

Computational Analysis & Development Of A Web Based Database For Functional Enrichment

Dissertation submitted in partial fulfillment of the requirement for the degree of

BACHELOR OF TECHNOLOGY IN BIOINFORMATICS

By

ARSH CHADHA

AAYUSHI JAIN

TARUNA THAKUR

VASUNDHRA SINGH

UNDER THE GUIDANCE OF

DR. TIRATHA RAJ SINGH

To



DEPARTMENT OF BIOTECHNOLOGY & BIOINFORMATICS

JAYPEE UNIVERSITY OF INFORMATION AND TECHNOLOGY

WAKNAGHAT, SOLAN (H.P) 173 234

**JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY
WAKNAGHAT, HIMACHAL PRADESH**



CERTIFICATE

This is to certify that the work titled ” **Computational Analysis & Development Of A Web Based Database For Functional Enrichment**” submitted by **Arsh Chadha(121512), Ayushi Jain(1213804), Taruna Thakur(121507)** and **Vasundhra Singh(121514)** in partial fulfilment for the award of degree of B.Tech Bioinformatics of **Jaypee University Of Information Technology, Wagnaghat** has been carried out under my supervision. This work has not been submitted partially or wholly to any other university or institute for the award of this or any other degree or diploma.

Signature of Supervisor :



Name of Supervisor : Dr. Tiratha Raj singh

Designation : Assistant Professor, JUIT

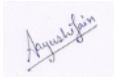
Date : 27th May, 2016

DECLARATION

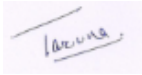
We hereby declare that the work reported in the B-Tech thesis entitled “**Computational Analysis & Development Of A Web Based Database For Functional Enrichment**” submitted at **Jaypee University Of Information Technology** is an authentic record of my work carried out under the supervision of Dr. Tiratha Raj Singh . I have not submitted this work elsewhere for any other degree or diploma.



Arsh Chadha (121512)



Aayushi Jain (123806)



Taruna Thakur (121507)



Vasundhra Singh (121514)

Department of Biotechnology & Bioinformatics

Jaypee University of Information Technology, Waknaghat, India

Date: 27th May,2016

ACKNOWLEDGEMENT

This dissertation would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study.

First and foremost , my utmost gratitude to **Dr.Tiratha Raj Singh** , Senior Lecturer of the Department of Biotechnology and Bioinformatics (JUIT) , for the unfailing support as our project adviser, and for his patience and steadfast encouragement to complete this study. He has been an inspiration as we hurdled through all the obstacles in the completion of this research work.

Apart from these countless events, countless people and several incidents have made contribution to this project that is indescribable. We again express our gratitude to them. We indebt to all those who provided reviews and suggestions for improving the results and topics covered in our project, extend our apologies to any one whom we have failed to recognize in this effort of ours.

All copyrights that are cited in this document remain the property of their respective owner(s).

ABSTRACT

Databases hold vast importance in the field Research and Development . This statement implies that databases are important for the retrieval of data . Any kind of data that the scientific community works upon is usually downloaded from the internet. Hence, we aim to develop such kind of a database that helps the scientific community to extract and utilize the data in an effective way.

We aim to present a comprehensive database which could help in efficient extraction of information regarding the plants. The reason for developing such kind of a database is that it is the first database to incorporate such information. Hence, such kind of a database would be helpful to provide information that could be easily accessed and utilized.

LIST OF FIGURES

Figure 1. Waterfall Model	14
Figure 2. Database Structure	47
Figure 3. Options in GUI.....	50
Figure 4. Home Page GUI.....	51
Figure 5. About Page GUI.....	52
Figure 6. Search Option GUI.....	53
Figure 7. Economical Plant Database.....	57
Figure 8. Model Plants DB Structure	58
Figure 9. Model Plants DB Structure	58
Figure 10. PHP Code For Download-1	58
Figure 11. PHP Code For Download-2	59
Figure 12. PHP Code For Download-3	59
Figure 13. PHP Code For Download-4	60
Figure 14. Specific Search.....	62
Figure 15. Result For Specific Search.....	62
Figure 16. Advanced Search.....	63
Figure 17. Advanced Search Example	63
Figure 18. Search Results.....	64

LIST OF TABLES

Table 1. Software Requirements	16
Table 2. Information Source Table.....	18
Table 3. Gene Table	44
Table 4. Protein Table	45
Table 5. Taxonomy Table	45
Table 6. Miscellaneous Features Table	46
Table 7. Gene Ontology Table	46

CONTENTS

CERTIFICATE	ii
DECLARATION.....	iii
ACKNOWLEDGEMENT.....	iv
ABSTRACT.....	v

LIST OF FIGURES

LIST OF TABLES

CHAPTER -1

INTRODUCTION

1.1 GENERAL BACKGROUND.....	11
1.2 OBJECTIVE.....	11
1.3 DESCRIPTION OF THE PROJECT.....	11

CHAPTER-2

PROJECT STUDY

2.1 STUDY OF EXISTING WEBSITES.....	12
2.2 DEVELOPED WEBSITE.....	12
2.3 WATERFALL METHODOLOGY IN WEB DEVELOPMENT.....	12

CHAPTER-3

PROJECT REQUIREMENT ANALYSIS

3.1 PROJECT DEFINITION.....	15
3.1.1 WEBSITE REQUIREMENT	15

3.1.2 FUNCTIONALITY OF DATABASE.....	15
3.1.3 SYSTEM ENVIRONMENT.....	16
3.2 FEASIBILITY ANALYSIS.....	16
3.3PROJECT PLAN.....	17
3.4 PROGRAMMING LANGUAGES AND DEVELOPMENT TOOLS.....	17

CHAPTER-4

DATA COLLECTION AND NORMALIZATION

4.1 CATEGORIZATION O F PLANT SPECIES.....	18
---	----

CHAPTER-5

DATABASE DEVELOPMENT

5.1 DATABASE DESCRIPTION.....	44
5.2 TABLES DESCRIPTION.....	44
5.3 DATABASE STRUCTURE.....	47

CHAPTER-6

DEVELOPMENT AND DESIGN

6.1 SYSTEM REQUIREMENTS.....	48
6.2 LANGUAGES USED.....	49
6.3 GUI DEVELOPMENT.....	50
6.4 BACK END WEB DESIGINING.....	53

CHAPTER-7

DATABASE CONNECTIVITY AND DATA RETRIEVAL

7.1 DATABASE CONNECTIVITY.....57

7.2 DATA RETRIEVAL.....58

CHAPTER-8

RESULTS AND CONCLUSION

8.1 BASIC SEARCH.....61

8.2 ADVANCE SEARCH.....63

REFERENCES.....65

CHAPTER 1

INTRODUCTION

1.1 GENERAL BACKGROUND

Genomic research makes it possible to look at biological phenomena on a scale not previously possible, which lead to production of massive quantities of data. In our project we are creating a database named PLANTEA DATABASE for diverse plant species with GUI. The PLANTEA Database provides standardized information about the medicinal, economical and model plants. It includes names, genetic information, protein information, sequential information, taxonomy, classification, miscellaneous features and onward Web links. In addition the underlying data is usually available for download in a variety of formats. These formats include XML, CSV, TAB separated data.

1.2 OBJECTIVE

- User friendly database search.
- Comprehensive collection of data.
- Knowledge based and bulk analysis of diverse plant species.
- Making this application available for the researchers for further research conduction.

1.3 DESCRIPTION OF THE PROJECT

Requirement: Requirement of a website to store information of diversified plant species as a database, process this information by researchers (e.g. viewing, updating, deleting, adding new details etc.), dynamic searching and comprehensive collection of data is retrieved from one destination or website.

Deliverables : A Website Based on Diversified Plant Species

Project Site : Jaypee University of Information Technology, Wagnaghat

CHAPTER 2

PROJECT STUDY

This section outlines details of existing and proposed website.

2.1 Study of Existing Websites

The developed website is similar to any other resource or website i.e., it contains all the facilities for a user to view and access the data from this website.

2.2 Developed Website

The developed website contains important information of diversified plant species with different categories such as Economical, Medicinal and Model. The data entered into the website is of utmost importance in the botanical world and research.

Salient features of the Website

- It makes efficient use of the resources and time involved in the whole accessing of data on this website.
- Concurrent users can access the website simultaneously.
- Integrity and safety of data is maintained.
- .Data is computationally generated hence, no manual processing is required.

2.3 Waterfall Methodology in Web Development

A development Process model specifies some activities that according to the model should be performed and it also specifies the order in which they should be performed. These activities are related to the production of the website, for example, design, coding, testing. Generally a development process is phased.

The development process specifies the major development and quality assurance activities that need to be performed in the process, thus it forms the core of the website development process. Also the management step is also a part of the development process. A process model for software engineering is chosen based on the nature of the project and application, the methods and tools to be used and the control and the deliverables that are required

We have used the **Waterfall Model** as the web development Process model in this Project. The web environment provides a few challenges to the waterfall model and requires some

adaptation. The most of the same sequential phases are there, but their contents vary to suit this dynamic environment.

Naturally, the first phase is requirement gathering. Fortunately, there is only one requirement--a catchy domain name. Ideally, this would be a common English word, but absolutely all dictionary words, in all languages (with the exception of 7 obscure and unpleasant diseases in a little known Klingon dialect) are already taken so brainstorming is required to find and register a pronounceable misspelling. This is generally the most time consuming phase of web software development.

The design phase is generally far less time consuming. The only important design elements to tie down are those that affect the project's buzzword compliance. The project must be implemented using trendy technology and some forethought is required as trends may change over the three months it will take to implement.

The only thing set in stone about the implementation phase is that it must take the longest time. Interestingly nobody really knows how this number came about. Nutritionists speculate that it may be because that is the period that an average programmer can subsist solely on employer provided pizza and soda before critical body systems start shutting down from malnutrition. Fortunately, programmers are no longer required in a waterfall based web project after the implementation phase, so their longer term viability is not important.

Verification and testing of web systems works a little differently to most other systems. It has been naively suggested that web application developers simply write 'Beta' at the top of all pages and launch, but it is not quite as simple as that. You also need to generate buzz among alpha-nerds and only let people use your Beta system if they have been invited by a friend. That way, they will feel so special that they will only want to give positive feedback. Positive feedback is all you are interested in as there is little point in gathering defects that you have no intention of correcting before sale.

While a traditional waterfall model ends in a Maintenance phase, a web system instead enters the "Exit Strategy Phase. The waterfall model was derived from engineering models to put some order in the development of large software products. It consists of different stages which are processed in a linear fashion. Compared to other software development models it is more rigid and better manageable. The waterfall model is an important model which is the basis of many other models; however for many modern projects it has become a little outdated. It is still wildly used.

There are usually five stages in this model of website development:

1. **Requirements analysis and definition.** In this stage the requirements of the "to be developed software" are established. These are usually the services it will provide, its constraints and the goals of the project. Once these are established they have to be defined in such a way that they are usable in the next stage.

2. **Database and website design.** In this stage the established requirements, flowing from the first stage, are identified as software or hardware requirements. The project requirements are then translated in such a way that they can be readily transformed into computer programs.
3. **Implementation and unit testing.** This is the stage where the computer programs are created. Each program is called a unit, and unit testing is the verification that every unit meets its specification.
4. **System testing.** All the units are combined and now the whole is tested. When the combined programs are successfully tested the website product is finished.
5. **Operation and maintenance.** Most website products include this stage of the development. It involves correcting errors that have gone undetected before, improvement and other forms of support. This stage is part of the life cycle of a web product, and not of the strict development, although improvements and fixes can still be considered as "development".

