not to mention a country with rich biodiversity contains an enormous amount of species. For example, it is estimated Brazil, the country with richest biodiversity has 170,000 to 210,000 species (Lewinsohn & Prado, 2005). To systematically classify all the species, it would generate a big data known as biodiversity information.

1.3 Biodiversity Informatics

Recent developments in information and communication technology are allowing new experiences in the integration, analysis and visualization of biodiversity information, leading to a new field of research, biodiversity informatics (Canhos, Souza, Giovannia & Canhos, 2004). It includes application of information technologies to the management, algorithmic exploration, analysis and interpretation of primary data regarding life, particularly at the species level of organization (Soberon & Peterson, 2004).

Biodiversity informatics cope with managing information of unnamed taxa that are produced by environmental samples or sequencing of samples, and also cover the computational problems specific to the names of biological entities. Algorithms are developed to cope with those problems. For example, Ch'ng (2009) developed a segmentation algorithm for entity interaction, which increases the efficiency of real-time simulation that models the biotic interactions of large population datasets. Besides, biodiversity informatics involve development of the syntax and semantics for publishing and integrating biodiversity information. Darwin Core archive is the data standard developed to improve interoperability (Wieczorek et al., 2012).

1.4 Biodiversity Databases

Biodiversity databases store biodiversity information such as taxonomy data, descriptive summary, and geographical information of living things. Several global and regional efforts are organizing and providing data for conservation and sustainable development research, including the Global Biodiversity Information Facility, the European Biodiversity

Information Network, and the Inter-American Biodiversity Information Network (Canhos et al., 2004).

Some biodiversity databases stores data of all species globally, while some is constructed based on variety families such as fishes, birds, insects, reptiles, and others.

(i) Global Biodiversity Information Facility (GBIF)[1]

The Global Biodiversity Information Facility (GBIF) is an international organization that focuses on making biodiversity data available via online website. They encourage and help institutions to publish data according to common standards. The data provided from institutions across the world will be published in their portal, making the data accessible and searchable via online. The mission of GBIF is to facilitate free and open access to biodiversity data worldwide to underpin sustainable development. The total number of collection can be shown in their homepage. As of November 2014, GBIF has published data of 1,454,694 species from 644 data publishers.

[1] http://www.gbif.org/

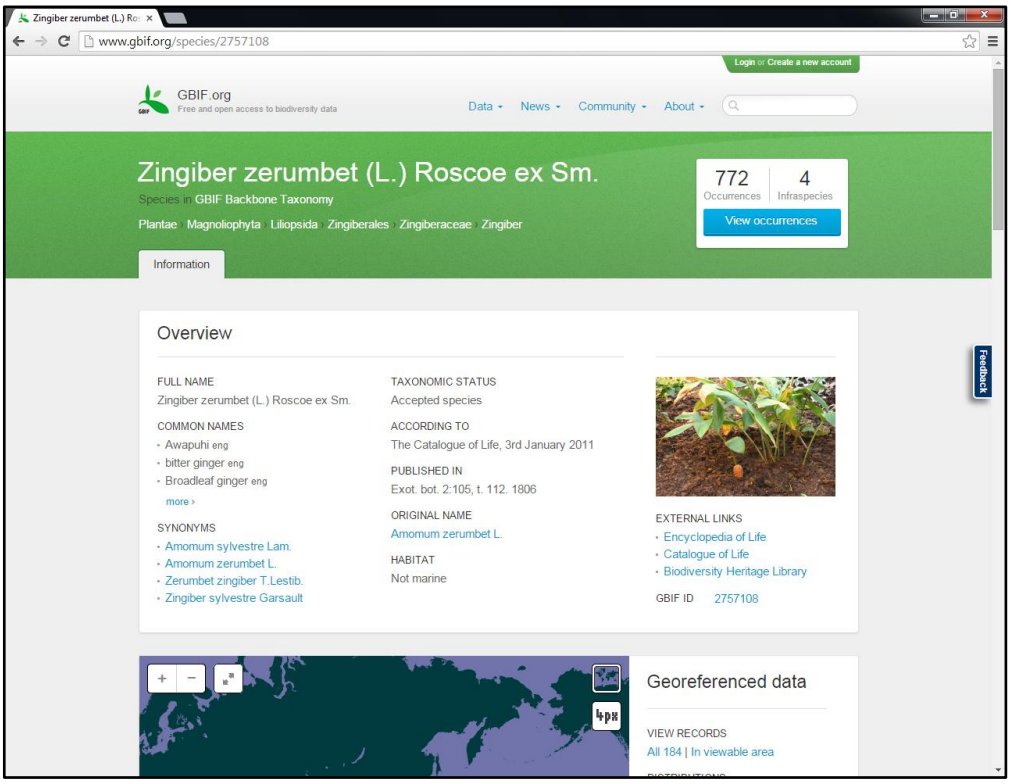


Figure 1.1: Global Biodiversity Information Facility (GBIF) result page.

(ii) Integrated Taxonomy Information System (ITIS)[2]

The Integrated Taxonomy Information System (ITIS) is American partnership of federal agencies. The goal of ITIS is to create an easily accessible database with reliable information on species names and their hierarchical classification. For each scientific name, ITIS will include the authority (author and date), taxonomic rank, associated synonyms and vernacular names where available, a unique taxonomic serial number, data source information (publications, experts, etc.) and data quality indicators. As of October 2014, ITIS contains 661,517 scientific names and 122,644 common names.

[2] http://www.itis.gov/

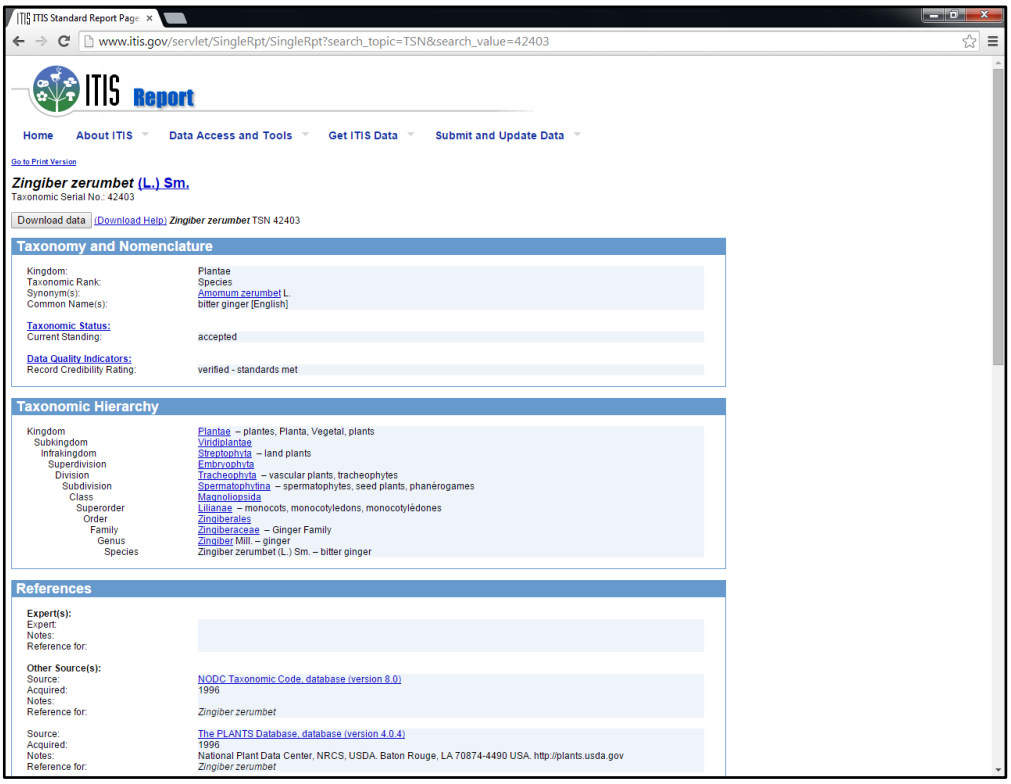


Figure 1.2: Integrated Taxonomy Information System (ITIS) result page.

(iii) Encyclopedia Of Life (EOL)[3]

The Encyclopedia of Life (EOL) is a free, online collaborative encyclopedia that collects and compiles biodiversity information from various institutions across the world. The mission of EOL to increase awareness and understanding of living nature through an Encyclopedia of Life that gathers, generates, and shares knowledge in an open, freely accessible and trusted digital resource. It has built a webpage for each taxon, which stores many information, including video, sound, images, graphics, as well as text. As of November 2014, it has a total of 1,362,137 pages.

[3] http://eol.org/

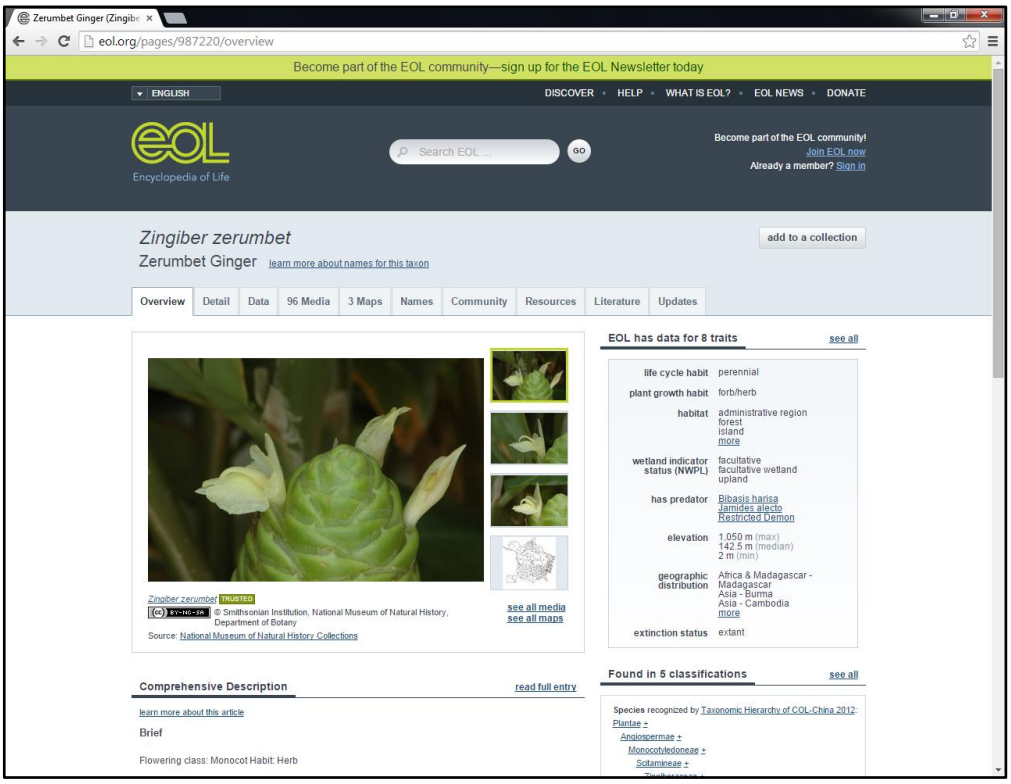


Figure 1.3: Encyclopedia Of Life (EOL) result page.

(iv) Kew Herbarium Catalogue[4]

Kew Herbarium Catalogue is a digital library of specimen records and images constructed by Royal Botanic Gardens Kew houses. As of November 2014, the Herbarium at the Royal Botanic Gardens Kew houses approximately 7 million specimens, collected from all around the world. Kew Herbarium Catalogue contains images of the specimens and information taken from their collection labels.

[4] http://www.kew.org/herbcat

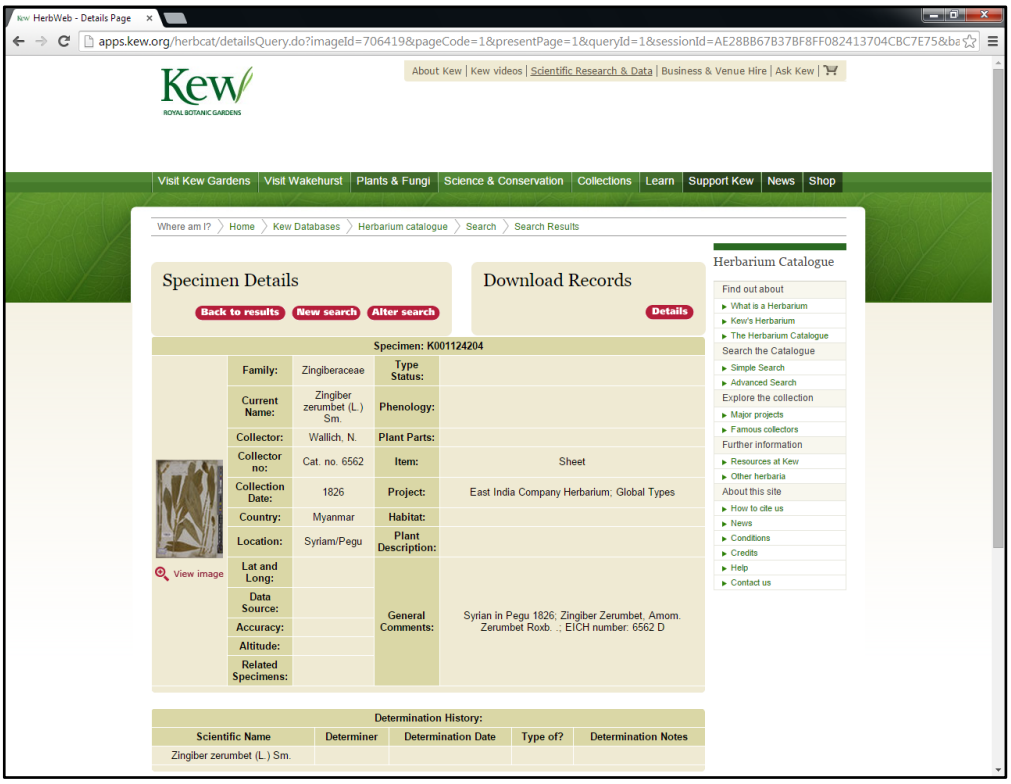


Figure 1.4: Kew Herbarium Catalogue result page.

(v) The Plant List[5]

The Plant List is a working list of all known plant species. It aims to be comprehensive for species of vascular plant (flowering plants, conifers, ferns and their allies) and of bryophytes (mosses and liverworts). It is an effort of Collaboration between the Royal Botanic Gardens, Kew and Missouri Botanical Garden. Version 1.1 contains 1,293,685 scientific plant names of which 350,699 are accepted species names. However, it does not include vernacular or common plant names.

[5] http://www.theplantlist.org/

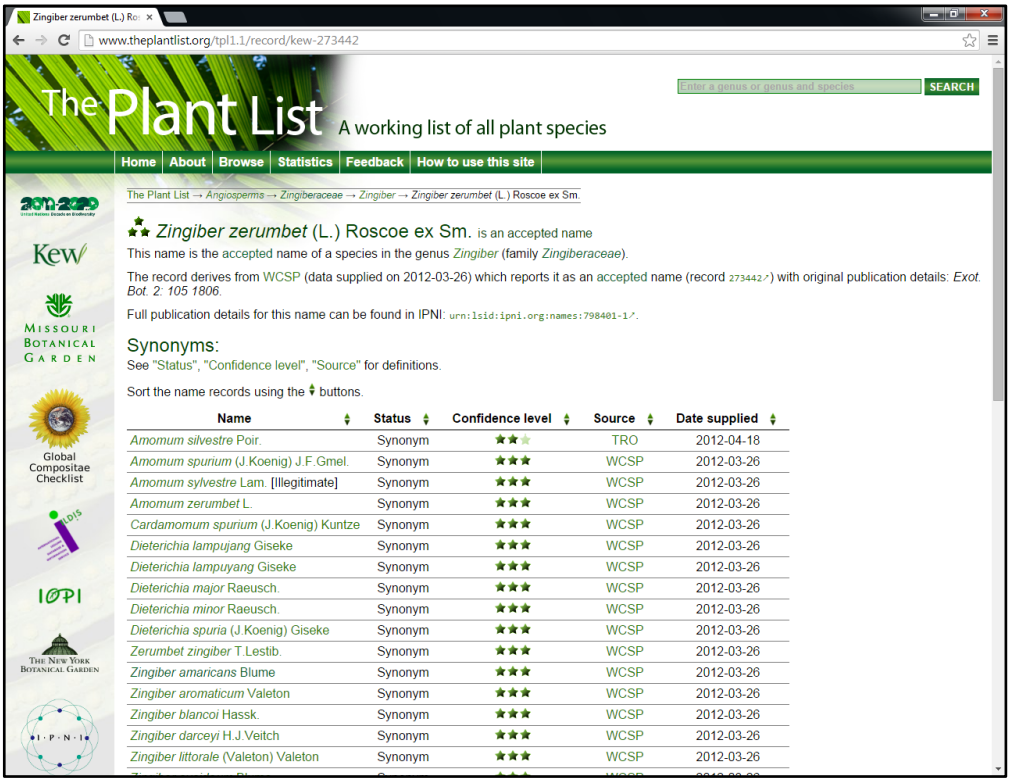


Figure 1.5: The Plant List result page.

(vi) Flora Malesiana[6]

Flora Malesiana is a systematic account of the flora of Malesia, the plant-geographical unit spanning six countries in Southeast Asia:Indonesia, Malaysia, Singapore, Brunei Darussalam, Philippines, and Papua New Guinea. It stores large amount of information (with literature references) on, e.g., taxonomy, variability, synonymy, typification, distribution, habitats and ecology, morphology and anatomy, phytochemistry, and uses. There are two series of flora in the database, seed plants and pteridophytes covering around 3,820 taxa belonging to 33 families.

[6] http://dev.e-taxonomy.eu/dataportal/flora-malesiana/

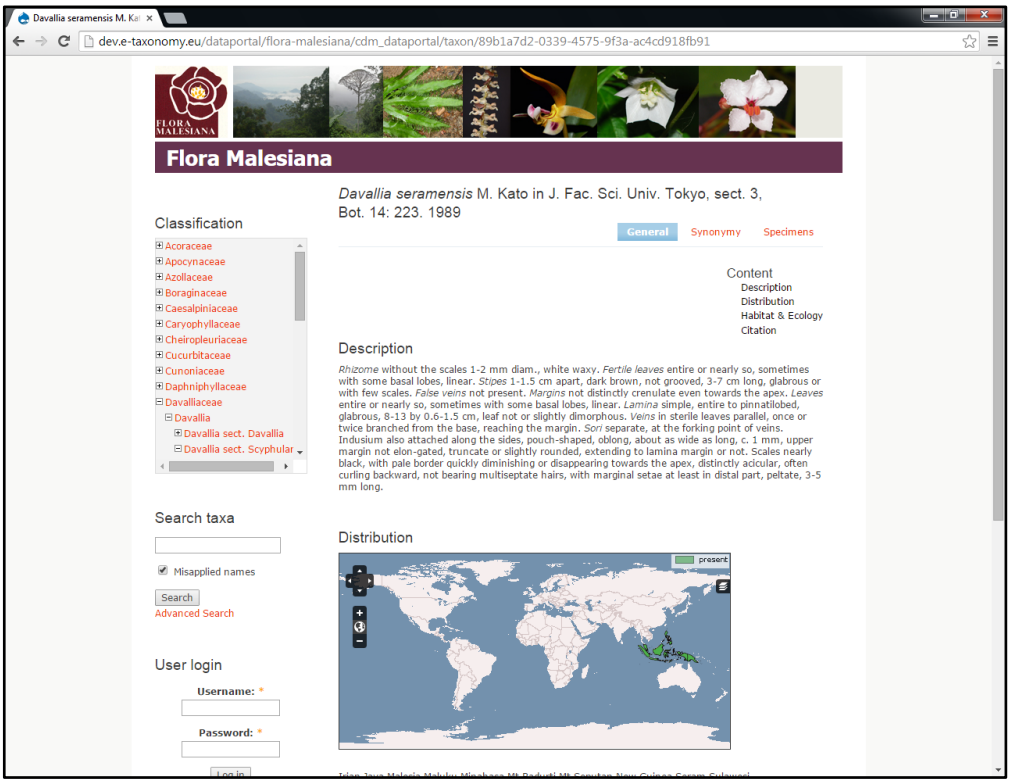


Figure 1.6: Flora Malesiana result page.

1.5 Importance of Biodiversity Informatics

The collection of biodiversity information plays a critical role in understanding biodiversity and ecosystem. There are few significance of biodiversity informatics that are not to be neglected.

1.5.1 Biodiversity and Environmental Change

Biodiversity informatics allows documentation of the biological diversity of life and demonstration of changes in the environment that have taken place through time (Baird, 2010). The presence or absence of a species as well as their frequency changes in a

geographic region over the course of time are documented by examining the records. If a gradual decline of an important species is observed, it may serve as an indicator whether a recovery program should be carried out to restore the population of the species. The collections that are well documented and deposited can offer proof that the program was successful, provide evidence that changes has taken place, and can help to effectively monitor rare, threatened and endangered species.

1.5.2 Invasive Alien Species and Biosecurity

Other than monitoring endangering species, biodiversity informatics is capable of revealing the invasive alien species that are threat for biodiversity. Increasing travel, trade, and tourism associated with globalization and expansion of the human population have facilitated intentional and unintentional movement of species beyond natural biogeographical barriers (Forest Research Institute Malaysia, 2011). These species are known as alien species and many of them have become invasive. Invasive alien species are known to cause biodiversity loss and lead to a nation's economic loss. Thus, it is critical to differentiate native from alien species, and respond rapidly to these threat, either to remove the species from a region, or suppress their population growth if unable to remove it.

Besides revealing invasive alien species, biodiversity informatics allows prediction of potential of a non-native species to be invasive alien species. Faulkner, Robertson, Rouget and Wilson (2014) has created and applied a simple, rapid methodology for developing invasive species watch lists. The watch lists are created using three predictors of three invasion success: history of invasion, environmental suitability and propagule pressure. The

authors claim that building the list may be an important step in developing biosecurity scheme, especially for resource poor regions.

1.5.3 Public Health and Wildlife Disease

There are many infections that manifest themselves in human and non-human populations, such as H1N1 Influenza, West Nile Virus, Lyme disease, tuberculosis, Ebola Virus, and SARS. For these and other zoonoses, approximately 70% of new important disease affecting human health are believed to have a wild animal source, and can have profound impacts on human health and economy (Blancou, Chomel, Belotto & Meslin, 2005).

Using the occurrence data from GBIF, Pigott et al. (2014) mapped the area potentially at risk from outbreaks of Ebola virus, based on the environmental niche of bat species believed to act as reservoir hosts of the disease. The authors determined the national population at risk, and claimed that this would be a strong rationale for improving, prioritizing, and stratifying surveillance for Ebola virus disease outbreaks and diagnostic capacity in these countries.

1.5.4 Economics and Regulatory Framework

In 16th April 1998, Ministry of Science, Technology and Environment (1998) declared Malaysia’s national policy on biological diversity. The policy statement is to conserve Malaysia's biological diversity and to ensure that its components are utilized in a sustainable manner for the continued progress and socio-economic development of the nation. To achieve that, the policy outlines 15 strategies. The first strategy is improve the

scientific knowledge base, which involve long term studies, research of species, and developing biodiversity database and effective information dissemination system. The development of biodiversity databases would ally the statement of policy.

1.5.5 Access and Benefit Sharing

The obvious significance of digitization of biodiversity information is dissemination of information globally via online. Before digitization, the information are held by institutes in developed countries, whereas researchers always do their field work at underdeveloped country. Access to information collected at those countries is limited. Thus, digitization allows researcher to remotely access to the information as long as there is internet connection. Besides, digitization allows sharing of information among researchers. The information could be integrated and analyzed to contribute in many study.

1.6 Research Objective

The aim of this research is to construct a database framework to store taxonomic information of gingers, ferns and bats in Malaysia and create a portal with information retrieval facilities. This would provide a platform for researchers to search, collect, store and share biodiversity information among researchers. The databases are expected to ally the statement of Malaysia's national policy on biological diversity, to help conserve Malaysia's biodiversity and to help utilize it in a sustainable manner.

The objectives of this study are:

(i) To build a database framework for storing, managing, visualizing and reporting of biodiversity sciences

(ii) To provide a platform of biodiversity information in Malaysia where information can be efficiently disseminated among researchers.

1.7 Research Scope

This study involves the building of database framework, which will be used in development of three databases: The Herbarium of Zingiberaceae where ginger species information is stored, Malaysian Indigenous Fern Database where fern species information is stored and Database of Bats Species where bat species information is stored. It focuses on the data preparation process on taxonomic information of species, design and implementation of databases and information retrieval portal.

CHAPTER 2

SYSTEM REQUIREMENTS AND ANALYSIS

2.1 System Development Life Cycle

System development life cycle (SDLC) methodology is used to develop the system in this study. It is a conceptual model describing process of planning, creating, testing, and deploying an information system. There are many types of model in SDLC and the model utilized in this study is waterfall model ("Selecting a development approach", 2008). This model is linear-sequential life cycle model, and very simple to understand and use. Each phase has to be completed before starting the next phase. Figure 2.1 shows the five phases involved in this study: preliminary analysis, system requirements and analysis, system design, system implementation, and system testing.

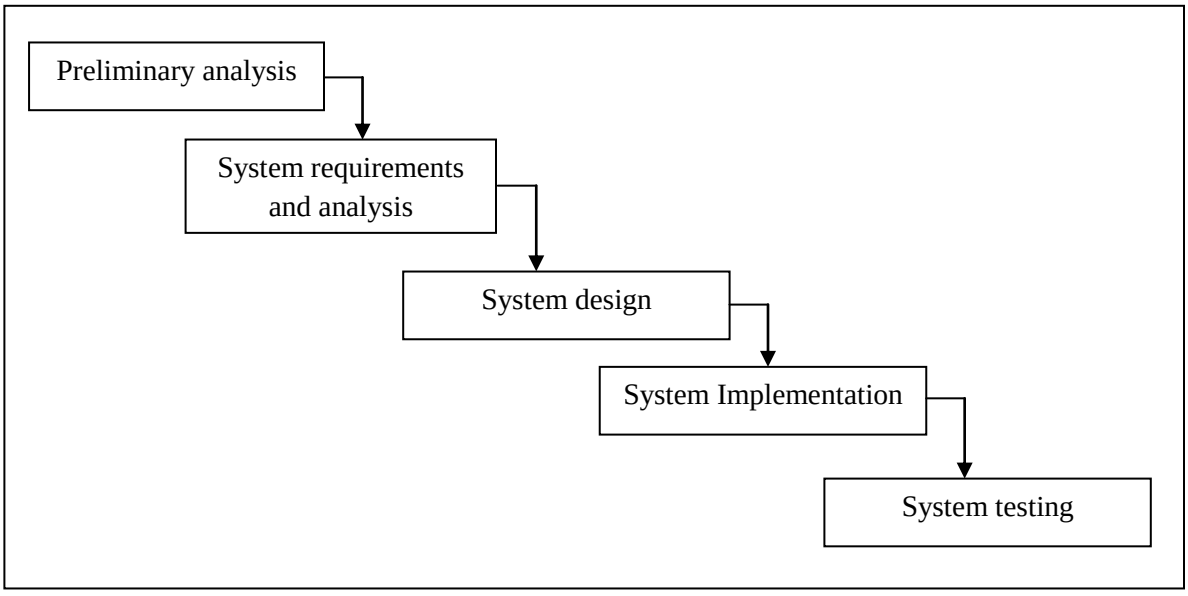


Figure 2.1: The flow of system development life cycle (SDLC) methodology.

2.1.1 Preliminary Analysis

This is the first phase in SDLC. Existing databases are studied and the information is usually gathered via online. The information gathered is essential to identify the scope of this study as well as its importance. This phase has been conducted and described in Chapter 1.

2.1.2 System Requirements and Analysis

During this phase, possible requirements of the system development and project basic architecture is analyzed. Research objectives, research scopes and project requirements are determined with the information gained from previous analysis. Requirements on functionality of system are determined and will be described in Chapter 3. Meanwhile, any problem aroused and suggestion for solution are listed out and act as a guidelines for the next phase.

2.1.3 System Design

The structure of system is designed in this phase. The design is based on the information studied from previous phase. This includes defining the relational database structure, to show how the data are linked in relational database environment. The module of user interface is also designed, to show how the process is involved in the system.

2.1.4 System Implementation

Functionalities and features of the system designed from previous stage is developed during this phase. The coding transforms the logical modules into physical modules and serves as the primary input for design elements described previously. The optimization of the codes is conducted for the purpose of resource saving.

2.1.5 System Testing

Source code generated is then subjected for testing. This phase reviews functionality of the system. The features in this system are checked and any error observed is reviewed. A testing checklist is created and will be filled up by users during this phase. The level of system effectiveness is also determined before the system is allowed to use.

2.2 Data and System Analysis

An analysis of the system was performed to identify the requirements needed by the system. Data analysis is performed to identify the appropriate system architecture for system development.

Three data files on ginger, fern and bat species were given by biodiversity experts (see Table 2.1).

Table 2.1: Data files, the name of professors who collected the data and their respective institute.

Data files	Name of Professor	Institute
Ginger species	Prof. Halijah Ibrahim	University Malaya (UM)
Fern species	Prof. Haji Mohamed Abd. Majid	University Brunei Darussalam (UBD)
Bat species	Prof. Dr. Mohd Tajuddin Abdullah	University Malaysia Sarawak (UNIMAS)

Three data files contain different data schemes. Due to limited data from the files and time limitation of this project, it is infeasible to produce a data integration module. Hence, a database framework will be produced and used to develop three databases for each data files. Data cleaning and validation will be performed before developing the databases.

2.3 Functional Requirements

The functionality of the system is described by what processes are involved, how the system works and what work flow is needed to produce desired output when users generate input. These functions include information search function, administrative functions (admin login, adding record, editing record and deleting record) and reporting functions (single species and multiple species). Table 2.2 describes specification of each function of the system.

Table 2.2: Functional specification of the database system.

Specification	**Description**
Information Search	**Outline** This function allows user to search desired information in the database. **Processes** 1. Select search menu. 2. Insert keywords in the search box. 3. If incomplete string is entered, suggestion box containing complete string will appear at the bottom of search box. 4. Users may click on the suggestion in suggestion box to complete the string and click Search, or click Search directly without choosing any option in suggestion box. 5. The system will perform search process in the database. 6. Related information will be displayed. **Pre-condition** None

Table 2.2, continued.

Specification	Description
Admin Login	**Outline** Users have to register to use the functions in the system for managing the database. **Processes** 1. Users enter username and password at the top banner of any page. 2. Click LOGIN button. 3. System will validate the users' login by matching with stored username and password in database. 4. If login is valid, users will be directed to Admin Home Page. Otherwise, error message will be displayed. **Pre-condition** None
Add Record	**Outline** This function allows users to add new record into database. **Processes** 1. Users select Add Record. 2. Add Record Form is displayed. 3. Users insert new information. Fields that are necessary to fill is validated by form validation. Any invalid input or blank input will prompt an error message. 4. Users click Add button. 5. The system stores new information into the database. **Pre-condition** Only admins are allowed to add record.

Table 2.2, continued.

Specification	Description
Delete Record	**Outline** This function allows user to delete existing record from database. **Processes** 1. Users select Edit/Delete Record and search for the desired record. 2. Users click Delete button. 3. The system will remove the record from database. **Pre-condition** Only admins are allowed to delete record.
Edit Record	**Outline** This function allows users to edit existing record in the database. **Processes** 1. Users select Edit/Delete Record and search for the desired record. 2. Users click Edit button. 3. Edit Record Form is displayed. 4. Users insert updated information. Fields that are necessary to fill is validated by form validation. Any invalid input or blank input will prompt an error message. 5. Users click Submit button. 6. The system will update the information in database. **Pre-condition** Only admins are allowed to edit record.

Table 2.2, continued.

Specification	Description
Reporting (Single Species)	**Outline** This function allows users to produce one-page PDF file containing information of a single species. **Processes** 1. Users search for desired species record and click to details page. 2. Click View in PDF File button. 3. A new tab with PDF file containing the information will appear. **Pre-condition** None
Reporting (Multiple Species)	**Outline** This function allows admins to produce PDF or excel file containing information of multiple species. **Processes** 1. Users click Reporting. 2. Reporting page is displayed. 3. Users choose which field and which species they wish to include in the file. 4. If users click Excel button, an excel file will be downloaded. 5. If users click PDF File button, a new tab with PDF file containing the information in table form will appear. **Pre-condition** Only admins are allowed to use this function.