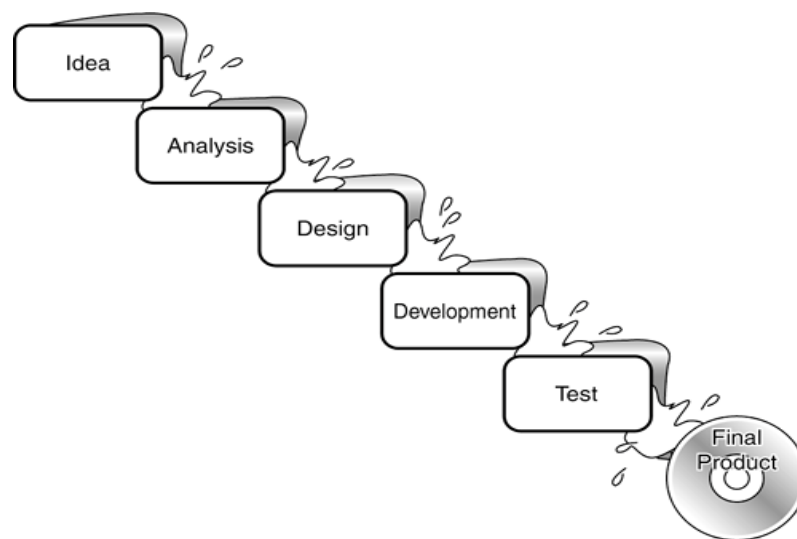


Figure 1. Waterfall Model

CHAPTER 3

PROJECT REQUIREMENT ANALYSIS

3.1 Project Definition

This section projects the scope of the project and also gives an idea to what kind of solution strategies were adopted for the completion of this project.

3.1.1 Website Requirement

To make a user friendly website.

Input required by the Website

A Database (First Requirement):-

- Details of the plants.
- Genetic information, protein information, sequential information, taxonomy, classification, information regarding Gene Ontology etc.
- A GUI template.
- The website comprises three databases for managing the data : Medicinal plants, Economical Plants & Model Plants.

Output to be Produced by the Website

This Website displays the details of the desired plants. The user can view the details, but neither can change the format of the details nor modify the data or text.

3.1.2 Functionality of PLANTAE Database

The PLANTAE Database has got five tables:

1. Taxonomy Table
2. Gene ontology Table
3. Gene Table
4. Protein Table
5. Miscellaneous features

Server-Side Functionality:

The server application involves storing of various tables used in the project and the various procedural calculations that are done on the server side in order to improve the throughput of the system.

Client-Side Functionality:

The client application consists of various forms for data retrieval and downloading which are generated on the basis of given requirements.

3.1.3 System Environment

Minimum hardware requirements: PC or laptop, 1GB RAM

Software requirements:

Client Environment	Server Environment	Development Platform
Any Browser	WAMP Server 2.0 MySQL ErWin	Windows 7/8/8.1/10/Vista/XP Oracle D2K (Forms 4.5 & Reports 2.5)

Table 1. Software Requirements

Network:

It is an offline website. We need wampServer and a browser to use this website.

3.2 Feasibility Analysis

Requirement analysis comprises of initial understanding of why the system is needed and pros and cons of the system. It consists of what is called Feasibility Study. All projects are feasible given unlimited resources and infinite time. Unfortunately the development of the computer based system or product is more likely to be plagued by a scarcity of resources and difficult delivery dates. Thus it is both necessary and prudent to evaluate the feasibility of the project at the earliest possible time. During Feasibility analysis, our team took the opportunity and talked to the respective supervisor on various topics including the cost effectiveness, time allotted for the completion of the project.

Technical feasibility

The proposed system was assessed to be technically viable as the technologies decide to be used in the project were available and were capable of accomplishing the desired function and performance. The scope of extensibility was also taken into Consideration as the IT industry is still on a speedy highway and various new technologies come up with a blink of eye. Technical feasibility centres on the existing computer system (hardware, software, etc.) and to what extent it can support the proposed addition. It may involve financial and other considerations to accommodate technical enhancements.

3.3 Project Plan

A Website Project Plan is produced at the culmination of the planning tasks. It provides baseline cost and scheduling information that will be used throughout the designing of a website. The plan concentrates on a general statement of what and a specific statement of how much and how long. The Project Plan consisting of all the relevant details about this project is as follows:

3.4 Programming Languages and Development Tools

The GUI for the project and database development is done using HTML, CSS, JavaScript, PHP and MySql. Much of the procedural calculations are done using PHP/SQL procedures at the server side. The details for the same are provided in later chapters.

CHAPTER 4

DATA COLLECTION AND NORMALIZATION

This chapter contains all the information regarding the different categories of plants such as economical, medicinal and model. There are five tables containing different features in three different databases.

4.1 Data collection

We have used various sources for the data collection. Here the list is provided for data collected and their sources:

Sr. No	Source	Information	Links
1	UNIPROT	Ontology	http://www.uniprot.org
2	NCBI	Taxonomy	http://www.ncbi.nlm.nih.gov/taxonomy
3	NCBI	Protein(FASTA) Sequence	http://www.ncbi.nlm.nih.gov/protein
4	NCBI	Gene Information	http://www.ncbi.nlm.nih.gov/gene
5	TAIR	<i>A. thaliana</i> Information	https://www.arabidopsis.org
6	PlantSeed	Protein(FASTA) Sequence	http://bioseed.mcs.anl.gov
7	PlantSeed	Genomic Data(FASTA)	http://bioseed.mcs.anl.gov
8	KEGG	Pathways	http://www.genome.jp/kegg/pathway.html

. Table 2. Information Source Table

4.2 Categorization of Plants Species

Plants have been categorized into three major categories while few plants share characteristics from all these categories. Details are given below:

1. MEDICINAL PLANTS

- *Acacia concinna (Shikakai)*

Kingdom:Plantae

(unranked):Angiosperms

(unranked):Eudicots

(unranked):Rosids

Order:Fabales

Family:Fabaceae

Genus:Acacia

Species:A.concinna

Description :Acacia concinna is a climbing shrub native to Asia, common in the warm plains of central and south India. The tree is food for the larvae of the butterfly *Pantoporia hordonia*. Alkaloids are found in the tree's fruit.

- *Aegle marmelos (Bael)*

Kingdom:Plantae

(unranked):Angiosperms

(unranked):Eudicots

(unranked):Rosids

Order:Sapindales

Family:Rutaceae

Subfamily:Aurantioideae

Tribe:Aurantieae

Genus:AegleCorrêa

Species:A.marmelos

Description : The leaves are compound, trifoliolate and with a peculiar fragrance. The leaflets are oval or lancet shaped. The lateral leaflets are without petiole and the terminal one has a long one. The flowers are greenish white and fragrant. The fruit is both edible and medicinal. It is spherical or oval. The seeds are small, hard. Multiple seeds are covered in a viscid material in the pulp.

- *Aloe barbadensis (Aloe vera)*

Kingdom: Plantae

(unranked): Angiosperms

(unranked): Eudicots

(unranked):Rosids

Order: Brassicales

Family: Brassicaceae

Genus:Arabidopsis

Species:A.thaliana

Description : Arabidopsis thaliana (arabidopsis) is a small flowering plant. It is a popular model organism in plant biology and genetics. For a complex multicellular eukaryote, Arabidopsis thaliana has a relatively small genome of approximately 135 megabase pairs (Mbp). It was long thought to have the smallest genome of all flowering plants, but the smallest flowering plants' genomes are now considered to belong to plants in the genus Genlisea, order Lamiales, with Genlisea tuberosa, a carnivorous plant, showing a genome size of approximately 61 Mbp. Arabidopsis thaliana was the first plant to have its genome sequenced, and is a popular tool for understanding the molecular biology of many plant traits, including flower development and light sensing

- *Alstonia scholaris (Satwin, chattin)*

Kingdom: Plantae

(unranked): Angiosperms

(unranked): Eudicots

(unranked): Asterids

Order: Gentianales

Family: Apocynaceae

Tribe: Plumeriae

Subtribe: Alstoniinae

Genus: Alstonia

Species: A. scholaris

Description : It is a medium sized to a large evergreen tree with tall, straight often buttressed and fluted stem. Whorled branches and dark – grey shining leaves also in whorls. Bark is blackish grey, rough with shallow cracks. Leaves 4-10 in a whorl, obovate & obtuse, gradually narrowed into the petiole, dark green above, pale beneath. Flowers greenish white, in panicles of cymes on stout peduncles, follicular fruit.

- *Anthocephalus chinensis* (*Bur flower tree*)

Kingdom: Plantae

(unranked): Angiosperms

(unranked): Eudicots

(unranked): Asterids

Order: Gentianales

Family: Rubiaceae

Subfamily: Cinchonoideae

Tribe: Naucleaeae

Genus: Neolamarckia

Species: N. cadamba

Description : Kadam is a large, tall tree, with a straight cylindrical bole. Foliage consists of elliptic-oblong leaves. Flowers are small, orange-coloured in dense terminal globose heads. The fruit is a pseudo carp of closely packed capsules each containing minute, singular seeds.

- *Araucaria angustifolia* (*Candelabra tree*)

Kingdom: Plantae

Division: Pinophyta

Class: Pinopsida

Order: Pinales

Family: Araucariaceae

Genus: Araucaria

Species: *A. angustifolia*

Description : It is an evergreen tree .The leaves are thick, tough and scale like, triangular, and with razor-sharp edges and tip. They persist for 10-15 years, so cover most of the tree except for the trunk and older branches. It is usually dioecious, with the male and female cones on separate trees. The male (pollen) cones are oblong. The female (seed) cones, are globose, and hold about 100-150 seeds. The cones disintegrate at maturity to release the nut-like seeds.

- *Azadirachta indica* (*Neem tree*)

Kingdom: Plantae

(unranked): Angiosperms

(unranked): Eudicots

(unranked): Rosids

Family: Sapindales

Subfamily: Meliaceae

Genus: Azadirachta

Species : *A. indica*

Description : Neem is a fast-growing tree that can reach a height of 15–20 metres (49–66 ft), and rarely 35–40 metres (115–131 ft). It is evergreen, but in severe drought it may shed most

or nearly all of its leaves.. The (white and fragrant) flowers are arranged in more-or-less drooping axillary panicles which are up to 25 centimetres (9.8 in) long. Protandrous, bisexual flowers and male flowers exist on the same individual tree. The fruit is a smooth (glabrous) olive-like drupe which varies in shape from elongate oval to nearly roundish. The fruit skin (exocarp) is thin and the bitter-sweet pulp (mesocarp) is yellowish-white and very fibrous. The mesocarp is 0.3–0.5 centimetres (0.12–0.20 in) thick. The white, hard inner shell (endocarp) of the fruit encloses one, rarely two or three, elongated seeds (kernels) having a brown seed coat.

- *Bauhinia purpurea* (*Purple Orchid Tree*)

Kingdom: Plantae

(unranked): Angiosperms

(unranked): Eudicots

(unranked): Rosids

Order: Fabales

Family: Leguminosae

Genus: Phanera

Species: P.purpurea

Description : Moderate sized evergreen tree. Young parts are covered with brown pubescence. Leaves cordate, nerved, the twin leaflets fold at night. Flowers large, showy, fragrant, rosy purple or lilac, in short peduncled terminal panicles. Seeds oblong, dark- brown, smooth and compressed.

- *Castanospermum australe* (*Moreton Bay chestnut*)

Kingdom: Plantae

(unranked): Angiosperms

(unranked): Eudicots

(unranked): Rosids

Order: Fabales

Family:Fabaceae

Subfamily:Faboideae

Tribe:Angylocalyceae

Genus:Castanospermum

Species:C.australe

Description : It is a large evergreen tree growing to 40 m tall, though commonly much smaller. The leaves are pinnate. The flowers are bicoloured red and yellow, produced in racemes. The fruit is a cylindrical pod, the interior divided by a spongy substance into three to five cells, each of which contains a large chestnut-like seed.

- *Cinnamomum tamala (Tejapatta)*

Kingdom: Plantae

(unranked): Angiosperms

(unranked):Magnoliids

Order:Laurales

Family:Lauraceae

Genus: Cinnamomum

Species: C.tamala

Description :Cinnamomum tamala, Indian bay leaf, also known as tejpatta. It has aromatic leaves which are used for culinary and medicinal purposes. It is thought to have been one of the major sources of the medicinal plant leaves known in classic and medieval times as malabathrum (or malobathrum). They are often labeled as "Indian bay leaves," or just "bay leaf", causing confusion with the leaf from the bay laurel, a tree of Mediterranean origin in a different genus, and the appearance and aroma of the two are quite different. Bay laurel leaves are shorter and light to medium green in color, with one large vein down the length of the leaf, while tejpatleaves are about twice as long and wider, usually olive green in color, and with three veins down the length of the leaf. True tejpat leaves impart a strong cassia- or cinnamon-like aroma to dishes, while the bay laurel leaf's aroma is more reminiscent of pine and lemon. Indian grocery stores usually carry true tejpat leaves.