CHAPTER 3

SYSTEM DESIGN AND IMPLEMENTATION

3.1 System Design

System design refers to the activities of designing the database structure, which it will be used to store and manage data. In this study, the database was built using MySQL, a relational database management system. The structure of data storage in this relational database is described using relational design. Beside relational design, user interface design was performed to outline the process involved in accessing the system.

3.1.1 Relational Design

A relational model describes the process of designing a relational database where the data is stored into one or more tables of rows and columns, with a unique key for each row. When designing a database, each data element and their relationship should be clearly designed to ensure the smoothness of implementation process that will be developed.

Entity relationship diagram (ERD) is a data model describing the entities of the system and the relationship between each entity. The main components of ERD are entities, attributes and relationships. Entity is something about which we store data, while attributes are data that describe the entities (Harrington, 2009). Relationships are assigned for attributes that are relevant to each other. Entity relationship diagram builds a conceptual schema that

allows us to see how the data are linked and what is the relationship between those linked data: one-to-one, one-to-many, or many-to-many.

The complete ERD for three databases in this study is shown in Figure 3.1, Figure 3.2 and Figure 3.3. Due to the different data schemes of data files, ERD of each database is different. These ERDs will be used in the implementation of the prototype of the actual database.

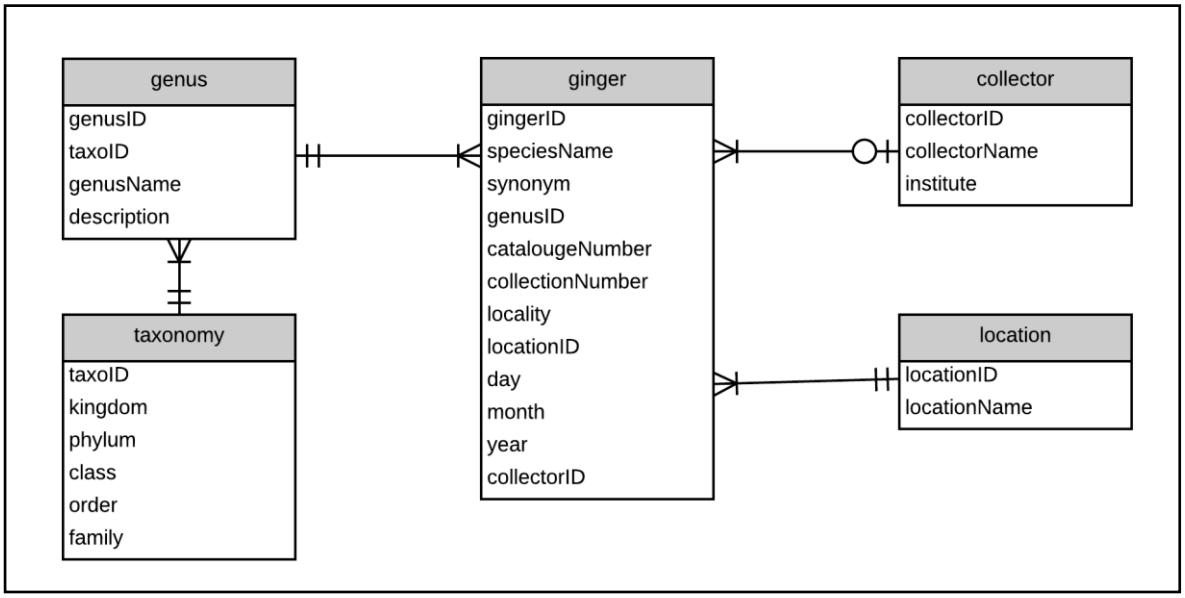


Figure 3.1: Entity relationship diagram used for developing ginger database.

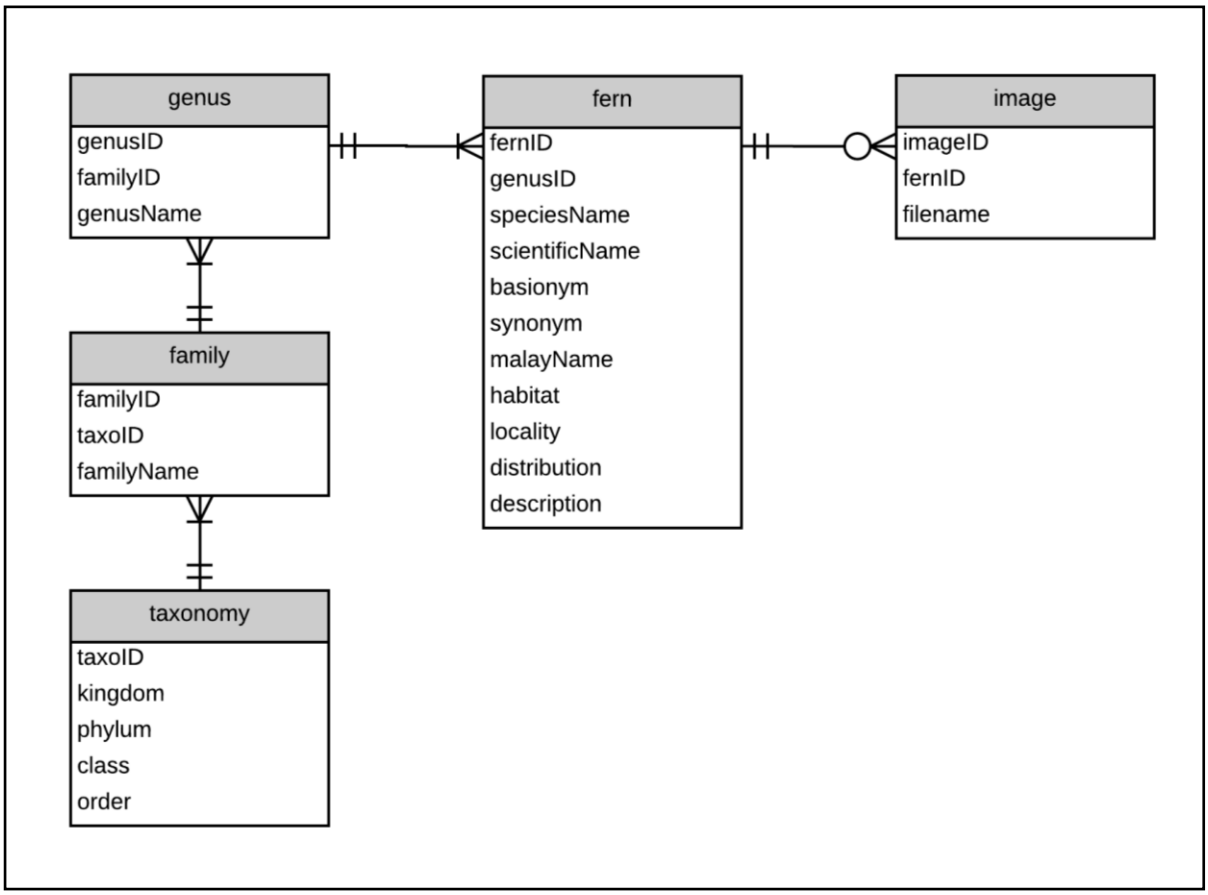


Figure 3.2: Entity relationship diagram used for developing fern database.

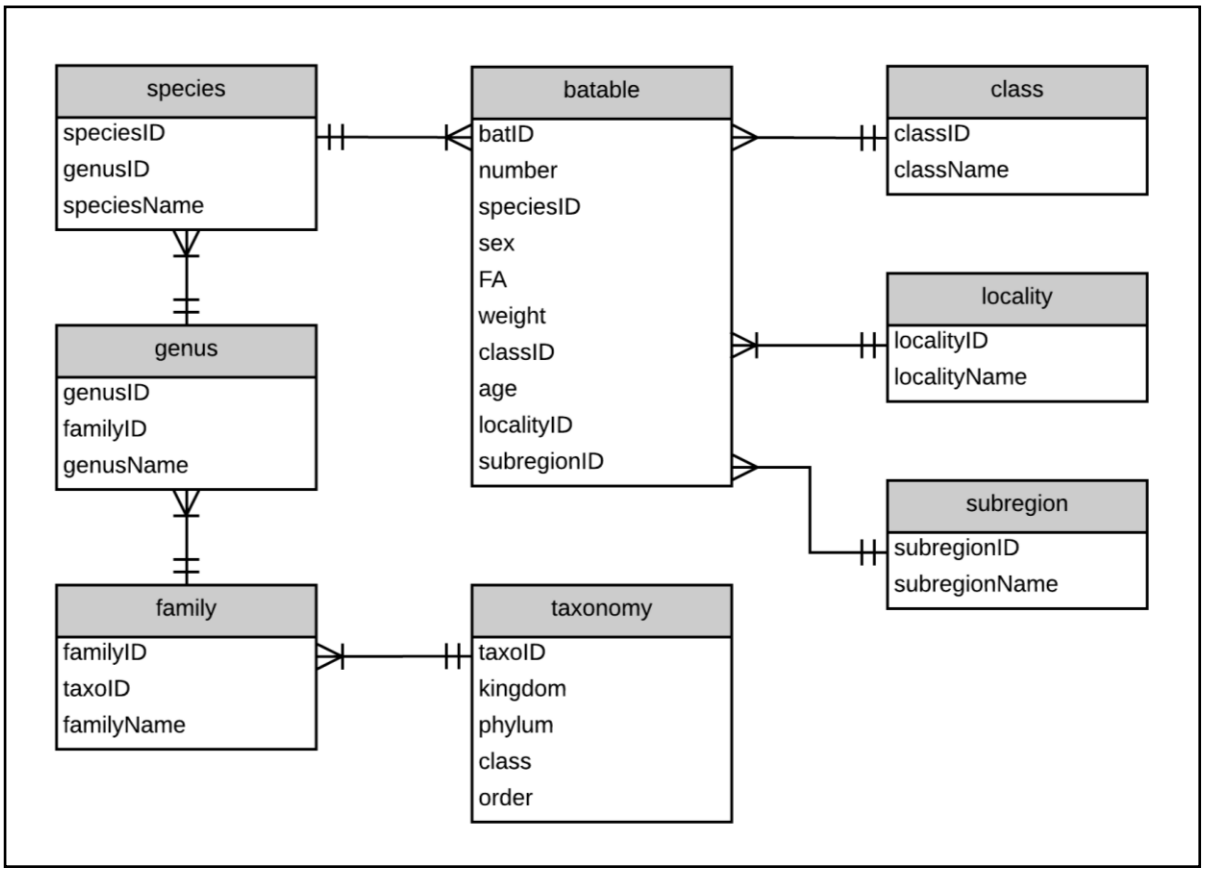


Figure 3.3: Entity relationship diagram used for developing bat database.

3.1.2 User Interface Design

User interface design describes the flow involved in the process of accessing the system. User interface is designed to draw the outline of the visual representation of the system before it is developed. The aim of designing the interface is to make user's interactions with system as simple and efficient as possible. It serves to facilitate the process of developing database system where required functionality is placed in the suitable part of the system. The development of database is based on database design and at the same time, based on user interface design too.

In this design, every users have access to the main page and search function. The search function can be accessed via the links provided in the main page. However, only the database administrators are allowed to manipulate data in the database. Each administrator is given a username and password, and is required to login to the system in order to add, remove or edit record. Login can be done via top banner on any webpage. Once logged in, administrator will be directed to admin home page, where they can manipulate the data in database and produce reports here. Figure 3.4 shows the user interface design structure for database system.

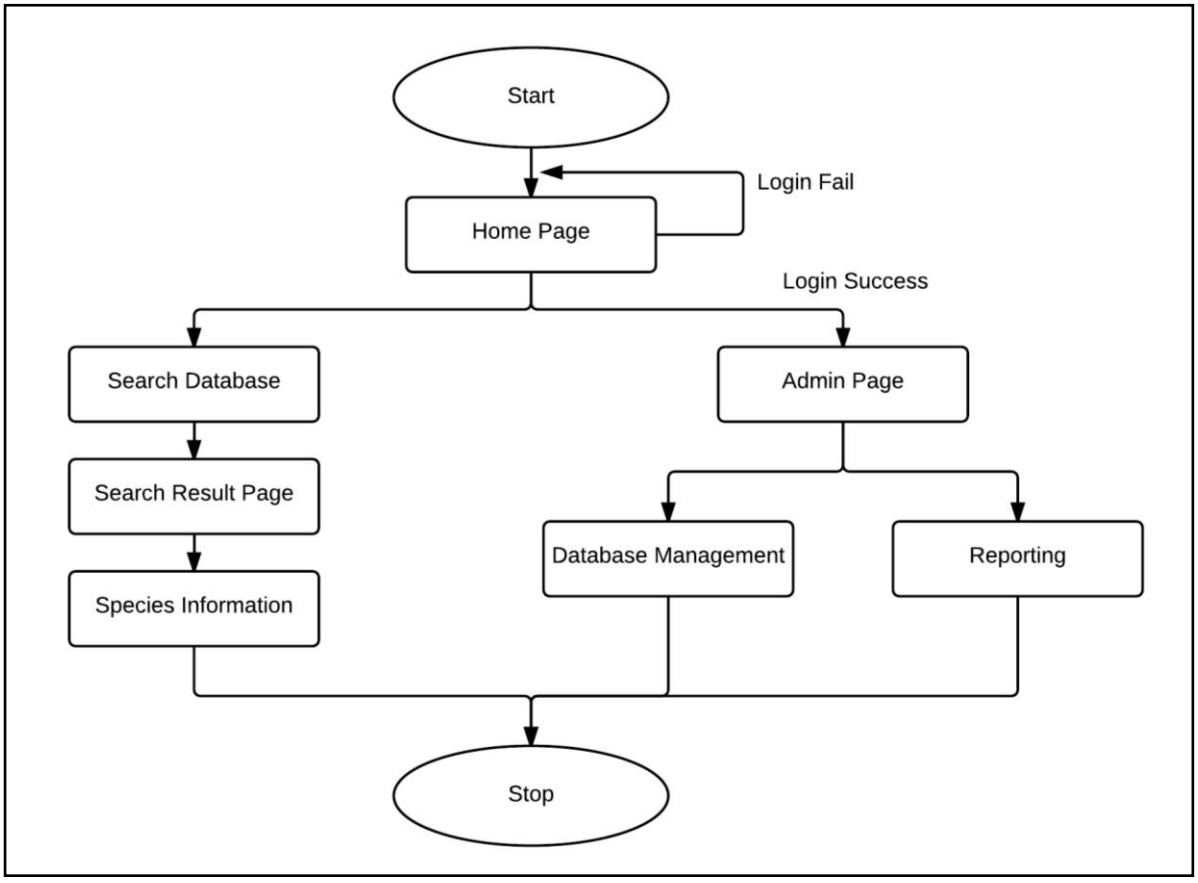


Figure 3.4: User interface design of database system.

3.2 Development Tools

Some software and hardware tools were used to develop the database system in this study. Hardware used has specification as many normal personal computer nowadays, while software used is open-source software which is free and can be downloaded online. Table 3.1 shows the details on the hardware and software tools used in developing the system.

Table 3.1: Development tools used in developing the database system.

Development Tools	Details
Hardware	• Intel® Core™ 2 Duo CPU processor • 1.00 GB free hard drive space • 2.00 GB RAM • 32-bit operating system
Software	• Genuine Windows® 7 Home Premium • MySQL Server 5.6.21 • PHPMyAdmin 4.2.7.1 • Notepad++ • PHP 5.5.19 • Apache Version 2.4.3 • Google Chrome

3.3 Data Preparation

Data preparation process organizes and arranges raw data into a suitable format to be entered into the system. The data used in building the databases are collected by other researchers and was directly delivered to us in electronic form. Thus, data collection and data digitization was not performed in this study. The data files come in different format and were converted into Microsoft Excel files, before performing data cleaning and validation.

Data cleaning was carried out to prevent data error and data inaccuracy. Each record was examined, and any error encountered was revised. The common errors encountered during data cleaning are:

(i) Spelling errors on scientific names, descriptions and locality;

(ii) Incorrect taxonomic classification, which have been validated by The Integrated Taxonomy Information System (ITIS);

(iii) Loss of data, incomplete data record;

(iv) Duplicated records.

As the data entry was done by human, errors occurred caused by human mistakes are inevitable. Those spelling errors and incorrect taxonomic classification, are amended based on the information from ITIS. Meanwhile, incomplete data records where there is no other information except binomial name, are omitted. Duplicated records are omitted too.

3.4 Database Development

Once the data is prepared, it is transferred and stored in the database. The database system is developed using MySQL database management system. PHP development environment is created by installing XAMPP in the workstation, and MySQL Server is used to create the database schema. Table 3.2 to Table 3.6 describe data definition for each table in the ginger database.

Table 3.2: Data field definition of table ginger.

Field Name	Field Definition	Primary/Foreign Key	Data Type
gingerID	A unique number that has been assigned to every ginger species	Primary Key	INT(5)
speciesName	A unique name given to a particular species formed by the combination of genus and species name		VARCHAR(50)
synonym	Other names that have been used for the same species		VARCHAR(50)
genusID	A unique number that has been assigned to every genus	Foreign Key	INT(5)
catalougeNumber	Catalogue number that has been given during collection		VARCHAR(10)
collectionNumber	Collection number that has been given during collection		VARCHAR(10)
locality	Specific geographic area where the specimen has been collected		VARCHAR(100)
locationID	A unique number that has been assigned to every location	Foreign Key	INT(5)
day	Date when the specimen collected, formed by day, month and year		INT(2)
month	Date when the specimen collected, formed by day, month and year		INT(2)
year	Date when the specimen collected, formed by day, month and year		INT(4)
collectorID	A unique number that has been assigned to every collector	Foreign Key	INT(5)

Table 3.3: Data field definition of table genus.

Field Name	Field Definition	Primary/Foreign Key	Data Type
genusID	A unique number that has been assigned to every genus	Primary Key	INT(5)
taxoID	A unique number that has been assigned to set of taxonomy classification from kingdom to family	Foreign Key	INT(5)
genusName	A unique name given to a particular genus		VARCHAR(25)
description	A unique number that has been assigned to every genus		TEXT

Table 3.4: Data field definition of table taxonomy.

Field Name	Field Definition	Primary/Foreign Key	Data Type
taxoID	A unique number that has been assigned to set of taxonomy classification from kingdom to family	Primary Key	INT(5)
kingdom	A unique name given to a particular kingdom (Plantae)		VARCHAR(25)
phylum	A unique name given to a particular phylum (Magnoliophyta)		VARCHAR(25)
class	A unique name given to a particular class (Liliopsida)		VARCHAR(25)
order	A unique name given to a particular order (Zingiberales)		VARCHAR(25)
family	A unique name given to a particular family (Zingiberaceae)		VARCHAR(25)

Table 3.5: Data field definition of table collector.

Field Name	Field Definition	Primary/Foreign Key	Data Type
collectorID	A unique number that has been assigned to every collector	Primary Key	INT(5)
collectorName	Name of collectors		VARCHAR(50)
institute	Name of institute where collector is working during collection		VARCHAR(100)

Table 3.6: Data field definition of table location.

Field Name	Field Definition	Primary/Foreign Key	Data Type
locationID	A unique number that has been assigned to every location	Primary Key	INT(5)
locationName	A unique name of each location		VARCHAR(20)

CHAPTER 4

RESULTS

4.1 Home Page

Three databases were developed in this study. However, as same framework is used in developing all three databases, only ginger database is showed in this chapter. Figure 4.1 shows the home page of the web portal. The banner at the top is login banner and it would appear on every web pages. Left side of web page embeds a menu which consists of links. The information search links are provided at this menu.

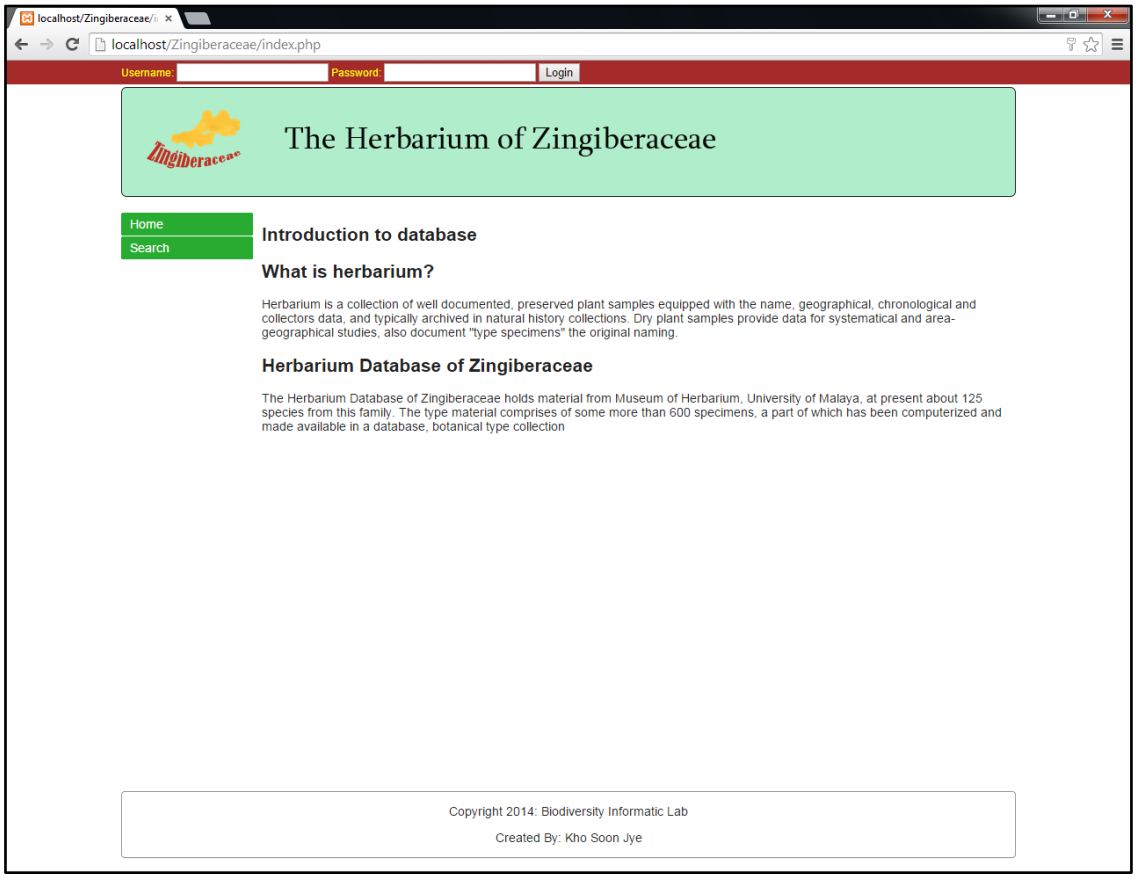


Figure 4.1: Home page of ginger database.

4.2 Information Search

Users will be directed to information search page via the links provided at the links menu. In this page, there is a text field that allows users to enter text query. If a user wants to have a view on all species record in database, the user may click ‘Search’ button by leaving the text field blank. This will retrieve every species record in database.

If a user wants to search for particular species record, the user may enter text query in the text field. When entering query, autocomplete suggestion feature will list out suggestions for users to choose and complete the text query. Users may or may not choose any suggestion before click the ‘Search’ button. Once Search button is clicked, search result will appear as shown in Figure 4.4. Click on any species record in the result table will direct users to the detail page of that species.

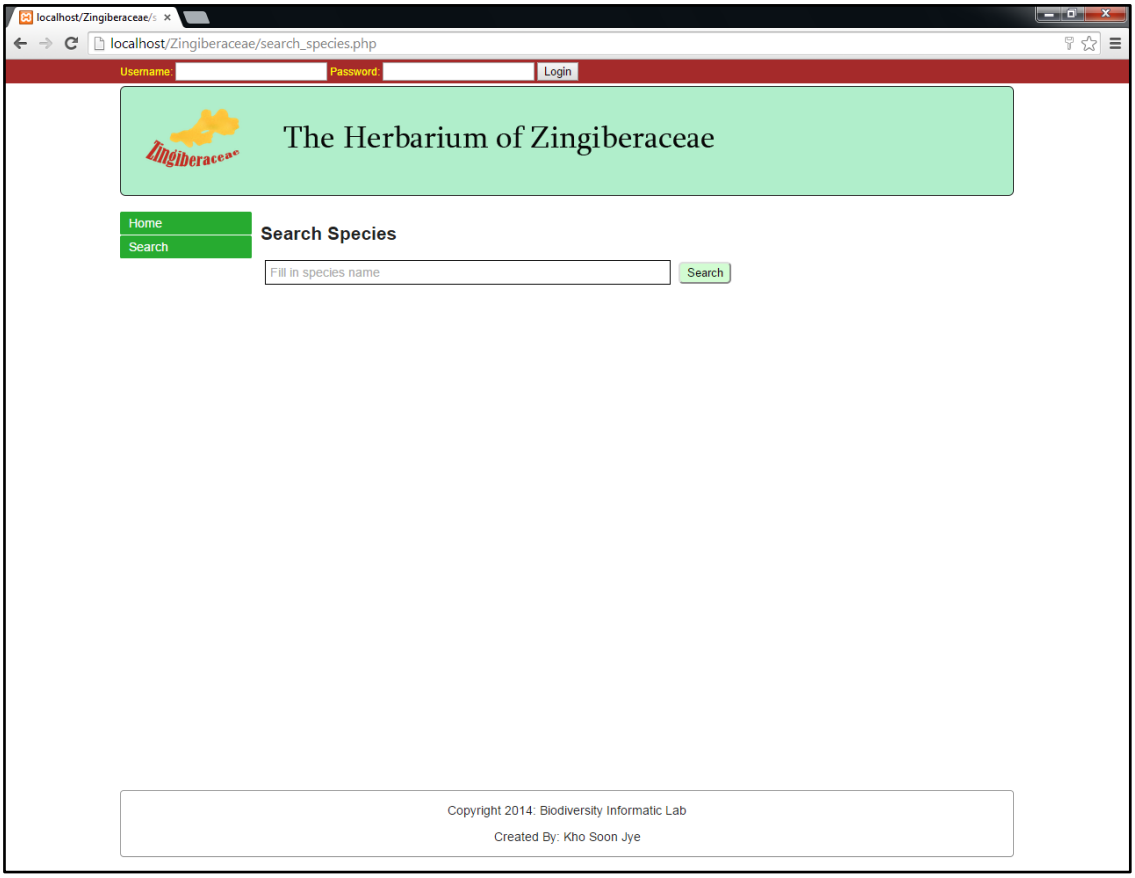


Figure 4.2: Information search page.

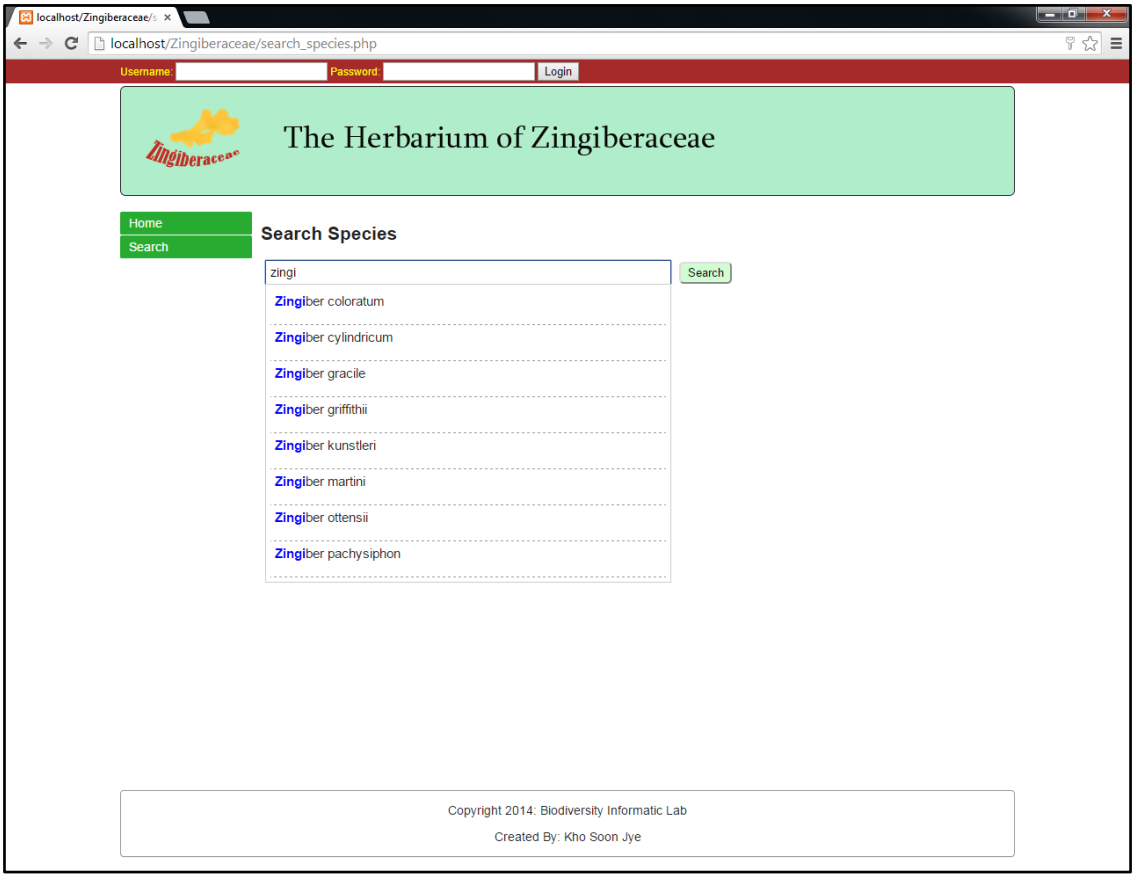


Figure 4.3: Suggestion lists provided by autocomplete suggestion feature, contain suggestion for users to complete the text query.

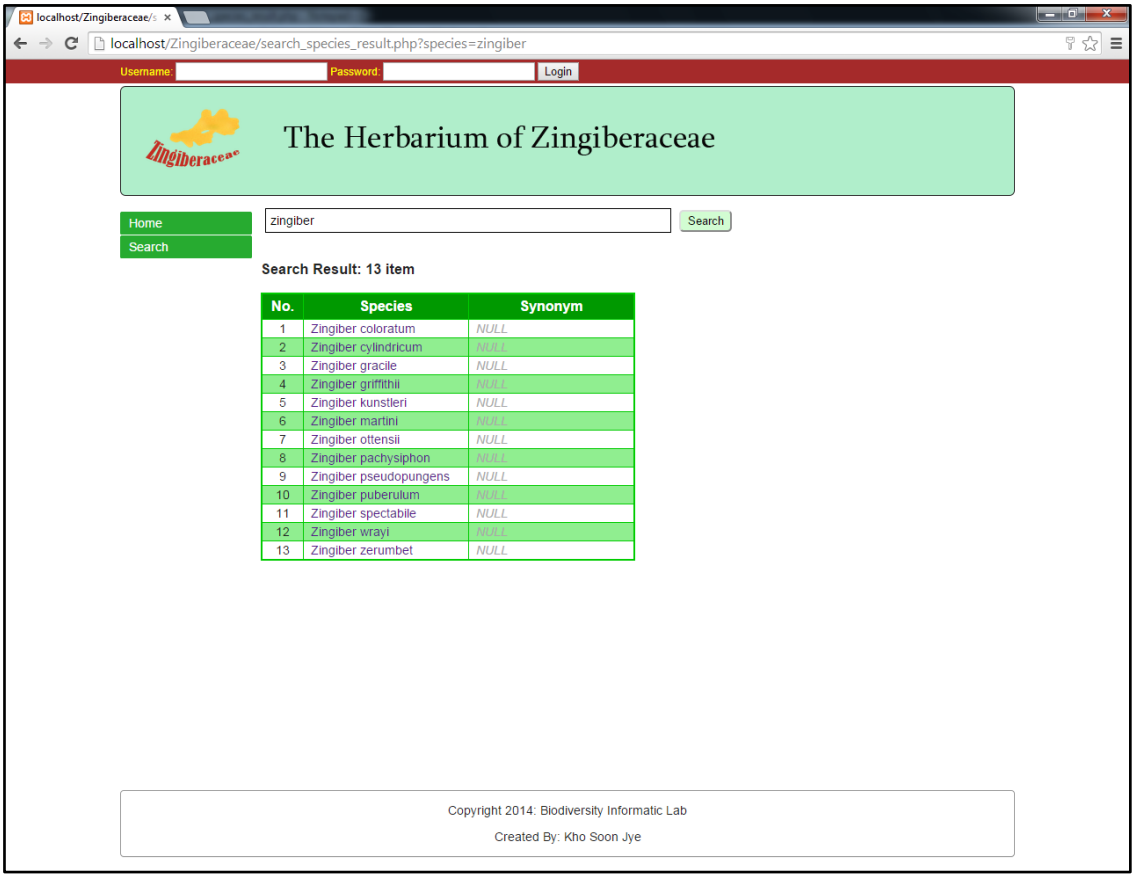


Figure 4.4: Search results appeared after users click on "Search" button. Text query used in this example is "zingiber". Click on any species record in the result table will direct users to the detail page.