- *Citrus maxima (Pomelo)*

Kingdom: Plantae

(unranked): Angiosperms

(unranked): Eudicots

(unranked): Rosids

Order: Rosids

Order: Sapindales

Family: Rutaceae

Genus: Citrus

Species: C. maxima

Description : Fruit is usually a pale green to yellow when ripe (but also pink or red), with sweet flesh and thick spongy rind. It is the largest citrus fruit.

- *Curcuma longa (Turmeric)*

Kingdom: Plantae

(unranked): Angiosperms

(unranked): Monocots

(unranked): Commelinids

Order: Zingiberales

Family: Zingiberaceae

Genus: Curcuma

Species: C. longa

Description : Turmeric (*Curcuma longa*) is a rhizomatous herbaceous perennial plant of the ginger family, Zingiberaceae. It requires temperatures between 20 and 30 °C (68 and 86 °F) and a considerable amount of annual rainfall to thrive. Plants are gathered annually for their rhizomes and propagated from some of those rhizomes in the following season. When not used fresh, the rhizomes are boiled for about 30–45 minutes and then dried in hot ovens, after which they are ground into a deep-orange-yellow powder commonly used as a spice in Bangladeshi cuisine, Indian cuisine, Pakistani cuisine and curries, for dyeing, and to impart color to mustard condiments. One active ingredient is curcumin, which has a distinctly earthy,

slightly bitter, slightly hot peppery flavor and a mustardy smell. India, a significant producer of turmeric, has regional names based on language and state.

- *Curcuma zedoaria* (*White turmeric*)

Kingdom: Plantae

(unranked): Angiosperms

(unranked): Monocots

(unranked): Commelinids

Order: Zingiberales

Family: Zingiberaceae

Genus: Curcuma

Species: C.zedoaria

Description : Curcuma zedoaria white turmeric or kentjur, is the name for a perennial herb and member of the genus Curcuma, family Zingiberaceae. The plant is native to India and Indonesia but now naturalized in other places including the US State of Florida. It was introduced to Europe by Arabs around the sixth century, but its use as a spice in the West today is extremely rare, having been replaced by ginger. Zedoary is a rhizome that grows in tropical and subtropical wet forest regions. The fragrant plant bears yellow flowers with red and green bracts and the underground stem section is large and tuberous with numerous branches. The leaf shoots of the zedoary are long and can reach 1 meter (3 feet) in height.

- *Delonix regia* (*Flame Tree, Gulmohar*)

Kingdom: Plantae

(unranked): Angiosperms

(unranked): Eudicots

(unranked): Rosids

Order: Fabales

Family: Fabaceae

Subfamily: Caesalpinioideae

Genus: Delonix

Species: D.regia

Description : Moderate sized fast growing, deciduous tree and light feathery foliage. Leaves bipinnate, at base of leaflet two stipules occur. Flowers appear in corymbs along and at the ends of branches. Pods are green and flaccid when young and turn dark-brown and woody.

- *Ginko biloba* (*Maidenhair tree*)

Kingdom: Plantae

Division: Ginkgophyta

Class: Ginkgopsida

Order: Ginkgoales

Family: Ginkgoaceae

Genus: Ginkgo

Species: G.biloba

Description : The tree has an angular crown. The leaves are unique among seed plants, being fan-shaped with veins radiating out into the leaf blade, sometimes bifurcating. The old popular name "Maidenhair tree" is because the leaves resemble some of the pinnae of the Maidenhair fern. Ginkgos are dioecious. Male plants produce small pollen cones with sporophylls each bearing two micro sporangia spirally arranged around a central axis. Female plants do not produce cones. Two ovules are formed at the end of a stalk, and after pollination, one or both develop into seeds.

- *Ocimum tenuiflorum* (*Holy Basil*)

Kingdom: Plantae

(unranked): Asterids

Order: Lamiales

Family: Lamiaceae

Genus: Ocimum

Species: O.tenuiflorum

Description : *Ocimum tenuiflorum*, also known as *Ocimum sanctum*, holy basil, or tulasi (also spelled thulasi), is an aromatic plant in the family Lamiaceae which is native to the Indian subcontinent and widespread as a cultivated plant throughout the Southeast Asian tropics. It is an erect, many-branched subshrub, 30–60 cm (12–24 in) tall with hairy stems and simple phyllotaxic green or purple leaves that are strongly scented. Leaves have petioles and are ovate, up to 5 cm (2.0 in) long, usually slightly toothed. The flowers are purplish in elongate racemes in close whorls. Tulasi is cultivated for religious and medicinal purposes, and for its essential oil. It is widely known across the Indian subcontinent as a medicinal plant and a herbal tea, commonly used in Ayurveda, and has an important role within the Vaishnava tradition of Hinduism, in which devotees perform worship involving holy basil plants or leaves. This plant is revered as an elixir of life.

- *Phyllanthus emblica* (*Amla tree*)

Kingdom:Plantae

(unranked):Angiosperms

(unranked):Eudicots

(unranked): Rosids

Order:Malpighiales

Family:Phyllanthaceae

Tribe:Phyllantheae

Subtribe:Flueggeinae

Genus:Phyllanthus

Species:*P.emblica*

Description : *Phyllanthus emblica*, also known as Indian gooseberry or amla from Sanskrit amalika is a deciduous tree of the family Phyllanthaceae. It is known for its edible fruit of the same name. The tree is small to medium in size, reaching 1–8 m (3 ft 3 in–26 ft 3 in) in height. The branchlets aren't glabrous or finely pubescent, 10–20 cm (3.9–7.9 in) long, usually deciduous; the leaves are simple, sessile and closely set along branchlets, light green, resembling pinnate leaves. The flowers are greenish-yellow. The fruit is nearly spherical, light greenish yellow, quite smooth and hard on appearance, with six vertical stripes or furrows. Ripening in autumn, the berries are harvested by hand after climbing to upper branches

bearing the fruits. The taste of Indian emblic is sour, bitter and astringent, and it is quite fibrous. In India, it is common to eat emblic steeped in salt water and red chilli powder to make the sour fruits palatable.

2. ECONOMICAL PLANTS

- *Acacia auriculiformis* (*Ear-leaf Acacia*)

Kingdom: Plantae

(unranked): Angiosperms

(unranked): Eudicots

(unranked): Rosids

Order: Fabales

Family: Fabaceae

Genus: Acacia

Species: A.auriculiformis

Description : Moderate sized ,evergreen tree, bark smooth, leaf stalks are modified into flattened blade called phyllode, which is narrowly oblong, slightly curved or sickle shaped, flower tiny, yellow, fragrant, crowded in axillary spikes. Pods are wide, flat, hard almost woody and much twisted in irregular coils. Initially green but turn brown on ripening.

- *Acer oblongum* - *Himalayan Maple*

Kingdom: Plantae

(unranked): Angiosperms

(unranked): Eudicots

(unranked): Rosids

Order: Sapindales

Family: Sapindaceae

Genus: Acer

Species: A.oblongum

Description : A middle sized tree, leaves oblong or ovate-lanceolate. Flowers pentamerous terminal or leaf bearing lateral corymbose, panicles pedicel and calyx hairy .

- *Adenanthera microsperma* (*Bead Tree*)

Kingdom: Plantae

(unranked): Angiosperms

(unranked): Eudicots

(unranked): Rosids

Order: Fabales

Family: Fabaceae

Genus: Adenanthera

Species: A. pavonina

Description : A medium- to large-size deciduous tree. The tree is generally erect, having dark brown to greyish bark, and a spreading crown. Multiple stems are common, as are slightly buttressed trunks in older trees. The leaves are bipinnate. The alternate leaflets, are oval-oblong with an asymmetric base and a blunt apex. The leaves yellow with age. Flowers are borne in narrow spike-like racemes, at branch ends. They are small, creamy-yellow in colour, and fragrant. The leathery pods curve and twist upon on dehiscence.

- *Allium sativum* (*Garlic*)

Kingdom: Plantae

Clade: Angiosperms

Clade: Monocots

Order: Asparagales

Family: Amaryllidaceae

Subfamily: Allioideae

Genus: Allium

Species: A. sativum

Description : Allium sativum, commonly known as garlic, is a species in the onion genus,

Allium. Its close relatives include the onion, shallot, leek, chive, and rakkyo. With a history of human use of over 7,000 years, garlic is native to central Asia, and has long been a staple in the Mediterranean region, as well as a frequent seasoning in Asia, Africa, and Europe. *Allium sativum* is a bulbous plant. It grows up to 1.2 m (4 ft) in height. Its hardiness is USDA Zone 8. It produces hermaphrodite flowers. Pollination occurs by bees and other insects.

- *Agathis robusta* (Kauri pine)

Kingdom: Plantae

Division: Pinophyta

Class: Pinopsida

Order: Pinales

Family: Araucariaceae

Genus: *Agathis*

Species: *A. robusta*.

Description : Monoecious tree, trunk clear for over half its length; crown dense. Bark is orange-brown, brown or grey-brown; smooth to slightly flaky. Bark exudates a clear or somewhat milky. Leaves spirally arranged on primary shoots, opposite to sub-opposite on lateral shoots; leaves entire on petioles, linear to elliptic, veins fine, longitudinal, more or less parallel. Juvenile leaves similar but oblong-lanceolate, acute, glabrous, green, shiny above and dull beneath, venation faint and longitudinal.

- *Anogeissus pendula* (Chakwa)

Kingdom: Plantae

(unranked): Angiosperms

(unranked): Eudicots

(unranked): Rosids

Order: Rosids

Order: Myrtales

Family: Combretaceae

Genus:Anogeissus

Description : The wood has a lot of mechanical strength and very tough. It is regarded as the 3rd toughest timber in the world. The timber is used for furniture, agricultural implements, tool handles of all kinds, poles, rafters, for carts, spokes of wheels, toys, shuttles and bobbins, tool handles etc.

- *Artocarpus heterophyllus (Jackfruit)*

Kingdom:Plantae

(unranked):Angiosperms

(unranked):Eudicots

(unranked):Rosids

Order: Rosales

Family:Moraceae

Tribe:Artocarpeae

Genus:Artocarpus

Species:A.heterophyllus

Description : The jackfruit (*Artocarpus heterophyllus*), also known as jack tree, jakfruit, or sometimes simply jack or jak is a species of tree in the mulberry and fig family (Moraceae). It is native to parts of South and Southeast Asia, and is believed to have originated in the southwestern rain forests of Indian Subcontinent, in present-day Goa, Kerala, Odisha, Tamil Nadu, Sri Lanka coastal Karnataka, and Maharashtra. The jackfruit tree is well suited to tropical lowlands, and its fruit is the largest tree-borne fruit, reaching as much as 35 kg (80 lb) in weight, 90 cm (35 in) in length, and 50 cm (20 in) in diameter. The Jackfruit tree can produce about 100 to 200 fruits in a year. The jackfruit is composed of hundreds to thousands of individual flowers, and it is the fleshy petals that are eaten.

- *Bombax ceiba (Silk Cotton Tree)*

Kingdom: Plantae

(unranked): Angiosperms

(unranked): Eudicots

(unranked): Rosids

Order: Malvales

Family: Malvaceae

Genus: Bombax

Species: B.ceiba

Description : This tree has a straight tall trunk and its leaves are deciduous in winter. Red flowers with 5 petals appear in the spring before the new foliage. It produces a capsule which, when ripe, contains white fibres like cotton. Its trunk bears spikes to deter attacks by animals.

- *Cedrus deodara* (*Deodar cedar*)

Kingdom: Plantae

Division: Pinophyta

Class: Pinopsida

Order: Pinales

Family: Pinaceae

Genus: Cedrus

Species: C.deodara

Description : *Cedrus deodara* (deodar cedar, Himalayan cedar, or deodar) is a species of cedar native to the western Himalayas in eastern Afghanistan, northern Pakistan (especially in Khyber Pakhtunkhwa) and India (Jammu and Kashmir, Himachal Pradesh, Sikkim, Darjeeling and Uttarakhand states), southwesternmost Tibet in (China) and western Nepal, occurring at 1,500–3,200 m (4,921–10,499 ft) altitude. It is a large evergreen coniferous tree reaching 40–50 m (131–164 ft) tall, exceptionally 60 m (197 ft) with a trunk up to 3 m (10 ft) in diameter. It has a conic crown with level branches and drooping branchlets. The leaves are needle-like, mostly 2.5–5 cm (0.98–1.97 in) long, occasionally up to 7 cm (2.8 in) long, slender (1 mm (0.039 in) thick), borne singly on long shoots, and in dense clusters of 20–30 on short shoots; they vary from bright green to glaucous blue-green in colour. The female cones are barrel-shaped, 7–13 cm (2.8–5.1 in) long and 5–9 cm (2.0–3.5 in) broad, and disintegrate when

mature (in 12 months) to release the winged seeds. The male cones are 4–6 cm (1.6–2.4 in) long, and shed their pollen in autumn.

- *Cocos nucifera* (*Coconut palm*)

Kingdom:Plantae

(unranked):Angiosperms

(unranked):Monocots

(unranked):Commelinids

Order:Arecales

Family: Arecaceae

Subfamily:Arecoideae

Tribe:Cocoeae

Genus:Cocos

Species:C.nucifera

Description :The term coconut can refer to the entire coconut palm, the seed, or the fruit, which, botanically, is a drupe, not a nut. The spelling cocoanut is an archaic form of the word. The term is derived from the 16th-century Portuguese and Spanish word coco meaning "head" or "skull", from the three indentations on the coconut shell that resemble facial features. The coconut is known for its great versatility as seen in the many uses of its different parts and found throughout the tropics and subtropics. Coconuts are part of the daily diets of many people. Coconuts are different from any other fruits because they contain a large quantity of "water" and when immature they are known as tender-nuts or jelly-nuts and may be harvested for drinking. When mature, they still contain some water and can be used as seednuts or processed to give oil from the kernel, charcoal from the hard shell and coir from the fibrous husk. The endosperm is initially in its nuclear phase suspended within the coconut water. As development continues, cellular layers of endosperm deposit along the walls of the coconut, becoming the edible coconut "flesh".

- *Coriandrum sativum* (*Coriander*)

Kingdom:Plantae

(unranked):Angiosperms

(unranked):Eudicots

(unranked):Asterids

Order:Apiales

Family:Apiaceae

Genus:Coriandrum

Species:C.sativum

Description :Coriander is also known as cilantro or Chinese parsley is an annual herb in the family Apiaceae. Coriander is native to regions spanning from southern Europe and northern Africa to southwestern Asia. It is a soft plant growing to 50 cm (20 in) tall. The leaves are variable in shape, broadly lobed at the base of the plant, and slender and feathery higher on the flowering stems. The flowers are borne in small umbels, white or very pale pink, asymmetrical, with the petals pointing away from the center of the umbel longer (5–6 mm or 0.20–0.24 in) than those pointing toward it (only 1–3 mm or 0.039–0.118 in long). The fruit is a globular, dry schizocarp 3–5 mm (0.12–0.20 in) in diameter. The seeds are generally used as a spice or an added ingredient in foods or recipes, although sometimes they are eaten alone.

- *Crocus sativus (Saffron)*

Kingdom:Plantae

(unranked):Angiosperms

(unranked):Monocots

Order:Asparagales

Family:Iridaceae

Subfamily:Crocoideae

Genus:Crocus

Species:C.sativus

Description :Crocus sativus, commonly known as saffron crocus, or autumn crocus, is a species of flowering plant of the Crocus genus in the Iridaceae family. It is best known for the spice saffron, which is produced from parts of the flowers. The cormous autumn-flowering perennial plant species is unknown in the wild. Human cultivation of saffron crocus and use

of saffron is known to have taken place for more than 3,500 years and spans different cultures, continents, and civilizations.

- *Dillenia indica* (*Elephant Apple*)

Kingdom: Plantae

(unranked): Angiosperms

(unranked): Eudicots

(unranked): Coreo-eudicots

Order: unplaced

Family: Dilleniaceae

Genus: Dillenia

Species: D.indica

Description : Middle sized to evergreen tree. Leaves fascicled at the end of the branches, oblong lanceolate, acuminate, sharply serrate, lateral nerves. Flowers large, white, fragrant. Petals obovate, delicate. Fruit always green, hard and enclosed by five thickened sepals.

- *Elettaria cardamomum* (*Green or true cardamom*)

Kingdom: Plantae

(unranked): Angiosperms

(unranked): Monocots

(unranked): Commelinids

Order: Zingiberales

Family: Zingiberaceae

Genus: Elettaria

Species: E.cardamomum

Description : Elettaria cardamomum, commonly known as green or true cardamom, is a herbaceous perennial plant in the ginger family, native to southern India. It is the most common of the species whose seeds are used as a spice called cardamom. It is cultivated widely in tropical regions and reportedly naturalized in Réunion, Indochina and Costa Rica.

- *Grevillea robusta* (*Silky oak*)

Kingdom: Plantae

(unranked): Angiosperms

(unranked): Eudicots

Order: Proteales

Family: Proteaceae

Genus: Grevillea

Species: G.robusta

Description : It is a fast growing evergreen tree with dark green delicately dented bipinnatifid leaves reminiscent of a fern frond. These leaves are greyish white or rusty undersides. Its flowers are golden-orange bottlebrush-like blooms. The fruits are with one or two flat, winged seeds.

- *Litchi chinensis* (*Lichi*)

Kingdom: Plantae

(unranked): Angiosperms

(unranked): Eudicots

(unranked): Rosids

Order: Sapindales

Family: Sapindaceae

Subfamily: Sapindoideae

Genus: Litchi

Species: L.chinensis

Description : It is a medium-sized evergreen tree, with alternate pinnate leaves, with 2-8 lateral leaflets; the terminal leaflet is absent. The newly emerging young leaves are a bright coppery red at first, before turning green as they expand to full size. The flowers are small, greenish-white or yellowish- white, produced in panicles. The fruit is a drupe. The edible flesh consists of a highly developed aril enveloping the seed. The centre contains a single glossy brown nut-like seed.

- *Mangifera indica* (Mango)

Kingdom:Plantae

(unranked):Angiosperms

(unranked):Eudicots

(unranked):Rosids

Order:Sapindales

Family:Anacardiaceae

Genus:Mangifera

Species:M.indica

Description : *Mangifera indica*, commonly known as mango, is a species of flowering plant in the sumac and poison ivy family Anacardiaceae. It is found in the wild in Bangladesh, India and Pakistan where it is indigenous and cultivated varieties have been introduced to other warm regions of the world. It is a large fruit-tree. The species was described for science by Linnaeus in 1753. It is believed to have been tasted by Alexander (4th century BCE) and Chinese pilgrim Hieun Tsang (7th century CE).

- *Murraya exotica* (Kamini)

Kingdom:Plantae

(unranked):Angiosperms

(unranked):Eudicots

(unranked):Rosids

Order:Sapindales

Family:Rutaceae

Genus:Murraya

Species:M.paniculata

Description : *Murraya paniculata* is a tropical, evergreen plant native to southern China, Taiwan, the Indian sub-continent (i.e. Bangladesh, India, Nepal, north-eastern Pakistan and Sri Lanka), south-eastern Asia (i.e. Cambodia, Laos, Myanmar, Thailand, Vietnam, Indonesia,

Malaysia and the Philippines) and northern Australia (i.e. the northern parts of the Northern Territory, far northern Queensland and northern Western Australia), bearing small, white, scented flowers, which is grown as an ornamental tree or hedge. *Murraya* is closely related to *Citrus*, and bears small orange to red fruit resembling kumquats, though some cultivars do not set fruit.

- *Musa accuminata* (*Banana*)

Kingdom:Plantae

Clade:Angiosperms

Clade:Monocots

Clade:Commelinids

Order:Zingiberales

Family:Musaceae

Genus:Musa

Species:M.acuminata

Description : *Musa acuminata* is a species of banana native to Southeast Asia. *Musa acuminata* is an evergreen perennial, not a tree. The trunk (known as the pseudostem) is made of tightly packed layers of leaf sheaths emerging from completely or partially buried corms.

- *Psidium guajava* (*Guava*)

Kingdom: Plantae

(unranked): Angiosperms

(unranked): Eudicots

(unranked):Rosids

Order: Myrtales

Family: Myrtaceae

Subfamily:Myrtoideae

Tribe: Myrteae

Genus:Psidium

Species:P.guajava

Description : A small tree with spreading branches, the guava is easy to recognize because of its smooth, thin, copper-colour bark that flakes off. The leaves, aromatic when crushed, are evergreen, opposite, short- petiole, oval or oblong-elliptic, leathery, with conspicuous parallel veins. Faintly fragrant, the white flowers, borne singly or in small clusters in the leaf axils. The fruit, exuding a strong, sweet, musky odour when ripe, may be round, ovoid, or pear-shaped, and thin, light-yellow skin, frequently blushed with pink.

- *Santalum album (White sandalwood)*

Kingdom:Plantae

(unranked):Angiosperms

(unranked):Eudicots

(unranked) :Core eudicots

Order:Santalales

Family:Santalaceae

Genus:Santalum

Species:S.album

Description :Santalum album or Indian sandalwood is a small tropical tree, and is the most commonly known source of sandalwood. Certain cultures place great significance on its fragrant and medicinal qualities. The high value of the species has caused its past exploitation, to the point where the wild population is vulnerable to extinction. Indian sandalwood still commands high prices for its essential oil, but due to lack of sizable trees it is no longer used for fine woodworking as before. The plant is widely cultivated and long lived, although harvest is viable after 40 years.

- *Zea mays (Maize)*

Kingdom:Plantae

(unranked):Angiosperms

(unranked):Monocots

(unranked):Commelinids

Order:Poales

Family:Poaceae

Subfamily:Panicoideae

Tribe:Andropogoneae

Genus:Zea

Species:Z.mays

Subspecies:Z.mays

Description :Maize commonly known as corn, is a large grain plant first domesticated by indigenous peoples in Mexico about 10,000 years ago. The six major types of corn are dent corn, flint corn, pod corn, popcorn, flour corn, and sweet corn. The leafy stalk of the plant produces separate pollen and ovuliferous inflorescences or ears, which are fruits, yielding kernels (often erroneously called seeds). Maize kernels are often used in cooking as a starch.

3. MODEL PLANTS

- *Arabidopsis thaliana* (*Mouseear cress*)

Kingdom:Plantae

(unranked):Angiosperms

(unranked):Eudicots

(unranked):Rosids

Order:Brassicales

Family:Brassicaceae

Genus:Arabidopsis

Species:A.thaliana

Description :Arabidopsis thaliana (mouse-ear cress or arabidopsis) is a small flowering plant native to Eurasia. A. thaliana is edible by humans and, as with other mustard greens, is used in salads or sautéed, like many species in the Brassicacea. A. thaliana is considered a weed; it is

found by roadsides and in disturbed lands. A winter annual with a relatively short life cycle, Arabidopsis is a popular model organism in plant biology and genetics. For a complex multicellular eukaryote, Arabidopsis thaliana has a relatively small genome of approximately 135 megabase pairs (Mbp). It was long thought to have the smallest genome of all flowering plants, but the smallest flowering plants' genomes are now considered to belong to plants in the genus Genlisea, order Lamiales, with Genlisea tuberosa, a carnivorous plant, showing a genome size of approximately 61 Mbp. Arabidopsis thaliana was the first plant to have its genome sequenced, and is a popular tool for understanding the molecular biology of many plant traits, including flower development and light sensing. In 2016, it was named as one of the organisms that would be involved in CRISPR-based genetic engineering.

- *Brachypodium distachyon* (Purple false brome)

Kingdom:Plantae

(unranked):Angiosperms

(unranked):Monocots

(unranked):Commelinids

Order:Poales

Family:Poaceae

Subfamily:Pooideae

Genus:Brachypodium

Species:B.distachyon

Description : Brachypodium distachyon, commonly called purple false brome or stiff brome, is a grass species native to southern Europe, northern Africa and southwestern Asia east to India. It has many qualities that make it an excellent model organism for functional genomics research in temperate grasses, cereals, and dedicated biofuel crops such as switchgrass. Brachypodium is emerging as a powerful model with a growing research community. The goal of the International Brachypodium Initiative (IBI) is to promote the development of B. distachyon as a model system and will develop and distribute genomic, genetic, and bioinformatics resources such as reference genotypes, BAC libraries, genetic markers, mapping populations, and a genome sequence database. Recently, efficient Agrobacterium-mediated transformation systems have been developed for a range of Brachypodium

genotypes, enabling the development of T-DNA mutant collections. The characterization and distribution of T-DNA insertion lines has been initiated to facilitate the understanding of gene function in grasses. By now, *Brachypodium distachyon* has established itself as an important tool for comparative genomics. It is now emerging as a model for crop plant disease, facilitating the model-to-crop transfer of knowledge on disease resistance. *Brachypodium distachyon* is also becoming a useful model system for studies of evolutionary developmental biology, in particular to contrast molecular genetic mechanisms with dicotyledon model systems, notably *Arabidopsis thaliana*.

CHAPTER 5

DATABASE DEVELOPMENT

In this section, we have described the detailed features of three different databases along with the description of tables.

5.1 Databases Description

In this project, we have prepared three databases according to the types of the plants :

1. Economical Plants Database
2. Medicinal Plants Database
3. Model Plants Database

Databases have been created separately for three major categories as we plan to incorporate more data for all three categories and it will be easy to update then through this agile approach.

5.2 Tables Description

In each database we have considered maximum possible useful data from different resources and framed the data in five different tables as follows :

1. GENE TABLE

Column Name	Data Type	Size
Gene Name (PK)	Varchar	100
Primary Gene Name	Varchar	100
Organism Name	Varchar	100
Gene encoded by	Varchar	100
Protein_ID	Varchar	30
Protein_length (FK)	Int	30
Protein_sequence	Varchar	1100
EC_Number	Varchar	30

Table 3. Gene Table

2. PROTEIN TABLE

Column Name	Data Type	Size
Protein_acc_id (PK)	Varchar	50
Organism_Name	Varchar	250
Protein_name	Varchar	250
Protein_length	Varchar	50
PIRsf_id	Varchar	250
UniRef_id	Varchar	250

Table 4. Protein Table

3. TAXONOMY TABLE

Column Name	Data Type	Size
Taxonomy_id (PK)	Varchar	30
Taxonomy_title	Varchar	50
Organism_name	Varchar	100
Organism_id	Varchar	30
Protein_name	Varchar	300
Proteome	Varchar	100
Taxonomy Lineage	Varchar	500

Table 5. Taxonomy Table

4. MISCELLANEOUS FEATURES

Column Name	Data Type	Size
Misc_feature_id (PK)	Varchar	50
Organism_name	Varchar	250
Ec_number	Varchar	50
Catalytic_activity	Varchar	250
Cofactor	Varchar	250
Kinetics	Varchar	250
Function	Varchar	250
Pathway	Varchar	250
Active Site	Varchar	250

Table 6. Miscellaneous Features Table

5. GENE ONTOLOGY

Column Name	Data Type	Size
Gene_ontology_id (PK)	Varchar	50
Organism_name	Varchar	250
Gene_ontology_title	Varchar	250
Biological_process	Varchar	250
Molecular_function	Varchar	250
Cellular_component	Varchar	250
Gene_ontology_ids	Varchar	250

Table 7. Gene Ontology Table

5.3 Database Structure

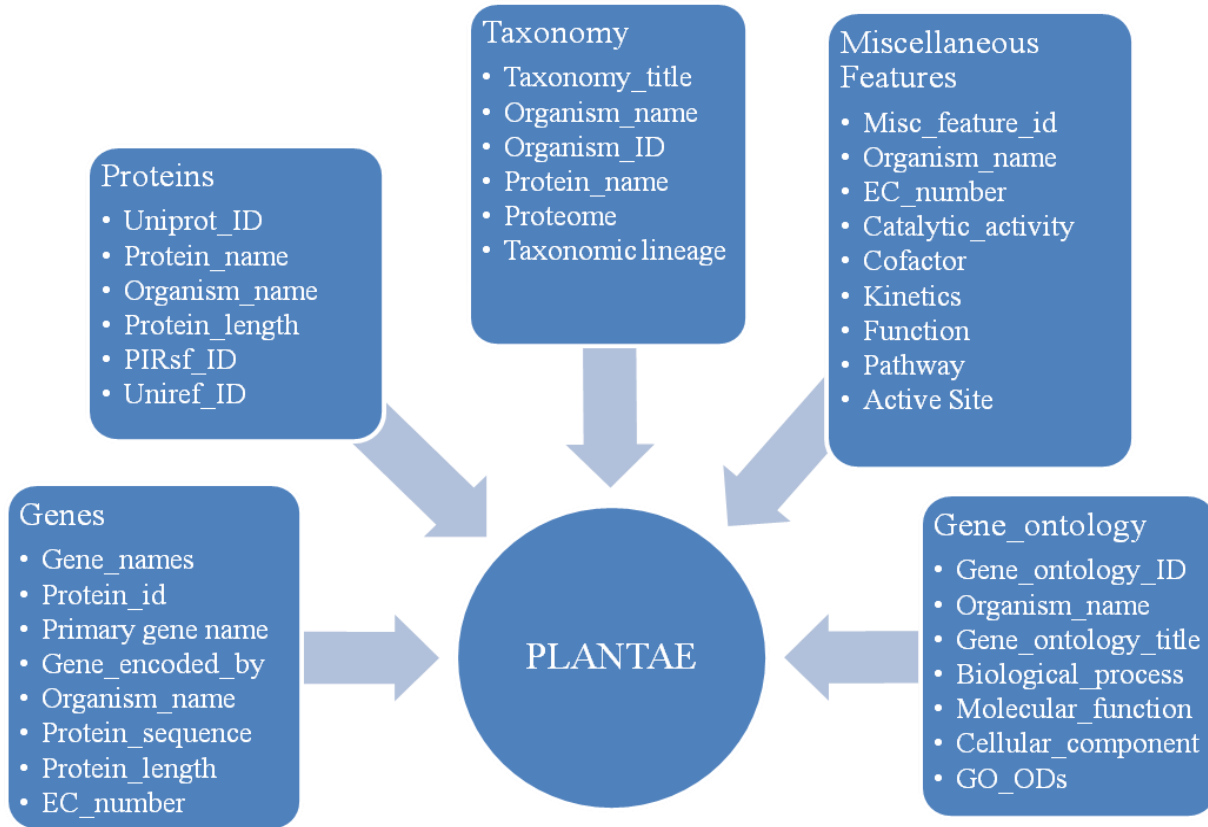


Figure 2. Database Structure

CHAPTER 6

DEVELOPMENT AND DESIGN

6.1 System Requirements

PLANTEA DATABASE Website is made using WampServer2.5 and my SQL, are used to store data. Much of the procedural calculations are done using PHP/SQL procedures at the server side.

Server Machine : Wamp Server 2.5

Client machine : PC with Server Connection

Software : Code editor, any browser(Google Chrome recommended), MS Office.

6.2 Languages Used

A front-end developer is a programmer that codes the front end of a website. Typically, the front-end developer's job is to convert website design files into raw HTML, JavaScript (JS) and/or CSS code. This includes the basic website design/layout, images, content, buttons, navigation and internal links. The end result is code that serves as the website's front-end structure, which is used by a back-end developer to add business logics and connect databases and processes, among other processes.

A front-end developer is responsible for ensuring that a website's visual front end is free of errors and looks exactly as designed. A front-end developer also ensures that a website has the same visibility across different computing and mobile Web browsers.

Similarly, in software applications, a front-end web developer creates the graphical user interface (GUI) that enables access to the software's back-end features and capabilities.

HTML

Hypertext Markup Language, commonly referred to as HTML, is the standard markup language used to create web pages. HTML is a cornerstone technology used to create web pages, as well as to create user interfaces for mobile and web applications. Web browsers can read HTML files and render them into visible or audible web pages. HTML describes the structure of a website semantically and, before the advent of Cascading Style Sheets (CSS), included cues for the presentation or appearance of the document (web page), making it a markup language, rather than a programming language. It is written in the form of HTML elements consisting of tags enclosed in angle brackets (like <html>). Web browsers can read HTML files and render them into visible or audible web pages. Browsers do not display the HTML tags and scripts, but use them to interpret the content of the page. HTML describes the structure of a website semantically along with cues for presentation, making it a markup language, rather than a programming language.

HTML elements are the building blocks of all websites. HTML allows images and objects to be embedded and can be used to create interactive forms. It provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists, links, quotes and other items. It can embed scripts written in languages such as JavaScript which affect the behavior of HTML web pages.

HTML can embed scripts written in languages such as JavaScript which affect the behavior of HTML web pages. HTML markup can also refer the browser to Cascading Style Sheets (CSS) to define the look and layout of text and other material.

CSS

Cascading Style Sheets (CSS) is a style sheet language used for describing the look and formatting of a document written in a markup language. Although most often used to change the style of web pages and user interfaces written in HTML and XHTML, the language can be applied to any kind of XML document, including plain XML, SVG and XUL. Along with HTML and JavaScript, CSS is a cornerstone technology used by most websites to create visually engaging webpages, user interfaces for web applications, and user interfaces for many mobile applications.

CSS is designed primarily to enable the separation of document content from document presentation, including elements such as the layout, colors, and fonts. This separation can improve content accessibility, provide more flexibility and control in the specification of presentation characteristics, enable multiple HTML pages to share formatting by specifying the relevant CSS in a separate CSS file, and reduce complexity and repetition in the structural content, such as semantically insignificant tables that were widely used to format pages before consistent CSS rendering was available in all major browsers. CSS makes it possible to separate presentation instructions from the HTML content in a separate file or style section of the HTML file. For each matching HTML element, it provides a list of formatting instructions. Changes to the graphic design of a document (or hundreds of documents) can be applied quickly and easily, by editing a few lines in the CSS file they use, rather than by changing mark up in the documents.

The CSS specification describes a priority scheme to determine which style rules apply if more than one rule matches against a particular element. In this so-called cascade, priorities (or weights) are calculated and assigned to rules, so that the results are predictable.

JavaScript

JavaScript (often shortened to JS) is a lightweight, interpreted, object-oriented language with first-class functions, most known as the scripting language for Web pages, but used in many non-browser environments as well such as node.js or Apache Couch DB. Alongside HTML and CSS, it is one of the three core technologies of World Wide Web content production; the majority of websites employ it and it is supported by all modern Web browsers without plug-

ins. It is a prototype-based, multi-paradigm scripting language that is dynamic, and supports object-oriented, imperative, and functional programming styles.

It has an API for working with text, arrays, dates and regular expressions, but does not include any I/O, such as networking, storage, or graphics facilities, relying for these upon the host environment in which it is embedded.

JavaScript is also used in environments that are not Web-based, such as PDF documents, site-specific browsers, and desktop widgets. Newer and faster JavaScript virtual machines (VMs) and platforms built upon them have also increased the popularity of JavaScript for server-side Web applications. On the client side, JavaScript has been traditionally implemented as an interpreted language, but more recent browsers perform just-in-time compilation. It is also used in game development, the creation of desktop and mobile applications, and server-side network programming with runtime environments.

PHP

PHP is a server-side scripting language designed for web development but also used as a general-purpose programming language. Originally created by Rasmus Lerdorf in 1994, the PHP reference implementation is now produced by The PHP Group. PHP originally stood for Personal Home Page, but it now stands for the recursive backronym PHP: Hypertext Pre processor.

PHP code may be embedded into HTML code, or it can be used in combination with various web template systems, web content management system and web frameworks. PHP code is usually processed by a PHP interpreter implemented as a module in the web server or as a Common Gateway Interface (CGI) executable. The web server combines the results of the interpreted and executed PHP code, which may be any type of data, including images, with the generated web page. PHP code may also be executed with a command-line interface (CLI) and can be used to implement standalone graphical applications.

6.3 GUI Development:

The development of Graphical User Interface requires coding in HTML, CSS and connecting the same via certain codes in PHP. We used a template from bootstrap.com. The website menu includes the following options.

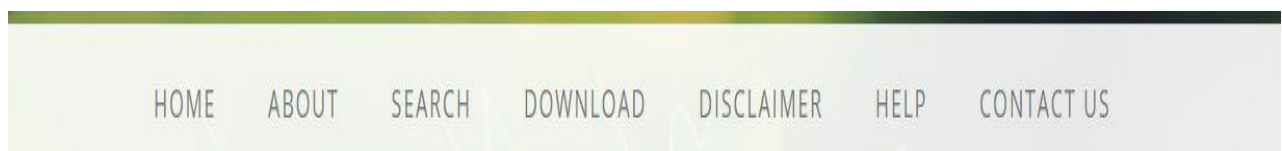


Figure 3. Options in GUI

1. HOME

The home page of this website has slider containing the 5 images of the plants that we have collected. The PLANTAE DATABASE is a database for diverse plant species. It contains standardised information about medicinal plants, agricultural plants, Ornamental plants and transgenic plants. It includes genomic, protein, taxonomy information. There are many similar websites but this website different from others as we are providing all the respective information on the same platform. The information provided is useful for research purposes, study purposes etc. We are trying to add more plants and with their useful information.

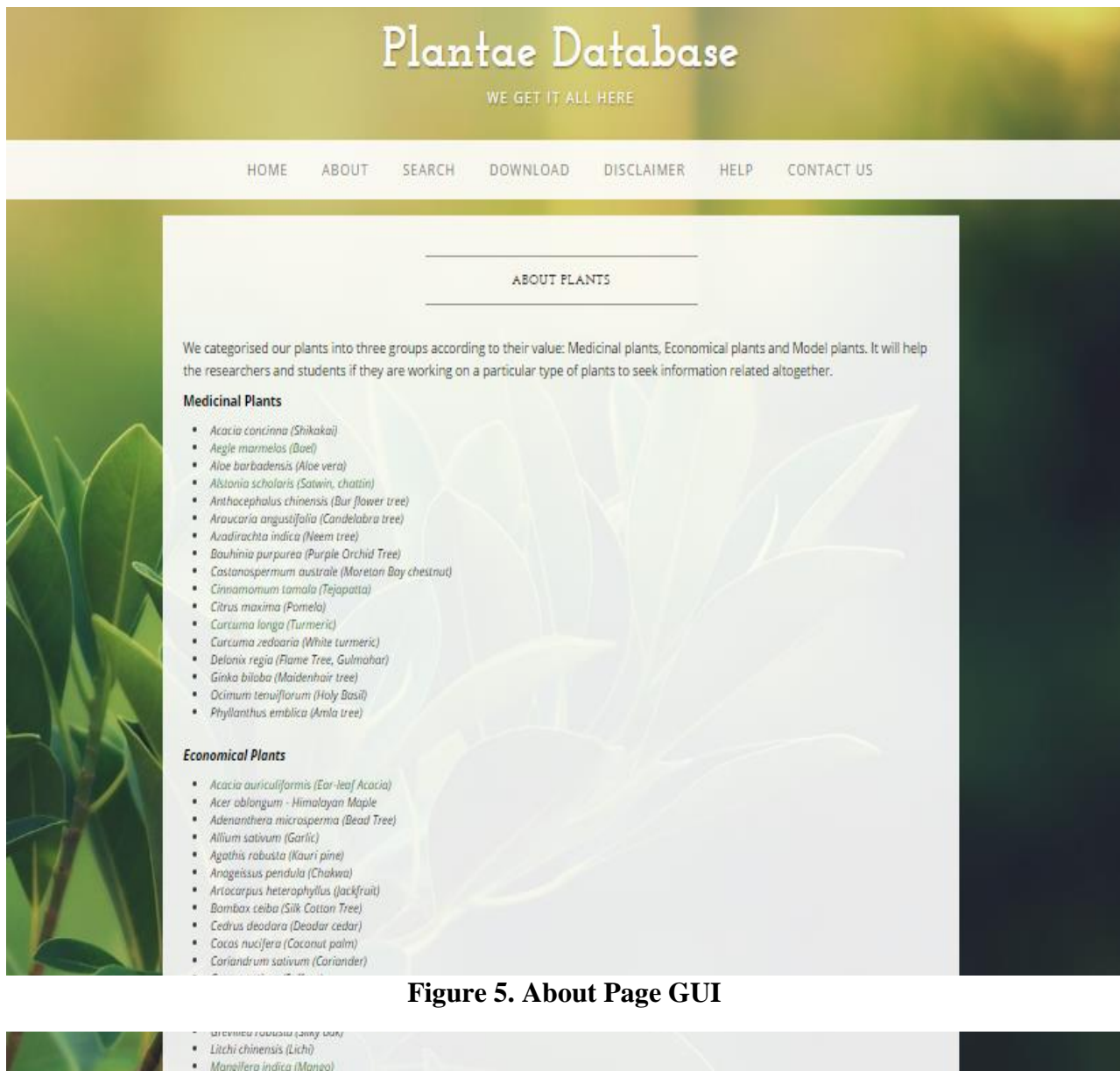


Figure 4. Home Page GUI

2. ABOUT

We categorized our plants into three groups according to their value: Medicinal plants, Economical plants and Model plants. It will help the researchers and students if they are

working on a particular type of plants to seek information related altogether. In this page basic information plant is given and have a NCBI taxonomy link.



3. SEARCH

There are two types of searches i.e. Basic Search and Advanced Search. In Basic Search, select the type of the plant, different kinds of data of different plants. In Advanced Search we can search with various keywords i.e. search acc. to gene name, gene id etc.

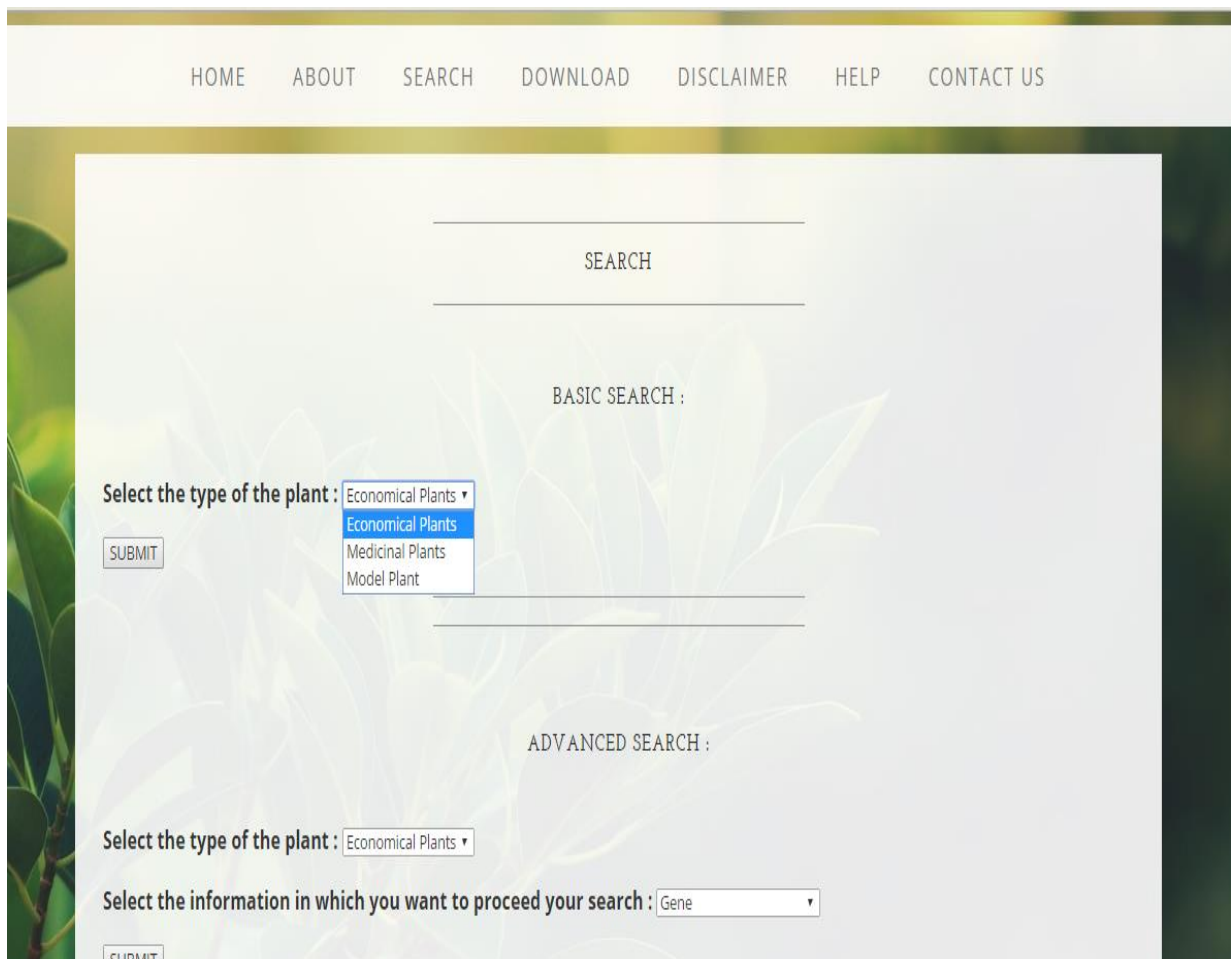


Figure 6. Search Option GUI

4. DOWNLOAD

The whole database is downloadable. According to Plant types and and the data.

5. DISCLAIMER

6. HELP

7. CONTACT US

This page contains the address and phone no and the map of the address. Also feedback form is there to contact.

6.4 Back End Web Designing

Back-end developers primarily develop and maintain the core functional logic and operations of a software or information system. The key job role of a back-end developer is to ensure that the data or services requested by the front-end system or software are delivered through programmatic means. Back-end developers also create and maintain the entire back-end, which consists of the core application logic, databases, data and application integration, API and other back-end processes. Moreover, a back-end developer also performs the testing and debugging of any back-end application or system.

Back End (Database)

A database server is the key to solving the problems of information management. In general, a server must reliably manage a large amount of data in a multiuser environment so that many users can concurrently access the same data. All this must be accomplished while delivering high performance. A database server must also prevent unauthorized access and provide efficient solutions for failure recovery.

The Wamp server provides efficient and effective solutions with the following features:

Client-Server (distributed processing) environments

To take full advantage of a given computer system or network, Oracle allows processing to be split between the database server and the client application programs. The computer running the database management system handles all of the database server responsibilities while the workstation running the database application concentrates on the interpretation and display of data.

- **Large databases and space management**
Oracle supports the largest of databases, potentially terabytes in size. To make efficient use of expensive hardware devices, it allows full control of space usage.
- **Many concurrent database users**
Oracle supports large number of concurrent users executing a variety of database applications operating on the same data. It minimizes data contention and guarantees data concurrency.
- **High transaction processing performance**
Oracle maintains the preceding features with a high degree of overall system performance. Database users do not suffer from slow processing performance.
- **High availability**
At some sites, oracle works 24 hours a day with no down time to limit database throughput. Normal system operations such as database backup and partial computer system failures do not interrupt database use.
- **Controlled availability**
Oracle can selectively control the availability of data, at the database level and sub-database level. For example, an administrator can disallow use of a specific application so that the application's data can be reloaded, without affecting other applications.
- **Openness, industry standards**
Oracle adheres to industry accepted standards for the data access language, operating systems, user interfaces, and network communication protocols. It is an "open" system that protects a customer's investment.
- Oracle also supports the Simple Network Management Protocol (SNMP) standard for system management. This protocol allows administrators to manage heterogeneous systems with a single administration interface.

- **Manageable security**

To protect against unauthorized database access and use, Oracle provides fail-safe security features to limit and monitor data access. These features make it easy to manage even the most complex design for data access.

- **Database enforced integrity**

Oracle ensures data integrity, “business rules” that dictate the standards for acceptable data. As a result, the costs of coding and managing checks in many database applications are eliminated.

- **Distributed systems**

For networked, distributed environments, Oracle combines the data physically located on different computers into one logical database that can be accessed by all network users. Distributed systems have the same degree of user transparency and data consistency as non-distributed systems, yet receive the advantages of local database managements.

Oracle also offers the heterogeneous system option that allows users to access data on some non-Oracle database transparently.

- **Portability**

Oracle software is ported to work under different operating systems. Applications developed for Oracle can be ported to any operating system with little or no modification.

- **Compatibility**

Oracle software is compatible with industry standards, including most industry standard operating system. Applications developed for Oracle can be used on virtually any system with little or no modification.

- **Connectivity**

Oracle software allows different types of computers and operating systems to share information across networks.

- **Replicated environments**

Oracle software lets you replicate groups of tables and their supporting objects to multiple sites. Oracle supports replication of both data and schema level changes to these sites. Oracle’s flexible replication technology supports basic primary site replication as well as advanced dynamic and shared–ownership models.

Features of SQL

Views : A view is a tailored presentation of the data contained in one or more tables (or other views), and takes the output of a query and treats it as a table. You can think of a view as a “stored query” or a “virtual table”. You can use views in most places where a table can be used.

Sequences : The CACHE operation pre-allocates a set of sequence numbers and keeps them in memory so that sequence numbers can be accessed faster. When the last of the sequence numbers in the cache has been used, Oracle reads another set of numbers into the cache.

Oracle might skip sequence numbers if you choose to cache a set of sequence numbers. For example, when an instance abnormally shuts down (for example, when an instance failure occurs or a SHUTDOWN ABORT statement is issued), sequence numbers that have been cached but not used are lost. Also, sequence numbers that have been used but not saved are lost as well.

Packages : Packages are PL/SQL constructs that allow related objects to be stored together. A package has two separate parts- the specification and the body. Each of them is stored separately in the data dictionary. Besides allowing related objects to be grouped together, packages are useful because they are less restrictive with respect to dependencies. They also have some performance advantages. A package is essentially a named declarative section. Anything that can go in the declarative part of a block can go in a package. This includes procedures, functions, cursors, types, and variables. One advantage of putting these objects into a package is the ability to reference them from other PL/SQL blocks, so packages also provide global variables to PL/SQL.

Cursors : To execute a multi-row query, Oracle opens an unnamed work area that stores processing information. A cursor lets you name the work area, access the information, and process the rows individually. A cursor is a handle or a pointer to the context area. Through the cursor a PL/SQL program can control the context area and what happens to it as it is processed.

Procedures and functions : procedure or a function is a PL/SQL program unit that has a name, can take parameters, and return values, is stored in the data dictionary, can be invoked by many users.

CHAPTER 7

DATABASE CONNECTIVITY AND DATA RETRIEVAL

This section involves two final steps: Database connectivity and Data Retrieval.

7.1 Database Connectivity

We have compressive collection of different types of data from different resources. In the multi dimensional model, data are organised into multiple dimensions and each dimension contains multiple levels of abstraction defined by the concept hierarchies. The data is stored in 3 different databases : Medicinal_plants, economical_plants and model_plants. And each database is containing five different tables as follows :

1. Economical Plant Database

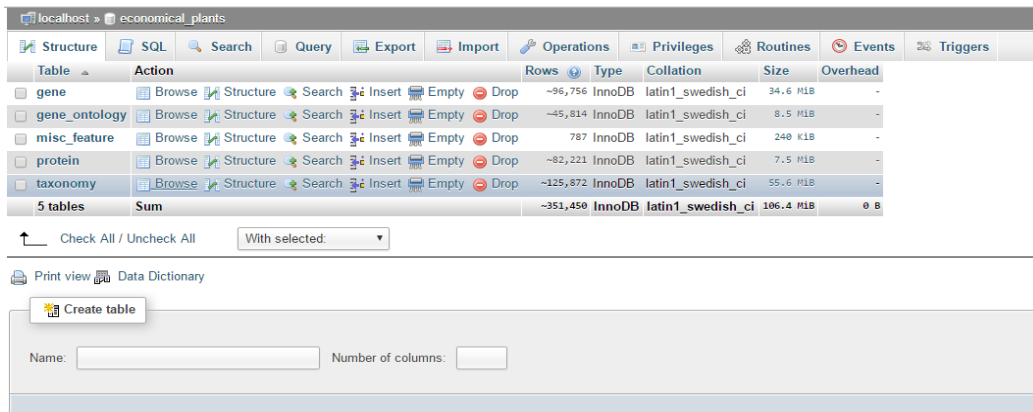


Table	Action	Rows	Type	Collation	Size	Overhead
gene	Browse Structure Search Insert Empty Drop	~96,756	InnoDB	latin1_swedish_ci	34.6 MiB	-
gene_ontology	Browse Structure Search Insert Empty Drop	~45,814	InnoDB	latin1_swedish_ci	8.5 MiB	-
misc_feature	Browse Structure Search Insert Empty Drop	787	InnoDB	latin1_swedish_ci	248 KiB	-
protein	Browse Structure Search Insert Empty Drop	~82,221	InnoDB	latin1_swedish_ci	7.5 MiB	-
taxonomy	Browse Structure Search Insert Empty Drop	~125,872	InnoDB	latin1_swedish_ci	55.6 MiB	-
5 tables	Sum	~351,450	InnoDB	latin1_swedish_ci	106.4 MiB	0 B

Check All / Uncheck All With selected: ▾

Print view Data Dictionary

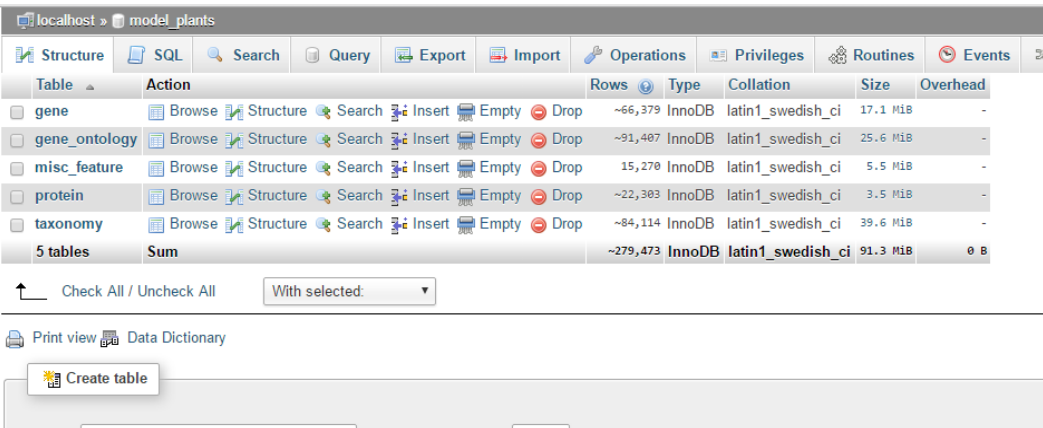
Create table

Name: Number of columns:

Figure 7. Economical Plant Database

2. Medicinal Plants Database Structure

3. Model Plants Database Structure



The screenshot shows the MySQL Workbench interface for the 'model_plants' database. The 'Structure' tab is active, displaying a table list with columns: Table, Action, Rows, Type, Collation, Size, and Overhead. The tables listed are: gene (~66,379 rows, 17.1 MiB), gene_ontology (~91,487 rows, 25.6 MiB), misc_feature (15,278 rows, 5.5 MiB), protein (~22,383 rows, 3.5 MiB), and taxonomy (~84,114 rows, 39.6 MiB). A summary row shows 5 tables with a total of ~279,473 rows and 91.3 MiB size. Below the table list, there are options for 'Check All / Uncheck All' and 'With selected:'. At the bottom, there are buttons for 'Print view', 'Data Dictionary', and 'Create table'.

Table	Action	Rows	Type	Collation	Size	Overhead
gene	Browse Structure Search Insert Empty Drop	~66,379	InnoDB	latin1_swedish_ci	17.1 MiB	-
gene_ontology	Browse Structure Search Insert Empty Drop	~91,487	InnoDB	latin1_swedish_ci	25.6 MiB	-
misc_feature	Browse Structure Search Insert Empty Drop	15,278	InnoDB	latin1_swedish_ci	5.5 MiB	-
protein	Browse Structure Search Insert Empty Drop	~22,383	InnoDB	latin1_swedish_ci	3.5 MiB	-
taxonomy	Browse Structure Search Insert Empty Drop	~84,114	InnoDB	latin1_swedish_ci	39.6 MiB	-
5 tables	Sum	~279,473	InnoDB	latin1_swedish_ci	91.3 MiB	0 B

Figure 8. Model Plants DB Structure

7.2 Data Retrieval

PHP code used to download the data as follows :

```
1 <?php
2     $username = "root";
3     $hostname = "localhost";
4     $dbname = "medicinal_plants";
5
6 /* Attempt MySQL server connection. Assuming you are running MySQL
7 server with default setting (user 'root' with no password) */
8 $link = mysqli_connect($hostname,$username,"",$dbname);
9
10 // Check connection
11 if($link === false)
12 {
13     die("ERROR: Could not connect. " . mysqli_connect_error());
14 }
15 else
16 {
17 }
18 ?>
```

Figure 10. PHP Code For Download-1

```

96 <?php
97
98     $username = "root";
99     $hostname = "localhost";
100     $dbname = "economical_plants";
101
102     /* Attempt MySQL server connection. Assuming you are running MySQL
103     server with default setting (user 'root' with no password) */
104     $link = mysqli_connect($hostname,$username,"",$dbname);
105
106     // Check connection
107     if($link === false)
108     {
109         die("ERROR: Could not connect. " . mysqli_connect_error());
110     }
111     else
112     {
113         mysqli_select_db('economical_plants');
114         // output headers so that the file is downloaded rather than displayed
115         header('Content-Type: text/csv; charset=utf-8');
116         header('Content-Disposition: attachment; filename=eco_data.csv');
117
118         // create a file pointer connected to the output stream
119         $output = fopen('php://output', 'w');
120
121         // output the column headings
122         fputs($output, array('go_id','organism_name','go_title', 'biological_process', 'molecular_function', 'cellular_component', 'go_ids'));
123
124         // fetch the data
125         $rows = mysqli_query('SELECT go_id,organism_name,go_title,biological_process,molecular_function,cellular_component,go_ids FROM gene_ontology');
126
127         // loop over the rows, outputting them
128         while ($row = mysqli_fetch_assoc($rows)) fputs($output, $row);
129     }
130 ?>

```

Figure 11. PHP Code For Download-2

```

112
113     $string1 = $_GET["species"];
114     $string2 = $_GET["category"];
115     echo " <h4>You have searched for the $string2 details of $string1. </h4>";
116
117     if($string2 == 'gene')
118     {
119
120         // Attempt select query execution :: gene
121         $sql = "SELECT * FROM gene WHERE organism_name='$string1'";
122         if($result = mysqli_query($link, $sql))
123         {
124             if(mysqli_num_rows($result) > 0)
125             {
126                 echo "<table border=2>";
127                 echo "<tr>";
128                 echo "<th>GENE NAME</th>";
129                 echo "<th>PRIMARY GENE NAME</th>";

```

Figure 12. PHP Code For Download-3

```
551
552 $mf_keyword = $_GET["mf_id"];
553 // Attempt select query execution :: misc_feature
554 $sql = "SELECT * FROM misc_feature WHERE misc_feature_id='$mf_keyword'";
555 if($result = mysqli_query($link, $sql))
556 {
557     if(mysqli_num_rows($result) > 0)
558     {
559         echo "<table border=2>";
560         echo "<tr>";
561         echo "<th>MISC FEATURE ID</th>";
```

Figure 13. PHP Code For Download-4

Chapter 8

Results And Conclusion

This interface provides users with the flexibility to view data from different perspectives. Now next our step is to retrieve the data.

In this website, we have two options to retrieve information on the basis of the requirement.

1. Basic Search
2. Advanced Search

There are manual retrieval techniques and data mining techniques are used. The techniques use the computational processes of discovering patterns in large data sets ("big data") involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems.

8.1 Basic Search

In this basic search way, we have to choose the type of plant of interest, i.e. Economical plant, Medicinal plant or Model plant. After clicking on submit button, following page is opened. Now as the screen is displaying, we need to select the name of the plant of interest, and the type of information we need to retrieve (gene data, gene ontology data, protein data, taxonomy data or miscellaneous data).

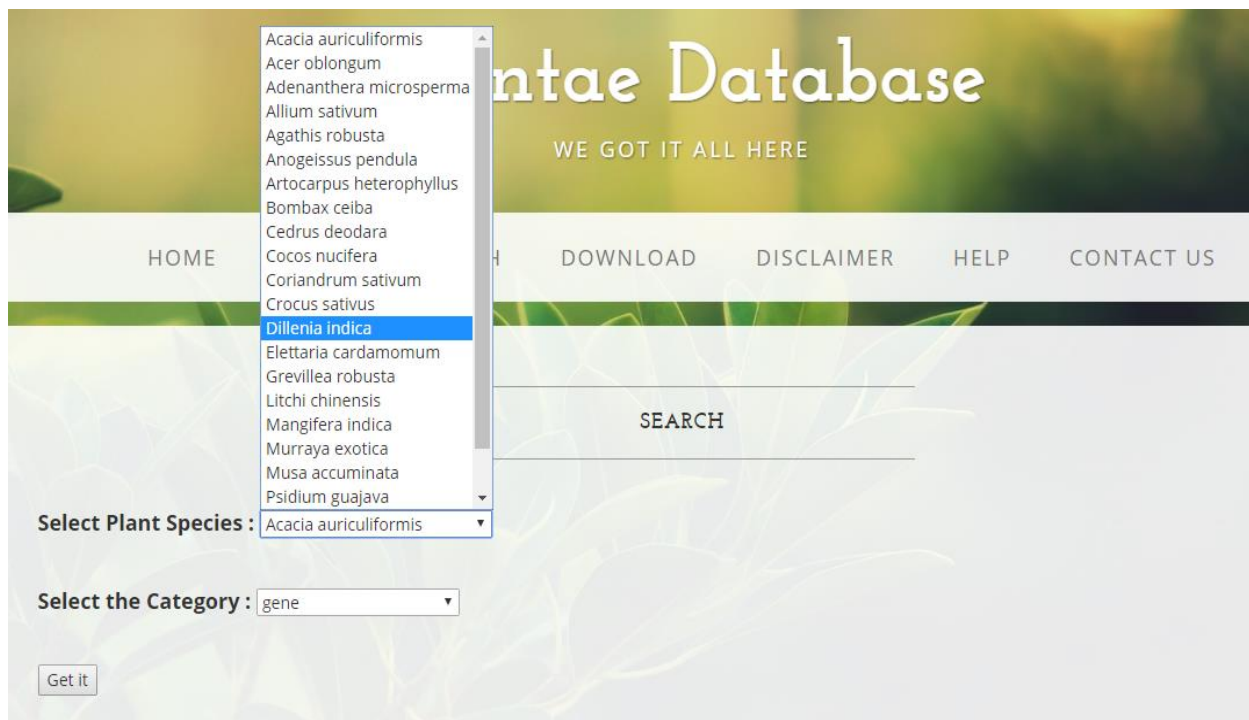


Figure 14. Specific Search

If we search for economical plant < *Cocos nucifera* < gene ontology data : The results are as follows :

GENE ONTOLOGY ENTRY	ORGANISM NAME	GENE ONTOLOGY TITLE	BIOLOGICAL PROCESSES	MOLECULAR FUNCTION	CELLULAR COMPONENT	GENE ONTOLOGY IDs
A0A068EW60	Cocos nucifera	ATP binding [GO:0005524]	Not Available	ATP binding [GO:0005524]	Not Available	GO:0005524
A0A088S6A3	Cocos nucifera	nucleic acid binding [GO:0003676]	Not Available	nucleic acid binding [GO:0003676]	Not Available	GO:0003676
A0A088S6Q5	Cocos nucifera	Not Available	Not Available	Not Available	Not Available	Not Available
A0A088S6T7	Cocos nucifera	RNA-directed RNA polymerase activity [GO:0003968]	Not Available	RNA-directed RNA polymerase activity [GO:0003968]	Not Available	GO:0003968
A0A088S6U2	Cocos nucifera	Not Available	Not Available	Not Available	Not Available	Not Available
A0A088S8C2	Cocos nucifera	RNA-directed RNA polymerase activity [GO:0003968]	Not Available	RNA-directed RNA polymerase activity [GO:0003968]	Not Available	GO:0003968
A0A088SQF2	Cocos nucifera	Not Available	Not Available	Not Available	Not Available	Not Available

Figure 15. Result For Specific Search

The first column is Gene Ontology Id : By clicking that user will be taken to online database where all the information regarding the ID is displayed.

8.2 Advanced Search

In Advanced Search option, the user can search using a keyword like gene_name, gene_id, protein_id etc. following is the glimpse for that :

First we need to select the type of the plant (i.e. Economical plant, Medicial plant or Model plant) and the type of the data (gene data, gene ontology data, protein data, taxonomy data or miscellaneous data) that the user is looking for.

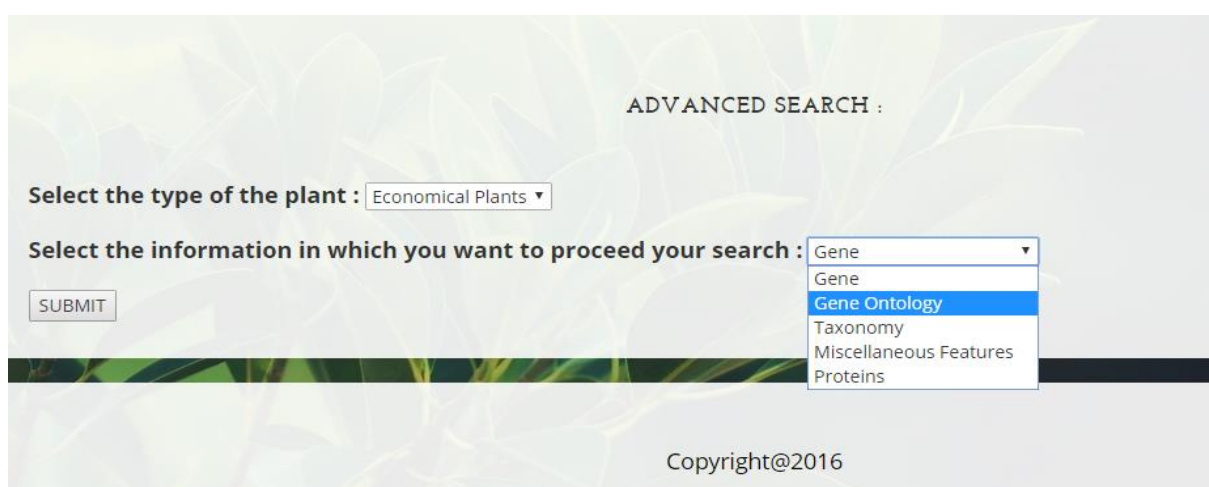


Figure 16. Advanced Search

After submitting the first form, the user will be asked for the keyword.

For Example : If user is looking for gene data by entering the gene ontology ID :



Figure 17. Advanced Search Example

Similarly, the user can search by many options : Miscellaneous feature ID, Protein ID, Taxonomy ID etc.

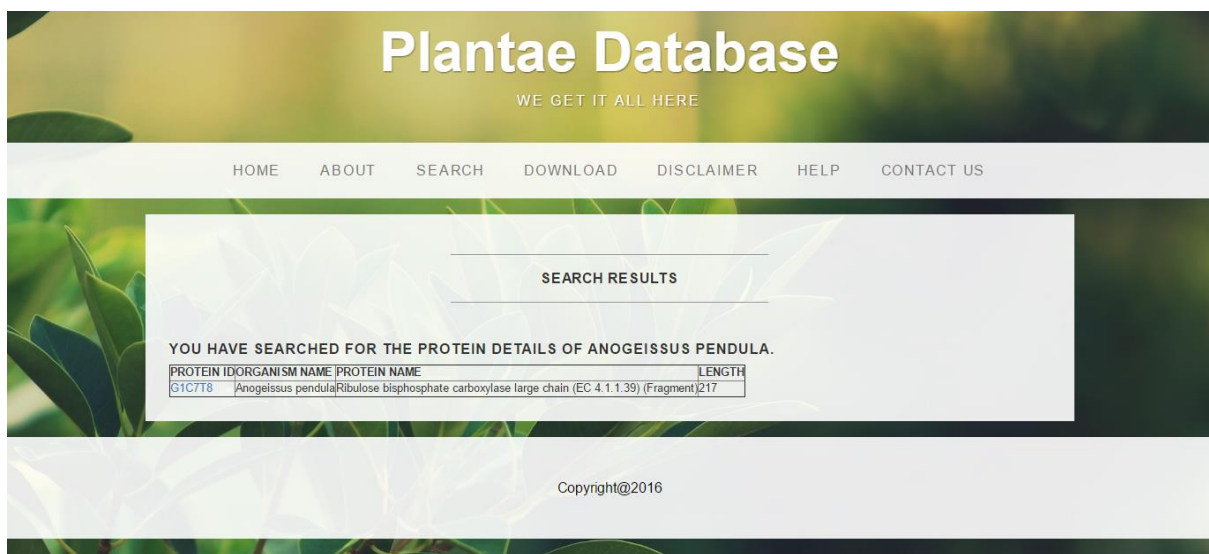
The outcomes look like :

If we search for the details of protein through Protein ID search option :

Plant Type : Economical_plants

Plant Name : *Anogeissus pendula*

Protein ID : G1C7T8



The screenshot shows the Plantae Database website. The header features the title "Plantae Database" and the tagline "WE GET IT ALL HERE". A navigation menu includes links for HOME, ABOUT, SEARCH, DOWNLOAD, DISCLAIMER, HELP, and CONTACT US. The main content area displays "SEARCH RESULTS" and a message: "YOU HAVE SEARCHED FOR THE PROTEIN DETAILS OF ANOGEISSUS PENDULA." Below this, a table lists search results:

PROTEIN ID	ORGANISM NAME	PROTEIN NAME	LENGTH
G1C7T8	Anogeissus pendula	Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (Fragment)	217

The footer of the page contains the text "Copyright@2016".

Figure 18. Search Results

REFERENCES

1. <http://www.uniprot.org>
2. Images in GUI: www.google.com
3. Code Reference: www.w3schools.com
4. Template : www.startbootstrap.com
5. www.rcsb.com
6. www.ebi.ac.uk
7. www.esemble.org
8. www.bgci.org
9. www.blast2go.com
10. www.genome.jp
11. <http://www.ncbi.nlm.nih.gov>
12. www.arabidopsis.org
13. <http://www.ncbi.nlm.nih.gov>
14. <http://plantreactome.gramene.org/>
15. <http://pir.georgetown.edu/>
16. <http://amigo.geneontology.org/amigo>
17. <http://www.expasy.org/>
18. www.brenda-enzymes.org
19. www.wwpdb.org
20. www.genome.jp
21. <http://www.wikipedia.org>
22. <http://www.genome.jp/kegg/pathway.html>
23. <http://bioseed.mcs.anl.gov>