4.3 Detail Page

Detail page contains all the information of a particular species. There are numerous tables in this page. The first three table display binomial name and synonym if any, taxonomic information, and collector's information. The fourth table contains information about country and specific locality of that species. In this table, there is an interactive map showing the locality, and users can interact with the map by scrolling or dragging. The last table is the largest table, it displays links and images retrieved from web which are relevant

to the particular species. Users can click on those links and images if he or she wants to know more about that species.

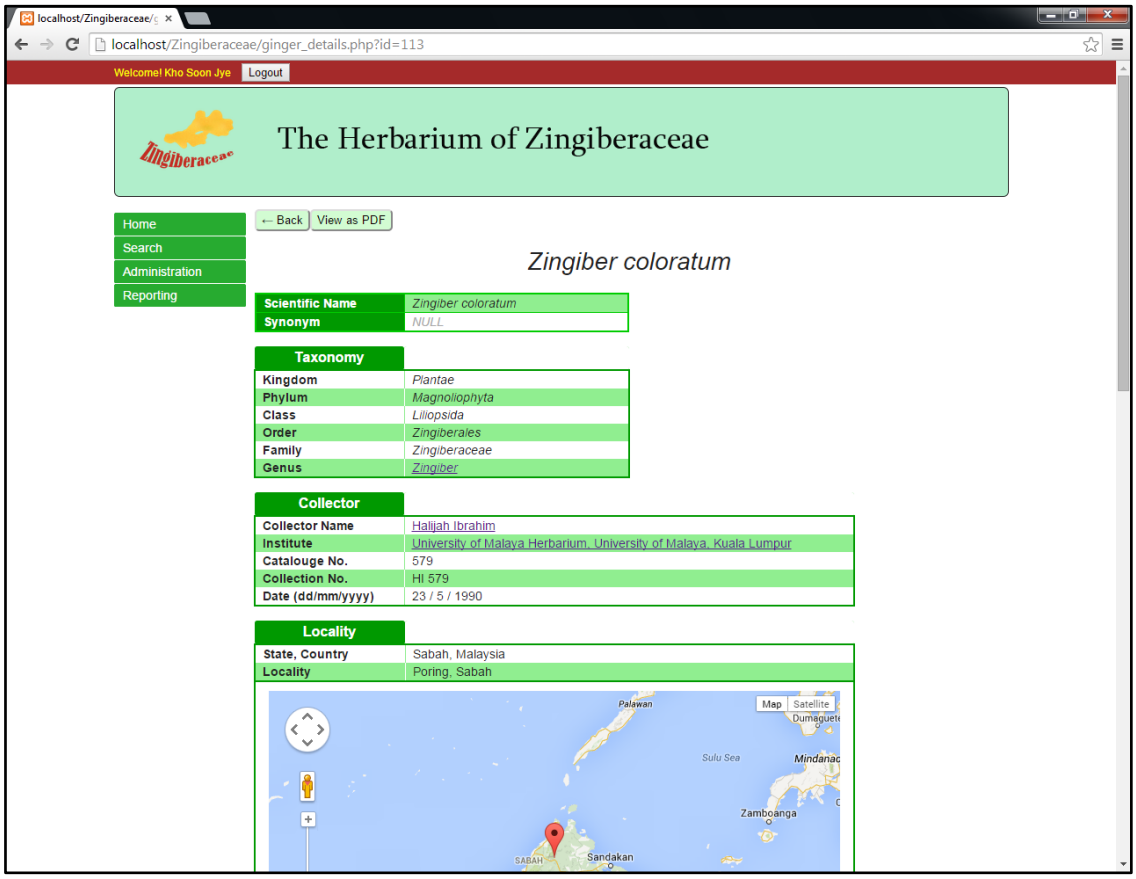


Figure 4.5: Detail page of species *Zingiber coloratum* showing the first four tables. These four tables display binomial name and synonym if any, taxonomic information, collector's information and locality.

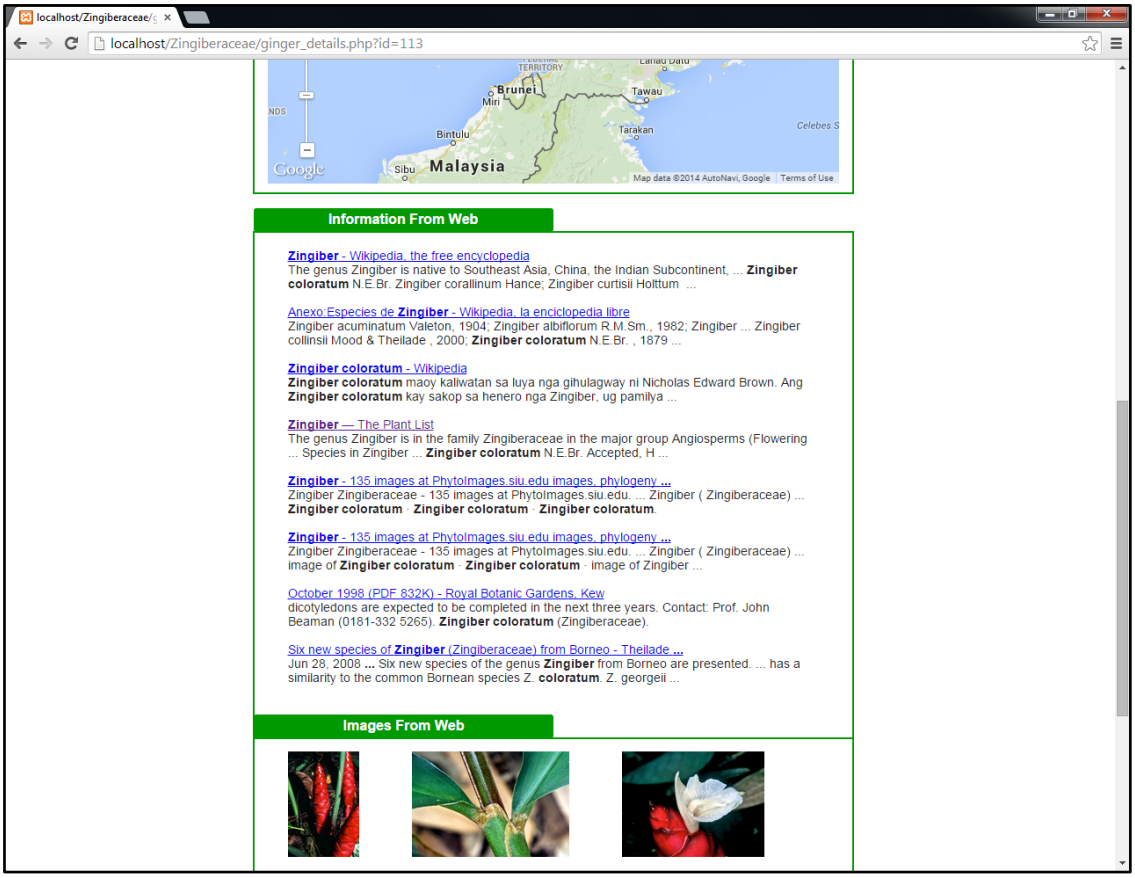


Figure 4.6: Detail page of species *Zingiber coloratum* showing the last table. Last table contains links and images retrieved from web.

4.4 Admin Login

Each administrator is given a combination of username and password for login. Login enables administrator to use function in administrating database and reporting. At the top of every web page, there is a brown banner with two text fields. Administrator is required to enter username in the first text field and password in the second to login. Once logged in, administrator will be directed to the admin home page as shown in Figure 4.8, where variety of administration function can be found here. Besides admin home page, those database management function can be accessed through the new links appeared at links menu.

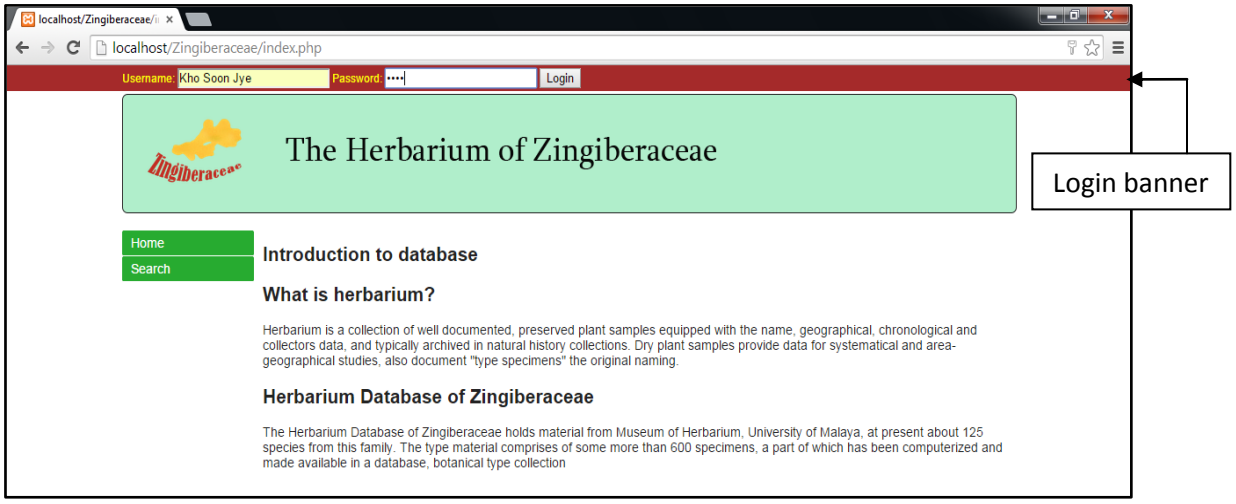


Figure 4.7: Login banner at the top of every web page.

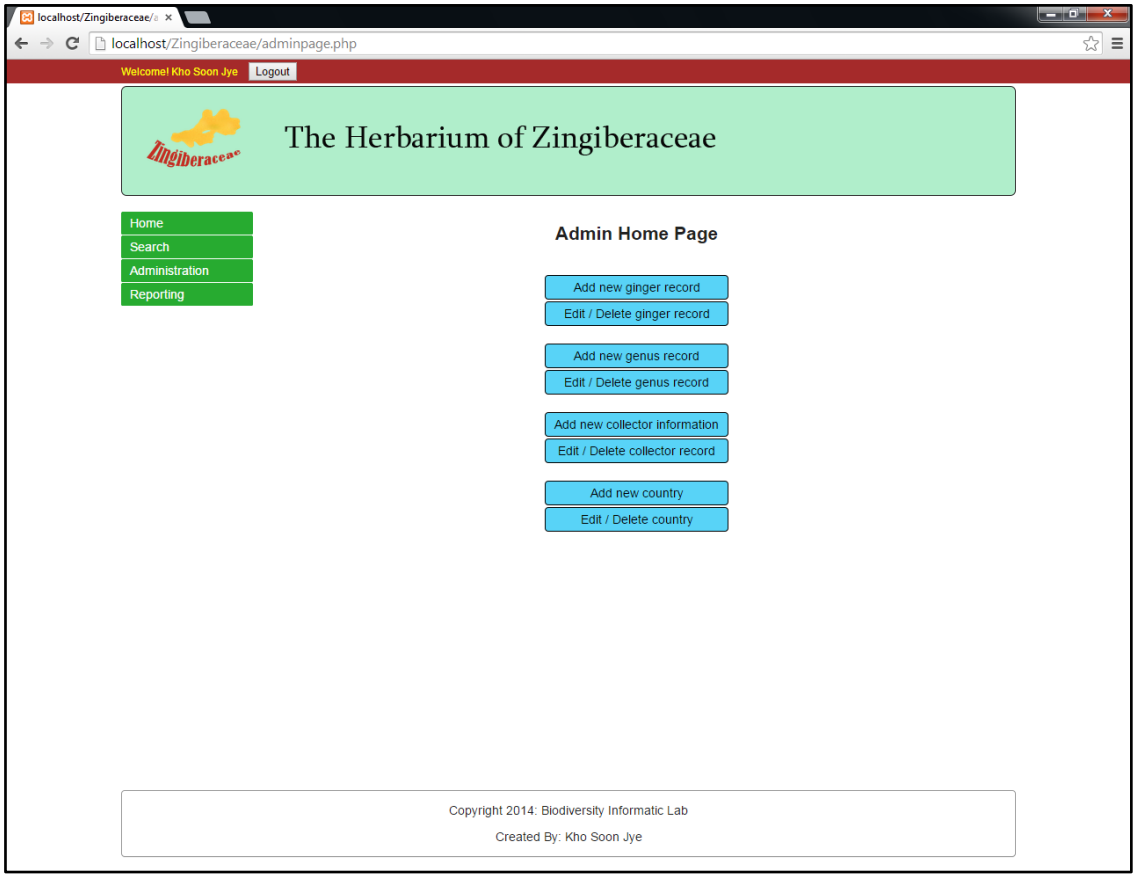


Figure 4.8: Admin home page where variety of database management function can be found here. Those database management functions can be accessed too via the new links appeared at links menu.

4.5 Adding Record

Adding record is one of database management function and can only be accessed by administrator after logging in. This function allows administrator to add new species record to the database. Administrator can access to this page from admin home page or links menu. In adding record web page, administrators have to fill in the information of new species. Once "Add Ginger" button is clicked, data validation will take place to ensure no invalid data has filled in or leaving blank on important data field. After validation, information of new ginger record will be added into database.

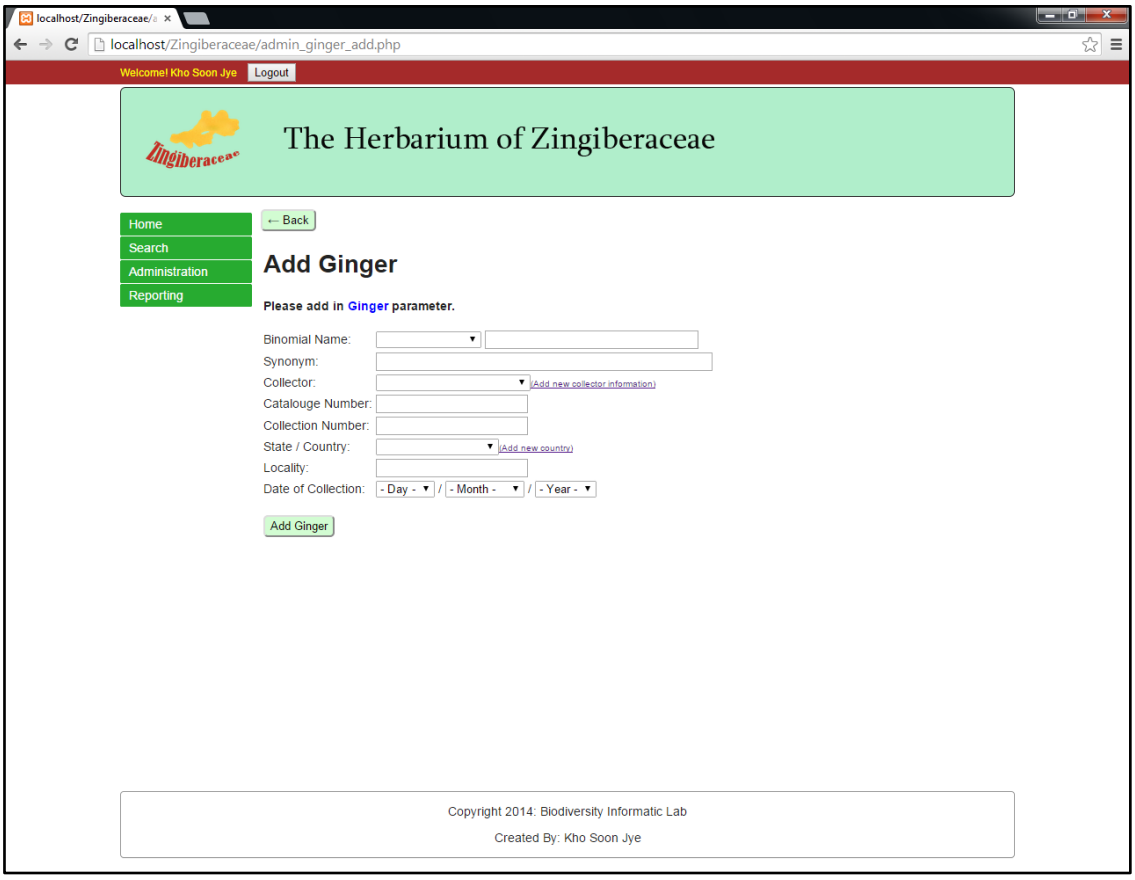


Figure 4.9: Adding record web page allows users to fill in information of new species record.

4.6 Editing Record

Editing record is one of database management function and can only be accessed by administrator. This function allows administrator to edit and update existing record in the database. Administrator can access this page from admin home page or links menu. To edit a record, administrator has to go to the detail page of the particular species record he or she wants to edit. In the detail web page, there are four buttons on the top of detail section. Click on the “Edit” button (second button from the left) and a form will prompt out. Administrator will need to fill in updated information in respective field. Once “Confirm” button is clicked, data validation will take place and the information of that particular species record in the database will be updated.

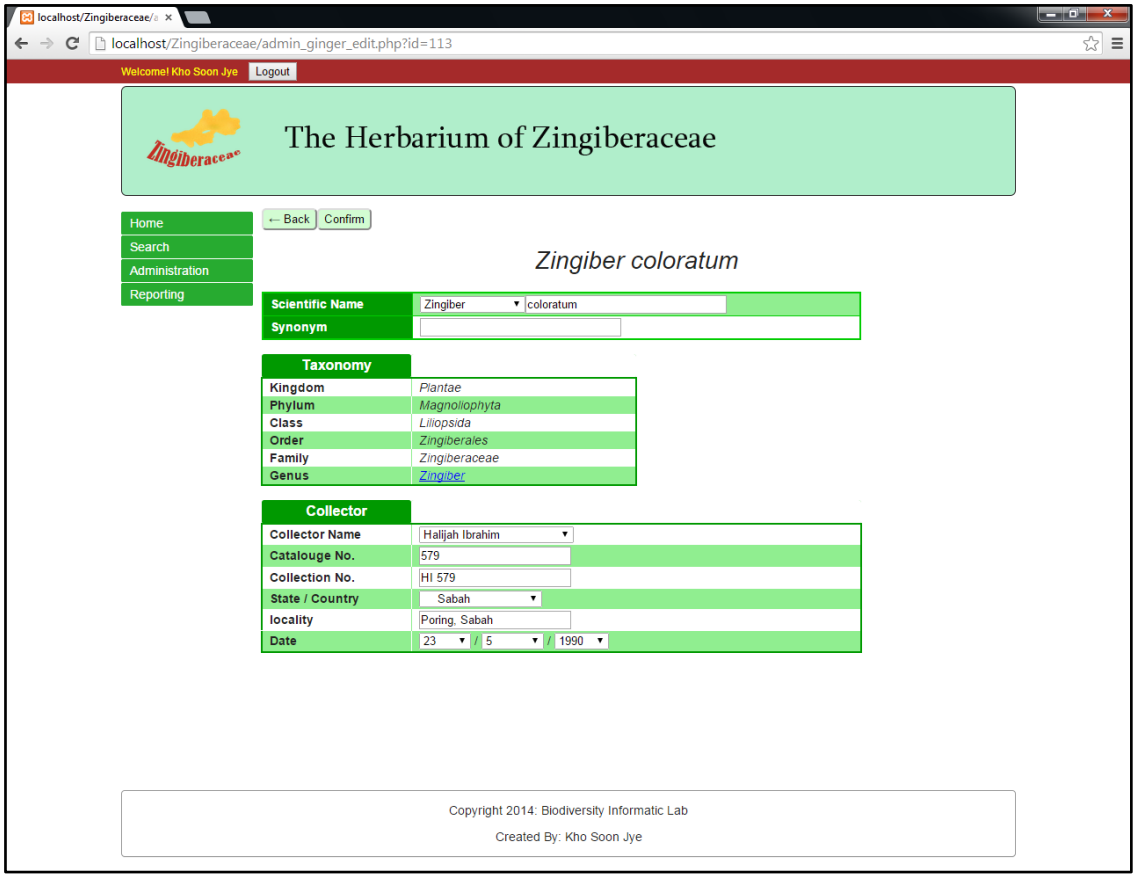


Figure 4.10: Editing record web page showing the information of existing record in database. Administrator are allowed to update relevant information of the record.

4.7 Deleting Record

Deleting record is one of database management function and can only be accessed by administrator after logging in. This function allows administrator to delete existing record in the database. Administrator can access this page from admin home page or links menu. To delete a record, administrator has to go to the detail page of the particular species record he or she wants to delete. In the detail web page, there are four buttons on the top of detail section. Click on the “Delete” button (third button from the left) and that particular species record will be removed from the database.

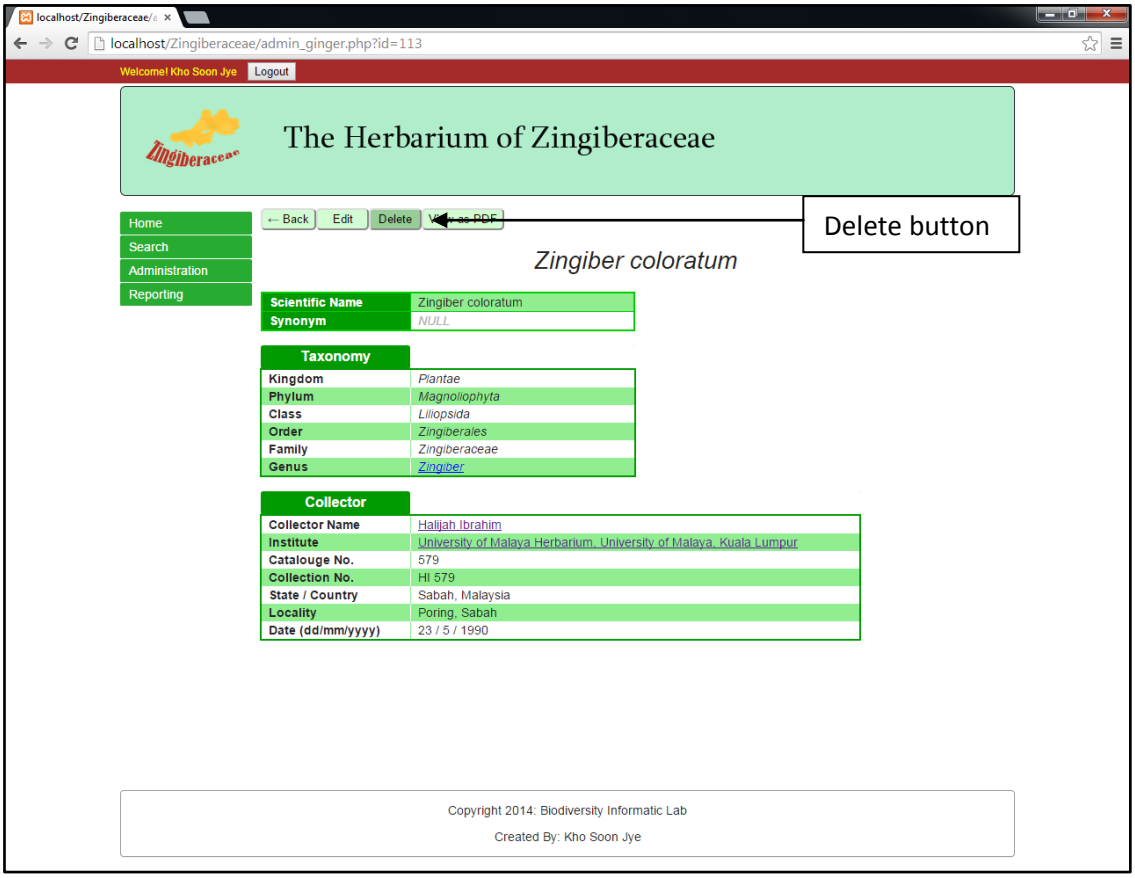


Figure 4.11: Detail page of *Zingiber coloratum*. Click on "Delete" button as pointed and the record will be removed from the database.

4.8 Reporting (Single Species)

This function allows users to produce one-page PDF file containing information of a particular species record. To produce the PDF file, users have to go to the detail page of the particular species record he or she wants to export. In the detail web page, there are four buttons on the top of detail section. Click on the "View as PDF" button (fourth button from the left), a new tab will open and a PDF file will appear, as shown in Figure 4.13.

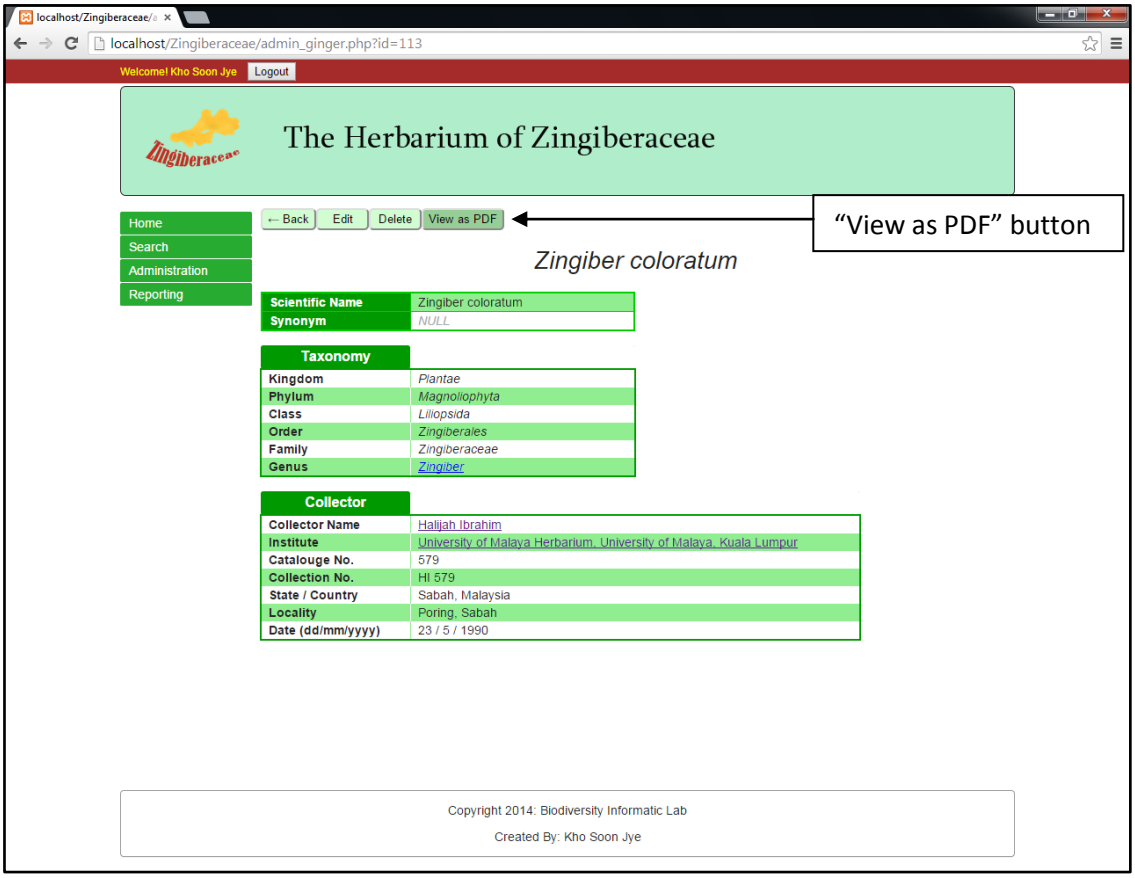


Figure 4.12: Detail page of *Zingiber coloratum*. Click on "View as PDF" button as pointed and one-page PDF file will be produced in a new tab.

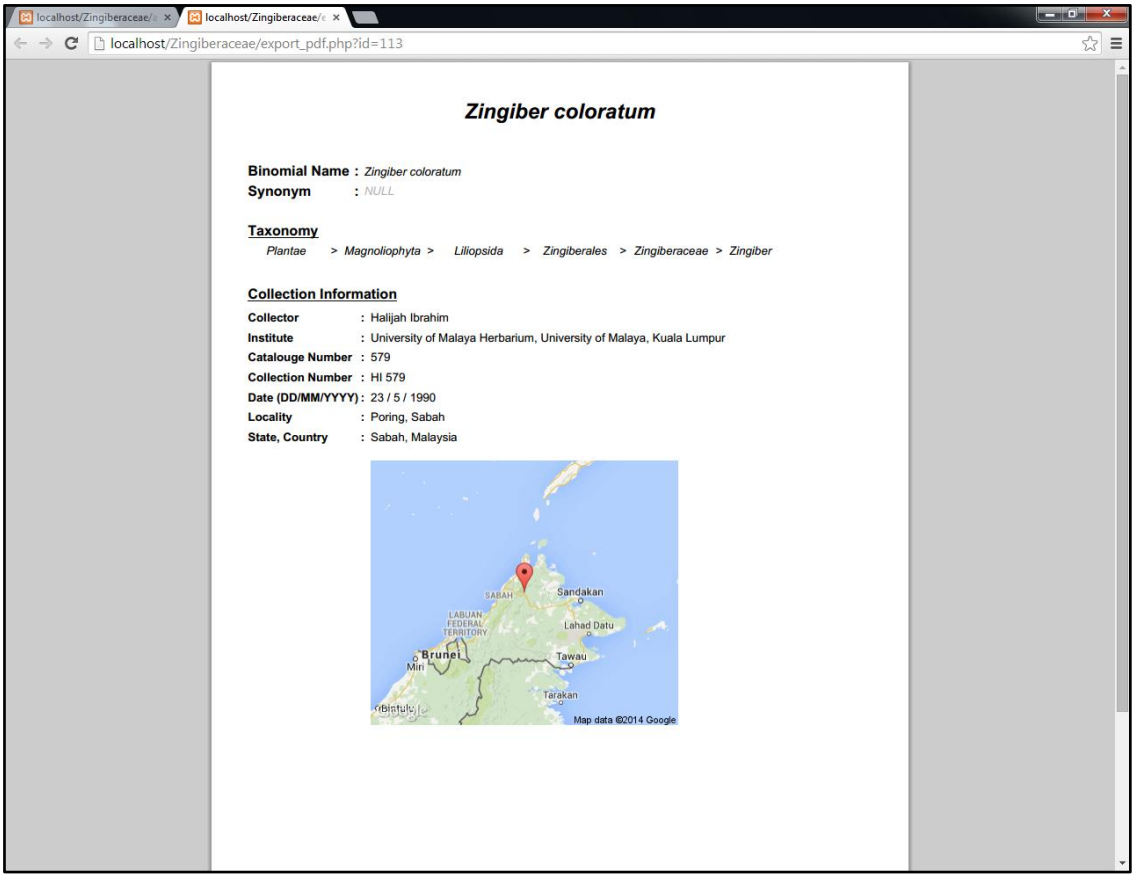


Figure 4.13: One-page PDF file containing information of species *Zingiber coloratum*.

4.9 Reporting (Multiple Species)

This function can only be accessed by administrator. Administrator can access this function by clicking 'Reporting' link in the links menu. This function allows administrator to produce report files containing information of multiple species record, in Microsoft Excel format or PDF format. In reporting page, administrator can choose which data field to be exported. Besides, administrator can also filter species record he or she wanted to be exported. After selection, administrator can choose to produce report in Microsoft Excel format or PDF format. There are two buttons at the bottom of reporting section, 'Excel' and 'PDF' button. Click on any button and the report will be produced accordingly. Figure 4.15

shows the print screen of Excel-format report while Figure 4.16 shows the print screen of PDF-format report.

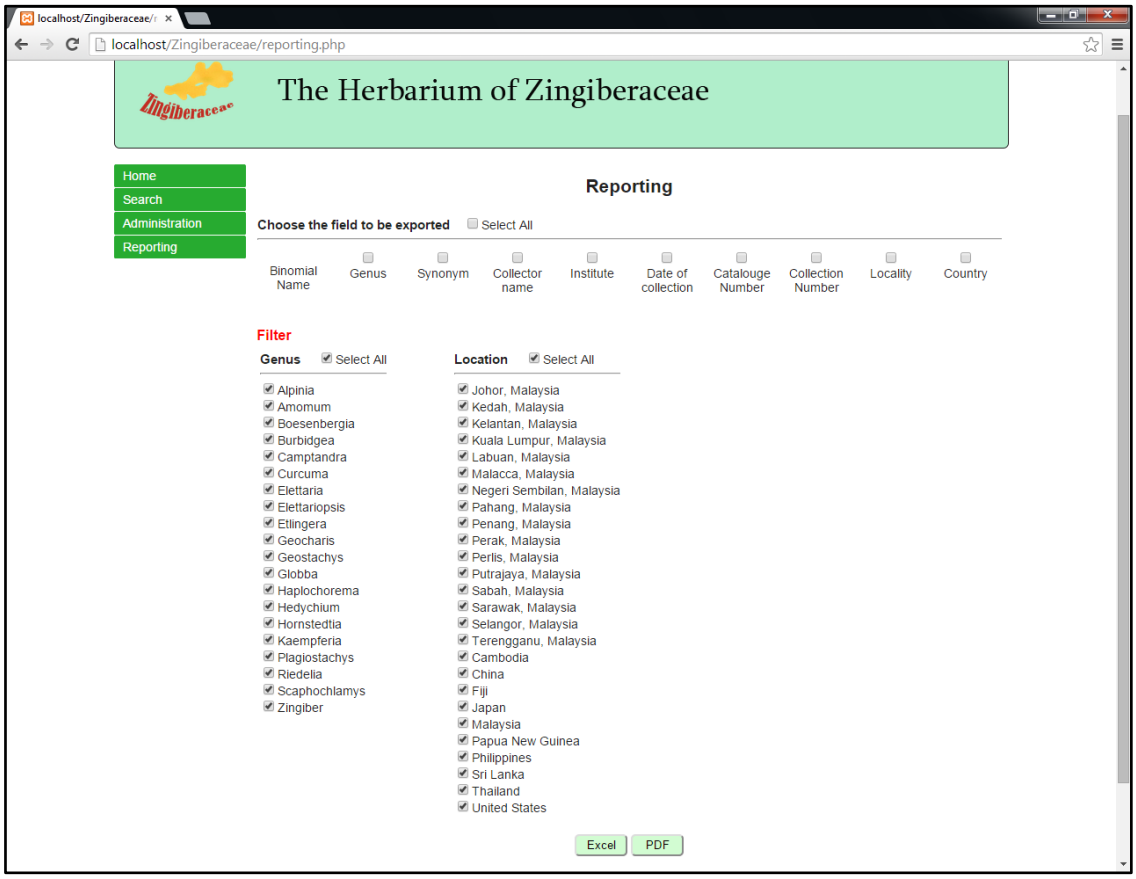
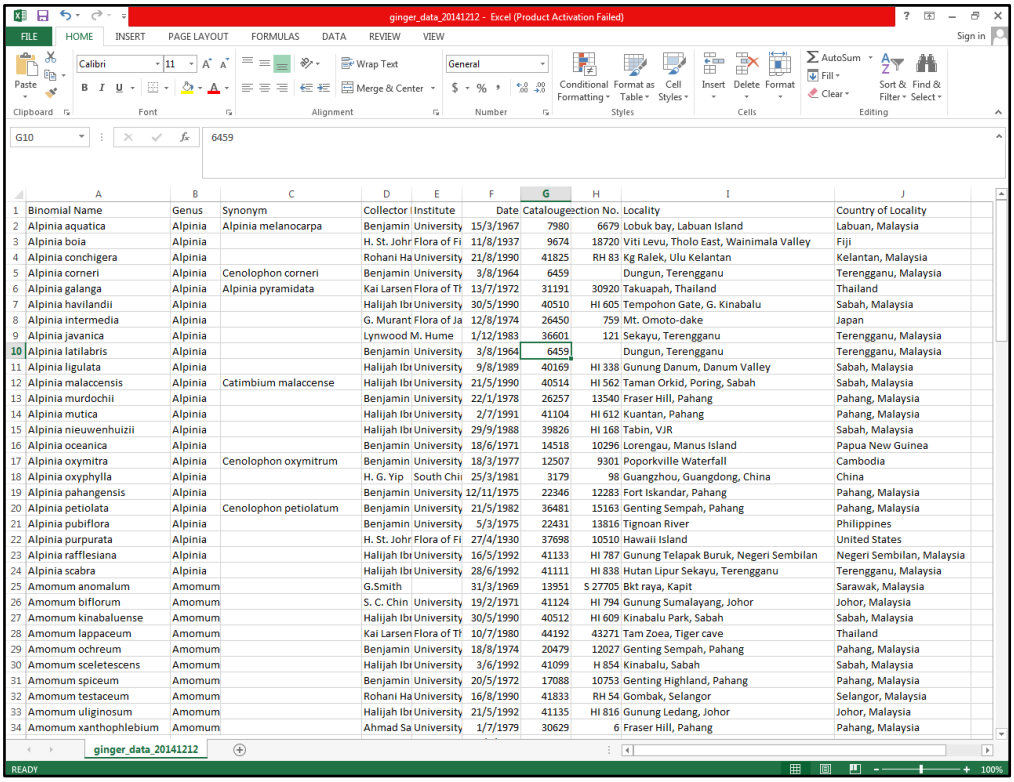
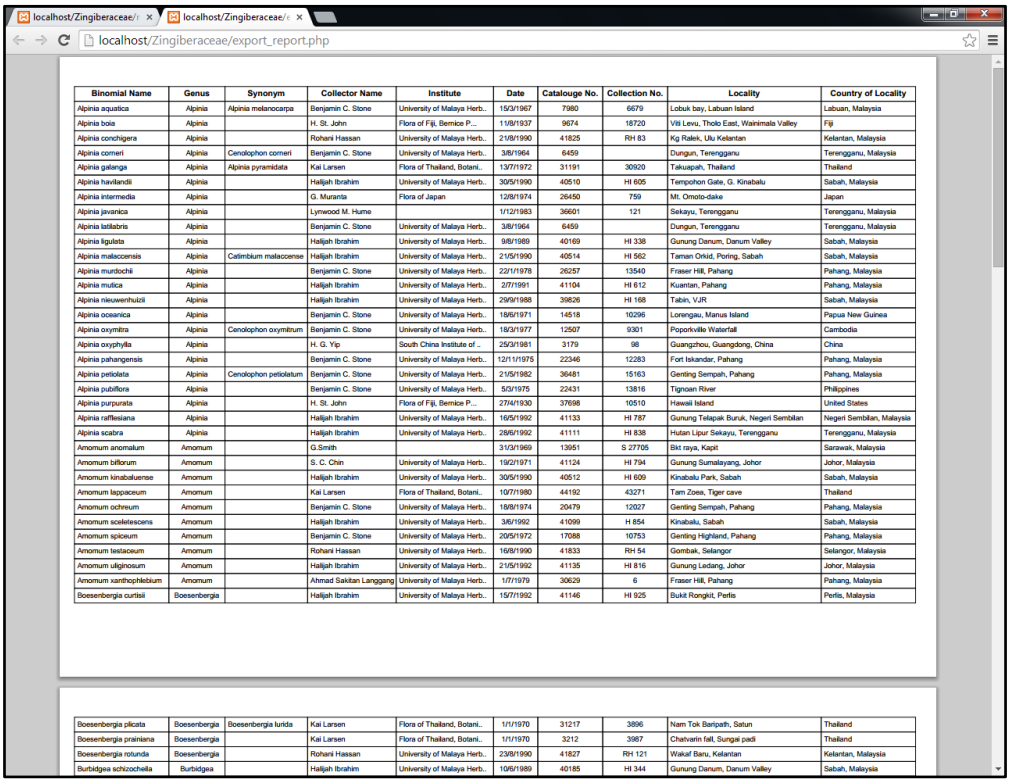


Figure 4.14: Reporting page of ginger database.



	A	B	C	D	E	F	G	H	I	J
1	Binomial Name	Genus	Synonym	Collector	Institute	Date	Catalougection No.		Locality	Country of Locality
2	Alpinia aquatica	Alpinia	Alpinia melanocarpa	Benjamin	University	15/3/1967	7980	6679	Lobuk bay, Labuan Island	Labuan, Malaysia
3	Alpinia boia	Alpinia		H. St. Johr	Flora of Fi	11/8/1937	9674	18720	Viti Levu, Tholo East, Wainimala Valley	Fiji
4	Alpinia conchigera	Alpinia		Rohani Ha	University	21/8/1990	41825	RH 83	Kg Ralek, Ulu Kelantan	Kelantan, Malaysia
5	Alpinia corneri	Alpinia	Cenolophon corneri	Benjamin	University	3/8/1964	6459		Dungun, Terengganu	Terengganu, Malaysia
6	Alpinia galanga	Alpinia	Alpinia pyramidata	Kai Larsen	Flora of Th	13/7/1972	31191	30920	Takuapah, Thailand	Thailand
7	Alpinia havilandii	Alpinia		Halijah Ibr	University	30/5/1990	40510	HI 605	Tempohon Gate, G. Kinabalu	Sabah, Malaysia
8	Alpinia intermedia	Alpinia		G. Murant	Flora of Ja	12/8/1974	26450	759	Mt. Omoto-dake	Japan
9	Alpinia javanica	Alpinia		Lynwood M. Hume		1/12/1983	36601	121	Sekayu, Terengganu	Terengganu, Malaysia
10	Alpinia latilabris	Alpinia		Benjamin	University	3/8/1964	6459		Dungun, Terengganu	Terengganu, Malaysia
11	Alpinia ligulata	Alpinia		Halijah Ibr	University	9/8/1989	40169	HI 338	Gunung Danum, Danum Valley	Sabah, Malaysia
12	Alpinia malaccensis	Alpinia	Catimbium malaccense	Halijah Ibr	University	21/5/1990	40514	HI 562	Taman Orkid, Poring, Sabah	Sabah, Malaysia
13	Alpinia murdochii	Alpinia		Benjamin	University	22/1/1978	26257	13540	Fraser Hill, Pahang	Pahang, Malaysia
14	Alpinia mutica	Alpinia		Halijah Ibr	University	2/7/1991	41104	HI 612	Kuantan, Pahang	Pahang, Malaysia
15	Alpinia nieuwenhuizii	Alpinia		Halijah Ibr	University	29/9/1988	39826	HI 168	Tabin, VJR	Sabah, Malaysia
16	Alpinia oceanica	Alpinia		Benjamin	University	18/6/1971	14518	10296	Lorengau, Manus Island	Papua New Guinea
17	Alpinia oxymitra	Alpinia	Cenolophon oxymitrum	Benjamin	University	18/3/1977	12507	9301	Poporkville Waterfall	Cambodia
18	Alpinia oxyphylla	Alpinia		H. G. Yip	South Chi	25/3/1981	3179	98	Guangzhou, Guangdong, China	China
19	Alpinia pahangensis	Alpinia		Benjamin	University	12/11/1975	22346	12283	Fort Iskandar, Pahang	Pahang, Malaysia
20	Alpinia petiolata	Alpinia	Cenolophon petiolatum	Benjamin	University	21/5/1982	36481	15163	Genting Sempah, Pahang	Pahang, Malaysia
21	Alpinia pubiflora	Alpinia		Benjamin	University	5/3/1975	22431	13816	Tignoan River	Philippines
22	Alpinia purpurata	Alpinia		H. St. Johr	Flora of Fi	27/4/1930	37698	10510	Hawaii Island	United States
23	Alpinia rafflesiana	Alpinia		Halijah Ibr	University	16/5/1992	41133	HI 787	Gunung Telapak Buruk, Negeri Sembilan	Negeri Sembilan, Malaysia
24	Alpinia scabra	Alpinia		Halijah Ibr	University	28/6/1992	41111	HI 838	Hutan Lipur Sekayu, Terengganu	Terengganu, Malaysia
25	Amomum anomalum	Amomum		G.Smith		31/3/1969	13951	S 27705	Bkt raya, Kapit	Sarawak, Malaysia
26	Amomum biflorum	Amomum		S. C. Chin	University	19/2/1971	41124	HI 794	Gunung Sumalayang, Johor	Johor, Malaysia
27	Amomum kinabaluense	Amomum		Halijah Ibr	University	30/5/1990	40512	HI 609	Kinabalu Park, Sabah	Sabah, Malaysia
28	Amomum lappaceum	Amomum		Kai Larsen	Flora of Th	10/7/1980	44192	43271	Tam Zoea, Tiger cave	Thailand
29	Amomum ochreum	Amomum		Benjamin	University	18/8/1974	20479	12027	Genting Sempah, Pahang	Pahang, Malaysia
30	Amomum sceletescens	Amomum		Halijah Ibr	University	3/6/1992	41099	H 854	Kinabalu, Sabah	Sabah, Malaysia
31	Amomum spiceum	Amomum		Benjamin	University	20/5/1972	17088	10753	Genting Highland, Pahang	Pahang, Malaysia
32	Amomum testaceum	Amomum		Rohani Ha	University	16/8/1990	41833	RH 54	Gombak, Selangor	Selangor, Malaysia
33	Amomum uliginosum	Amomum		Halijah Ibr	University	21/5/1992	41135	HI 816	Gunung Ledang, Johor	Johor, Malaysia
34	Amomum xanthophlebium	Amomum		Ahmad Sa	University	1/7/1979	30629	6	Fraser Hill, Pahang	Pahang, Malaysia

Figure 4.15: Excel-format report produced by reporting function.



Binomial Name	Genus	Synonym	Collector Name	Institute	Date	Catalouge No.	Collection No.	Locality	Country of Locality
Alpinia aquatica	Alpinia	Alpinia melanocarpa	Benjamin C. Stone	University of Malaya Herb..	15/3/1967	7980	6679	Lobuk bay, Labuan Island	Labuan, Malaysia
Alpinia boia	Alpinia		H. St. John	Flora of Fiji, Bernice P...	11/8/1937	9674	18720	Viti Levu, Tholo East, Wainimala Valley	Fiji
Alpinia conchigera	Alpinia		Rohani Hassan	University of Malaya Herb..	21/8/1990	41825	RH 83	Kg Ralek, Ulu Kelantan	Kelantan, Malaysia
Alpinia corneri	Alpinia	Cenolophon corneri	Benjamin C. Stone	University of Malaya Herb..	3/8/1964	6459		Dungun, Terengganu	Terengganu, Malaysia
Alpinia galanga	Alpinia	Alpinia pyramidata	Kai Larsen	Flora of Thailand, Botani..	13/7/1972	31191	30920	Takuapah, Thailand	Thailand
Alpinia havilandii	Alpinia		Halijah Ibrahim	University of Malaya Herb..	30/5/1990	40510	HI 605	Tempohon Gate, G. Kinabalu	Sabah, Malaysia
Alpinia intermedia	Alpinia		G. Muranta	Flora of Japan	12/8/1974	26450	759	Mt. Omoto-dake	Japan
Alpinia javanica	Alpinia		Lynwood M. Hume		1/12/1983	36601	121	Sekayu, Terengganu	Terengganu, Malaysia
Alpinia latilabris	Alpinia		Benjamin C. Stone	University of Malaya Herb..	3/8/1964	6459		Dungun, Terengganu	Terengganu, Malaysia
Alpinia ligulata	Alpinia		Halijah Ibrahim	University of Malaya Herb..	9/8/1989	40169	HI 338	Gunung Danum, Danum Valley	Sabah, Malaysia
Alpinia malaccensis	Alpinia	Catimbium malaccense	Halijah Ibrahim	University of Malaya Herb..	21/5/1990	40514	HI 562	Taman Orkid, Poring, Sabah	Sabah, Malaysia
Alpinia murdochii	Alpinia		Benjamin C. Stone	University of Malaya Herb..	22/1/1978	26257	13540	Fraser Hill, Pahang	Pahang, Malaysia
Alpinia mutica	Alpinia		Halijah Ibrahim	University of Malaya Herb..	2/7/1991	41104	HI 612	Kuantan, Pahang	Pahang, Malaysia
Alpinia nieuwenhuizii	Alpinia		Halijah Ibrahim	University of Malaya Herb..	29/9/1988	39826	HI 168	Tabin, VJR	Sabah, Malaysia
Alpinia oceanica	Alpinia		Benjamin C. Stone	University of Malaya Herb..	18/6/1971	14518	10296	Lorengau, Manus Island	Papua New Guinea
Alpinia oxymitra	Alpinia	Cenolophon oxymitrum	Benjamin C. Stone	University of Malaya Herb..	18/3/1977	12507	9301	Poporkville Waterfall	Cambodia
Alpinia oxyphylla	Alpinia		H. G. Yip	South China Institute of ..	25/3/1981	3179	98	Guangzhou, Guangdong, China	China
Alpinia pahangensis	Alpinia		Benjamin C. Stone	University of Malaya Herb..	12/11/1975	22346	12283	Fort Iskandar, Pahang	Pahang, Malaysia
Alpinia petiolata	Alpinia	Cenolophon petiolatum	Benjamin C. Stone	University of Malaya Herb..	21/5/1982	36481	15163	Genting Sempah, Pahang	Pahang, Malaysia
Alpinia pubiflora	Alpinia		Benjamin C. Stone	University of Malaya Herb..	5/3/1975	22431	13816	Tignoan River	Philippines
Alpinia purpurata	Alpinia		H. St. John	Flora of Fiji, Bernice P...	27/4/1930	37698	10510	Hawaii Island	United States
Alpinia rafflesiana	Alpinia		Halijah Ibrahim	University of Malaya Herb..	16/5/1992	41133	HI 787	Gunung Telapak Buruk, Negeri Sembilan	Negeri Sembilan, Malaysia
Alpinia scabra	Alpinia		Halijah Ibrahim	University of Malaya Herb..	28/6/1992	41111	HI 838	Hutan Lipur Sekayu, Terengganu	Terengganu, Malaysia
Amomum anomalum	Amomum		G.Smith		31/3/1969	13951	S 27705	Bkt raya, Kapit	Sarawak, Malaysia
Amomum biflorum	Amomum		S. C. Chin	University of Malaya Herb..	19/2/1971	41124	HI 794	Gunung Sumalayang, Johor	Johor, Malaysia
Amomum kinabaluense	Amomum		Halijah Ibrahim	University of Malaya Herb..	30/5/1990	40512	HI 609	Kinabalu Park, Sabah	Sabah, Malaysia
Amomum lappaceum	Amomum		Kai Larsen	Flora of Thailand, Botani..	10/7/1980	44192	43271	Tam Zoea, Tiger cave	Thailand
Amomum ochreum	Amomum		Benjamin C. Stone	University of Malaya Herb..	18/8/1974	20479	12027	Genting Sempah, Pahang	Pahang, Malaysia
Amomum sceletescens	Amomum		Halijah Ibrahim	University of Malaya Herb..	3/6/1992	41099	H 854	Kinabalu, Sabah	Sabah, Malaysia
Amomum spiceum	Amomum		Benjamin C. Stone	University of Malaya Herb..	20/5/1972	17088	10753	Genting Highland, Pahang	Pahang, Malaysia
Amomum testaceum	Amomum		Rohani Hassan	University of Malaya Herb..	16/8/1990	41833	RH 54	Gombak, Selangor	Selangor, Malaysia
Amomum uliginosum	Amomum		Halijah Ibrahim	University of Malaya Herb..	21/5/1992	41135	HI 816	Gunung Ledang, Johor	Johor, Malaysia
Amomum xanthophlebium	Amomum		Ahmad Sakitan Langgang	University of Malaya Herb..	1/7/1979	30629	6	Fraser Hill, Pahang	Pahang, Malaysia
Boesenbergia curtisii	Boesenbergia		Halijah Ibrahim	University of Malaya Herb..	15/7/1992	41146	HI 925	Bukit Rongkit, Perlis	Perlis, Malaysia

Boesenbergia plicata	Boesenbergia	Boesenbergia lurida	Kai Larsen	Flora of Thailand, Botani..	1/1/1970	31217	3896	Nam Tok Baripath, Satun	Thailand
Boesenbergia prainiana	Boesenbergia		Kai Larsen	Flora of Thailand, Botani..	1/1/1970	3212	3987	Chalvarin fall, Sungai padi	Thailand
Boesenbergia rotunda	Boesenbergia		Rohani Hassan	University of Malaya Herb..	23/8/1990	41827	RH 121	Wakaf Baru, Kelantan	Kelantan, Malaysia
Burbidgea schizochella	Burbidgea		Halijah Ibrahim	University of Malaya Herb..	10/6/1989	40185	HI 344	Gunung Danum, Danum Valley	Sabah, Malaysia

Figure 4.16: PDF-format report produced by reporting function.

4.10 System Testing

System Testing is the final phase in this project. This phase was conducted to ensure that all function in the system is working, and with no errors. Users are asked to surf the website and fill out a survey form. The survey form is like a checklist consists of several criteria (see Appendix), for user to tick whichever criteria they feel satisfactory or cross whichever they feel unsatisfactory.

As a result, there is no bug or error detected. All function in the system is working, and no any inaccuracy, misspelling, or dead links reported. However, users have provided us some valuable comments. These comments are useful, as they provide guide for system improvement works in future. Those comments have been categorized and are shown in Table 4.1.

Table 4.1: Users' comments in survey form

Category	Users' comments
Interface	• Larger font • Include further morphological description
Information search	• Better if have more functionality
Administration	• More form validation in adding and editing record • Provide guide on what to input in the form • Prompt out warning of deletion before actually delete the record
Reporting	• Include numbering in report to show the total records exported

CHAPTER 5

DISCUSSIONS

5.1 System Functionality

In this project, a database system was developed with several functions. These fundamental functions meet the functional requirements that have been determined and described in Chapter 2. Basically, there are three group of functions: information search, administrative, and reporting.

Information search function is accessible for all users. This function allows users to retrieve their desired information from database. By inserting text query as input, users may retrieve a particular species record from the database, given that the query matches the parameter. Several parameters can be used to perform the search and each database has different search parameter due to different data scheme of data files. For example, search parameters of ginger database are binomial name, genus, collector's name and locality, while search parameters of bat database are binomial name, genus, locality, class and subregion. These parameters enable users to search information based on different criteria whichever users prefer.

Administrative functions are intended to manage the database. These functions are accessible only by administrator after logging in and allow administrator to add, delete, and edit record. These functions are mainly used in managing species record, but are used in

managing other record as well, such as genus record or collector record. This enables administrator to update the database with the latest information.

Reporting functions allow users to export information of selected species record in a report. There are two types of reporting: single or multiple species. Single species reporting produces one-page PDF file containing all information of a particular species. This function is accessible to all users and can be accessed in detail page of species record. Multiple species reporting produces report containing information of multiple species and is accessible only for administrators. Before produce the report, administrators can choose which data they want to include in the report and which species record they want to export. Administrator can even choose to produce the report in either Microsoft Excel format or PDF format. These parameters enable administrator to conveniently produce a report in a desired format and contains only the information they want.

5.2 System Development Limitation

In this project, using the same framework, three databases were built. This is due to the difference in data schemes. For example, fern data file contain information of distribution while ginger data file does not. To integrate all data in three data files into one database, data integration module is required. However, it is infeasible to produce the module due to time limitation of this project.

Besides, most data is stored in the text format. This is because most of the data contains natural language term. For example, the distribution of fern species *Christensenia aesculifolia* is “from Assam southwards throughout Malaysia”. The boundary of

geographical area of this species is vaguely stated. Proper GPS coordinates are needed to document the precise locality. Most of the information documented by the experts in text files did not have precise locality information thus the database cannot store exact locality information.

Besides, ginger and bat databases do not contain image data. This is a limitation as image data is very resourceful in providing visual clue for researchers when searching information, thus increasing the efficiency of information sharing.

5.3 Future Work

Three databases were created to store and retrieve biodiversity information of ginger, fern and bat species. The basic function of information search and database management were built and provided. However, this system can be further enhanced via two approaches: addition of more useful data and improvement of the system.

In term of storing additional useful information, it is recommended to add molecular information regarding a species so that users could know about all the molecular work that has been done for the species and genetic composition of that species. Besides, medicinal value for certain ginger and fern species can be added to the database over time. Some species has medicinal value and has been used by local natives to treat illness. This type of information is useful for other researchers, especially in the field of drug discovery.

Besides molecular and medicinal value information, it would be beneficial if disease information is added into the database. Disease information includes the information of

diseases that could cause high mortality rate of a particular species, information of zoonotic pathogens where their natural reservoir is of a particular species, and risk of transmission of pathogens to human.

As for system improvement, there are many works could be done. Firstly, as there are three different databases storing information of three different groups of biological organisms, it would be more convenient to integrate information of all three databases into one. So that in the future, users will be able to conveniently search for any information in one database. This would save researchers a lot of time in retrieving information.

Secondly, it is recommended to add the Geographic Information System (GIS) module to the database system. The database system now is able to show the locality of a species on a map, but is not able to integrate and analyze locality information of multiple species in database. Geographical information of species is one of useful information and may provide many knowledge upon analysis. With the integration of databases and employment of GIS, it is possible for users to know whatever species that have occurred within a geographical area, observe different species' distribution pattern in a specific geographical area and deduce relationship among them.

Thirdly, the improvement work involves development of mobile application on smart phone. Within these few years, use of smart phone has become very common and mobile app development has been rapid since then. Thus, it is recommended to develop a mobile app that synchronize with the database system. Mobile app enables users to search information online using smartphone, and this would benefit field data collector significantly. Field data collectors collect specimen at remote area and often have no access

to laptop, but as long as there is internet coverage, they still can retrieve information from database using smartphone. Besides retrieving information, the mobile app can be developed to assist field data collectors in specimens' collection. Whenever collectors collect a specimen at field, they can use the mobile app to record the collection information and determine the coordinate of the spot where the species is found. This information will be input directly to the database and utilized by GIS implemented.

5.4 Summary

Three databases were built to store biodiversity information of ginger, fern and bat species respectively. These databases were developed with the intention of assisting researchers to search, collect, store and share biodiversity information. The database systems still have many works to be done, such as integration of three databases, adding image data, and adding geographic information system (GIS) modules. These further enhancements would increase the efficiency of usage of biodiversity information and possibly contribute to other research purposes, such as drug discovery, rare species conservation and invasive species monitoring.

REFERENCES

Baird, R. (2010). Leveraging the fullest potential of scientific collections through digitization. *Biodiversity Informatics,* 7, 130-136.

Blancou, J., Chomel, B. B., Belotto, A. &Meslin, F. X. (2005). Emerging or re-emerging bacterial zoonoses: factors of emergence, surveillance and control. *Veterinary Research, 36*(3), 507-522. doi:10.1051/vetres:2005008

Canhos, V. P., Souza, S., Giovannia, R., &Canhos, D. A. L. (2004). Global biodiversity informatics: setting the scene for "new world" of ecological modeling. *Biodiversity Informatics, 1,* 1-13.

Ch'ng, E. (2009). An efficient segmentation algorithm for entity interaction. *Biodiversity Informatics,6*, 5-17.

De Queiroz, K. (2007). Species concept and species delimitation. *Systematic Biology,56*(6), 879-886. doi:10.1080/10635150701701083

Encyclopedia of Life (EOL). Retrieved November 10, 2014. http://eol.org/

Faulkner, K. T., Robertson, M. P., Rouget, M. & Wilson, J. R. U. (2014). A simple, rapid methodology for developing invasive species watch lists. *Biological Conservation, 179,* 25–32. doi:10.1016/j.biocon.2014.08.014

Flora Malesiana. Retrieved November 10, 2014. http://dev.e-taxonomy.eu/dataportal/flora-malesiana/

Forest Research Institute Malaysia (FRIM) (2011). *Invasive alien species*. Retrieved November 14, 2014 from Biological Diversity Clearing House Mechanism: http://www.chm.frim.gov.my/

Global Biodiversity Information Facility (GBIF). Retrieved November 10, 2014. http://www.gbif.org/

Harrington, J. L. (2009). *Relational database design and implementation.* Burlington, Massachusetts: Morgan Kaufmann Publishers.

Integrated Taxonomic Information System (ITIS). Retrieved November 10, 2014. http://www.itis.gov/

Knapp, S. (2010). *What's in a name? A history of taxonomy*. Retrieved November 24, 2014 from Natural History Museum: http://www.nhm.ac.uk/

Kumar, S. (2005). *Molecular biology database/biological database/bioinformatics database - an overview*. Retrieved November 6, 2014 from Bioinformaticsweb: http://bioinformaticsweb.net/

Lewinsohn, T. M. & Prado, P. I. (2005). How many species are there in Brazil? *Conservation Biology, 19*(3), 619-624. doi:10.1111/j.1523-1739.2005.00680.x

Ministry of Science, Technology and Environment (1998). *Malaysia's national policy on biological diversity*. Retrieved from Office of the Prime Minister Putrajaya, Malaysia: http://www.pmo.gov.my/

Pigott, D. M., Golding, N., Mylne, A., Huang, Z., Henry, A. J., Weiss, D. J., & … Hay, S. I. (2014). Mapping the zoonotic niche of Ebola virus disease in Africa. *eLife 2014*. doi:10.7554/eLife.04395

Selecting a development approach (2008). Retrieved November 10, 2014 from Centers for Medicare & Medicaid Services: http://www.cms.gov/

Soberon, J. & Peterson, A. T. (2004). Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society B: Biological Sciences, 359,* 689–698. doi:10.1098/rstb.2003.1439

The Herbarium Catalogue, Royal Botanic Gardens, Kew. Retrieved November 10, 2014. http://www.kew.org/herbcat

The Plant List (2013). Version 1.1. Retrieved November 10, 2014. http://www.theplantlist.org/

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., & ... Vieglais, D. (2012). Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE*, *7*(1), e29715. doi:10.1371/journal.pone.0029715

Woese, C. R., Kandler, O., &Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America, 87*, 4576-4579.

APPENDIX

Survey form used in system testing phase.

System Testing Checklist

Reviewer:

Review Date:

INTERFACE

Format	✓/×	Comment
The format is standard for all pages?		
Content		
The fonts are same for all pages?		
The content has no spelling error?		
The content presents accurate information?		
The content is adequate? If no, what information you would like to add to the website?		
The content is appropriate? If no, what information you would like to remove from website?		
Images		
Do the images appear in the webpage?		
The images appeared are relevant?		
Links		
The links are recognizable?		
All links are in standard color?		
PDF file		
The information is neatly arranged?		
There is any information you would like to add into or remove from the PDF file?		

Appendix, continued

FUNCTIONALITY

Search	✓/×	Comment
The search feature retrieves accurate results?		
Map		
Does map show up in ginger information page?		
Does the map show the accurate location?		
Links		
All links direct to right webpage?		
Log in		
Login feature is functional? Is there any error?		
Does it show error message when invalid login is performed?		
Forms		
Does it reject invalid input?		
Administration		
Editing information is functional? Is there any error?		
Deleting a record is functional? Is there any error?		
Please comment on how would you like to improve in administration section?		
Reporting		
It produces report in chosen format?		
The report contains only the correct, chosen fields?		
The report contains only the correct, chosen records?		
Please comment on how would you like to improve reporting feature?		

Appendix, continued

ADDITIONAL COMMENT: