**Gene Curation Tool (GCT)**

**Design Document**

08/24/2015

Contents

[1 Introduction 5](#_Toc428313797)

[1.1 Purpose 5](#_Toc428313798)

[1.2 Basic Genetic knowledge 5](#_Toc428313799)

[1.3 Ontology 7](#_Toc428313800)

[1.4 Annotation 9](#_Toc428313801)

[1.5 Planteome Project 9](#_Toc428313802)

[1.6 Gene Curation Tool (GCT) 10](#_Toc428313803)

[2 Related Ontology Databases 11](#_Toc428313804)

[2.1 Amigo and Amigo2 11](#_Toc428313805)

[2.1.1 Introduction 11](#_Toc428313806)

[2.1.2 Features 12](#_Toc428313807)

[2.1.3 GO Annotation File (GAF) Format 12](#_Toc428313808)

[2.2 Gramene 13](#_Toc428313809)

[2.3 AgriGO 13](#_Toc428313810)

[2.4 Conclusion 13](#_Toc428313811)

[3 Software Requirement Specification 13](#_Toc428313812)

[3.1 Product Perspective 13](#_Toc428313813)

[3.2 Scope 14](#_Toc428313814)

[3.3 Operating Environment 14](#_Toc428313815)

[3.4 Role Based Access Control (RBAC) 14](#_Toc428313816)

[3.5 Product Functions 14](#_Toc428313817)

[3.6 User management subsystem 16](#_Toc428313818)

[3.6.1 Use Case: Register 16](#_Toc428313819)

[3.6.2 Use Case: Login 18](#_Toc428313820)

[3.6.3 Use Case: Ban User 18](#_Toc428313821)

[3.6.4 Use Case: User role Hierarchy 19](#_Toc428313822)

[3.6.5 Use Case: Credit of the contribution 20](#_Toc428313823)

[3.6.6 Use Case: Edit Specialty 21](#_Toc428313824)

[3.7 Annotation management subsystem 21](#_Toc428313825)

[3.7.1 Use Case: browse/edit/add annotation 21](#_Toc428313826)

[3.7.2 Use Case: Save annotation draft 22](#_Toc428313827)

[3.7.3 Use Case: Save note 22](#_Toc428313828)

[3.7.4 Use Case: flag annotation 22](#_Toc428313829)

[3.7.5 Use Case: Comment on annotation 22](#_Toc428313830)

[3.8 Object management subsystem 23](#_Toc428313831)

[3.9 Publication management subsystem 24](#_Toc428313832)

[4 API design 25](#_Toc428313833)

[4.1 Object import API 25](#_Toc428313834)

[4.2 Annotation import API 25](#_Toc428313835)

[4.3 Annotation export API 25](#_Toc428313836)

[4.4 Utilize API to get Ontology information 25](#_Toc428313837)

[5 User Interface Design 26](#_Toc428313838)

[6 Database Design 26](#_Toc428313839)

[6.1 ER diagram 26](#_Toc428313840)

[6.2 Tables design 26](#_Toc428313841)

[6.2.1 Table: Users 26](#_Toc428313842)

[6.2.2 Table: User\_banned 27](#_Toc428313843)

[6.2.3 Table: Specialty 27](#_Toc428313844)

[6.2.4 Table: User\_Specialty 27](#_Toc428313845)

[6.2.5 Table: Object 28](#_Toc428313846)

[6.2.6 Table: Synonyms 28](#_Toc428313847)

[6.2.7 Table: Species 28](#_Toc428313848)

[6.2.8 Table: Gene\_Species 28](#_Toc428313849)

[6.2.9 Table: Annotation 29](#_Toc428313850)

[6.2.10 Table: Annotation\_Validation 29](#_Toc428313851)

[6.2.11 Table: Annotation\_Approvement 30](#_Toc428313852)

[6.2.12 Table: Approved\_Annotations 30](#_Toc428313853)

[6.2.13 Table: Evidence 31](#_Toc428313854)

[6.2.14 Table: Annotation\_Evidence 31](#_Toc428313855)

[6.2.15 Table: Publications 31](#_Toc428313856)

[6.2.16 Table: Author 31](#_Toc428313857)

[6.2.17 Table: Author\_Publication 32](#_Toc428313858)

[6.2.18 Table: Xref 32](#_Toc428313859)

[6.2.19 Table: Xreference\_relation 32](#_Toc428313860)

[7 References 32](#_Toc428313861)

[Appendix A: Definitions and Abbreviations 33](#_Toc428313862)

# Introduction

* This long term project need a document to help participants to understand the background, terms, purpose, etc., So that they could comprehend the situation as soon as possible
* The developers and researchers, no matter come from computer science or biology, need a document to confirm the definition and verify the understanding of the whole project.
* This document will be the most authoritative handbook and reference while developing the project.
* The structure of this document and guide of reading
* Some foreword about this section, like purpose (intent to give enough definition and examples to help understanding)

## Planteome Project

Why Planteome project is important?

What problem we try to solve by this project?

Basic information about ‘Common Reference Ontologies and Applications (cROP) for Plant Biology’

## Gene Curation Tool (GCT)

What is gene curation?

What the purpose of developing GCT?

Who will be the beneficiary from GCT and how? (the meaning of developing GCT)

## Basic knowledge about Genes

Following terms are ordered by the inclusion relationship, from the smaller unit to bigger ones.

*Gene*

DNA

RNA

Chromosome

*Genome*

*Germplasm*

*QTL*

*Gene product*

Phenotype

## Ontology

What is ontology

Ontology in biology

*Gene Ontology (GO)*

What is GO? Some examples of GO

*Plant Ontology (PO)*

What is PO? Some examples of PO

Plant Trait Ontology (TO)

What is TO? Some examples of TO

Plant Environment Ontology (EO)

What is EO? Some examples of EO

Plant Stress Ontology (PSO)

What is PSO? Some examples of PSO

## Annotation

What is Annotation?

What’s the use of Annotation?

Where the Annotation data come from?

GO Annotation examples

# Existing Ontology Databases

## Amigo and Amigo2

Amigo has a deep relationship with our project. So in this section, we will demonstrate what is Amigo and how we will use utilize it in our project.

### Introduction

What is Amigo?

### Features

Build based on Gene Ontololgy

Justin is extending it to other ontology data (Amigo2)

Provide all ontology data

GCT is a backup and database could be easily modified

GCT would also how the data in process, unapproved, out date

### GO Annotation File (GAF) Format

Introduction of Annotation File Fields (GAF 2.1)

The explanation of the format

## Gramene

Introduction:

Features:

## AgriGO

Introduction:

Features:

## Conclusion

As above statement, we conclude following features which would be provide great convenient for the biologists to use GCT:

# Software Requirement Specification

Itemize the user requirements,

## Product Perspective

Who would need to use GCT?

The benefit of using GCT

The meaning of using GCT

## Scope

We descript the features in scope of GCT.

1. …
2. …

## Operating Environment

Database: MySQL

Server: Apache

## Role Based Access Control (RBAC)

What is RBAC?

How we use it in our system?

## Product Functions

GCT should support the following use cases:

|  |  |  |  |
| --- | --- | --- | --- |
| Class of use cases | RS\_ID | Use cases | description |
| unregistered user’s capabilities (basic contributor) | 1-1 | register |  |
| 1-2 | browse annotation information |  |
| 1-3 | browse ontology information |  |
| 1-4 | browse gene information |  |
| 1-5 | comment on annotations |  |
| registered user’s capabilities (expert contributor) |  | Login |  |
|  | profile management  (include change password) |  |
| 1-2,1-3,1-4 | browse annotation/ontology/gene information |  |
|  | export annotation information |  |
|  | browse credit |  |
|  | edit annotation |  |
|  | save annotation draft |  |
|  | save note to user self |  |
|  | flag annotation |  |
|  | add annotation |  |
|  | comment on annotation |  |
| Admin’s capabilities |  | all capabilities of registered user |  |
|  | manage user’s information  (credit, password, profile) |  |
|  | manage users’ role |  |
|  | ban user from activities |  |
|  | approve annotation modifications  (edit, flag and add) |  |
|  | edit publication |  |
|  | edit evidence |  |
| Super admin’s capabilities |  | All capabilities of admin |  |
|  | Import data (gene data, annotation data, etc.) |  |

## User role Hierarchy

The users are divided into four levels (0-3), different level correspond to different ability range. The admin (both super admin and admin) are able to manage the users’ roles. Also, the admins could be able to edit the personal information of the user in backstage.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Role | comment | edit | add | flag | approve | import |
| super admin | X | X | X | X | X | X |
| admin | X | X | X | X | X |  |
| contributor expert | X | X | X | X |  |  |
| basic contributor | X |  |  |  |  |  |

## Contribution Credit Scheme

Each contribution of the annotation may lead to an accumulation of credit. The users could see the ranking result and get to know who contribute more to the whole gene curation system.

|  |  |  |
| --- | --- | --- |
| **Credit Rule** | | |
| **action** | **score** | **description** |
| comments | **1** |  |
| make suggestion | **1** |  |
| suggestion got approved | **2** |  |
| edit annotation | **2** |  |
| add annotation | **2** |  |
| edit/add got approved | **2** |  |

## User management subsystem

In this section, we will introduce all use cases related to the user’s profile information.

### Use Case: Register

Unregistered user could only browse the information from the web, every users of could register and then login to get higher access of the system.

|  |  |  |
| --- | --- | --- |
| **USE CASE: Register** | | |
| **Description** | User could register to become an expert contributor | |
| **Main Actor** | non-registered user | |
| **Trigger** | click the register button | |
| **Typical case Scenario** | **Action** | **Response** |
| 1 fill all required information |  |
| 2 click submit button | 3 system will check the information been filled |
|  | 4 system will save the data |
| **Alternate Scenario** | 4: if 3 found the information provided is not correct, the system will prompt a dialogue to indicate the problem | |
| **Result** | successfully create a new user | |
| **Constraint** | the user’s name should be an email address  the user need to select specialty from a drop list  the user name should be unique  the password need to be input twice, and both of them should be same | |

|  |  |  |
| --- | --- | --- |
| **INPUT** | | |
| **NAME** | | **DESCRIPTION** |
| name | first name | User’s name will be shown on the pages. |
| last name |
| middle name |
| affiliation | institute |  |
| XXX |  |
| Specialty | | the user could only edit the annotation belong to specific specialty |
| user\_name | | Email address, used to login, need to check if there is exist a same user\_name and the format of the user\_name is correct. |
| phone | | contact information |
| country | |  |
| password | | need to be input twice to confirm |

|  |  |
| --- | --- |
| **OUTPUT** | |
| **NAME** | **DESCRIPTION** |
| success |  |
| user\_name occupied |  |
| miss required information |  |

### Use Case: Login

### Use Case: Ban User

…

### Use Case: Edit Specialty

…

## Annotation management subsystem

### Use Case: browse/edit/add annotation

### Use Case: Save annotation draft

### Use Case: Save note

### Use Case: flag annotation

### Use Case: Comment on annotation

## Object management subsystem

…

## Publication management subsystem

…

# API design

In this section, all the APIs related to Gene Curation Tool system will be descripted.

## Object import API

We need to develop an API to import the Object data

## Annotation import API

API to import the annotation

## Annotation export API

API to export the annotation

## Utilize API to get Ontology information

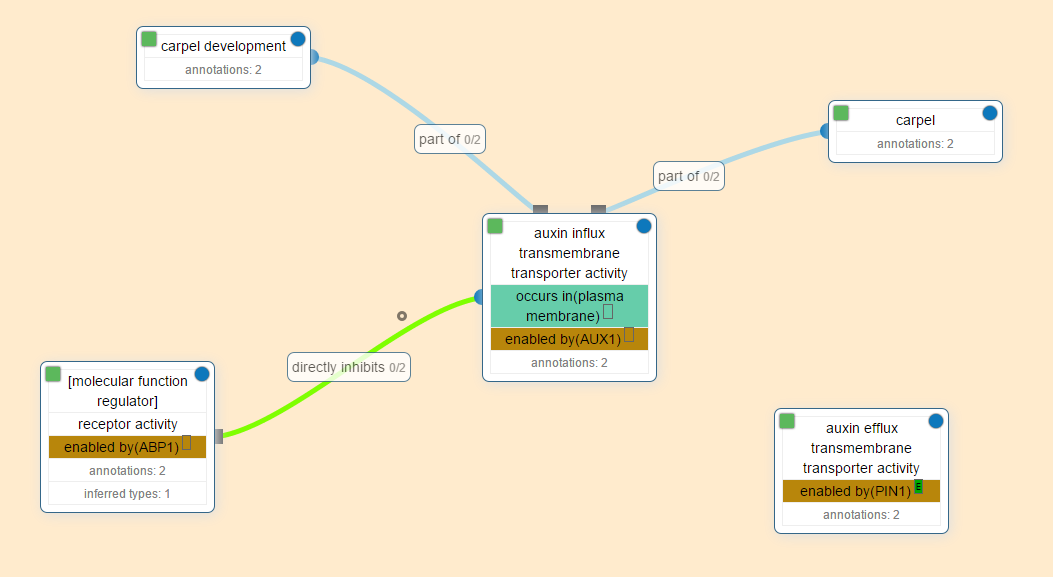
The ontology data should be accessed by using Amigo API

# User Interface Design

The hand drawing user interface design

## Web Pages Design

## Graphic Gene Curation



# Database Design

In this section, we will illustrate the database design of the GCT project.

## ER diagram



## Tables design

In this section, we will itemize all tables in the database and the characteristic of rows of the table.

### Table: Users

|  |  |  |  |
| --- | --- | --- | --- |
| **ATTRIBUTE NAME** | **DESCRIPTION** | **DATA TYPE** | **NULLABLE** |
| User\_ID | PK | Integers | no |
| Username | Unique Username |  |  |
| Password | Password |  |  |
| Email | User’s email |  |  |
| Last\_IP | The IP of last login |  |  |
| Last\_Login\_Timestamp | The time of last login |  |  |
| User\_Level | Can be 0,1 or 2 (Defines the level of the user) |  |  |
| Credit | the credit for the contribution of the system |  |  |

### Table: User\_banned

### Table: Specialty

### Table: User\_Specialty

### Table: Object

### Table: Synonyms

### Table: Species

### Table: Gene\_Species

### Table: Annotation

### Table: Annotation\_Comment

### Table: Annotation\_Note

### Table: Annotation\_Validation

### Table: Annotation\_Approvement

### Table: Approved\_Annotations

### Table: Evidence

### Table: Annotation\_Evidence

### Table: Publications

### Table: Author

### Table: Author\_Publication

### Table: Xref

### Table: Xreference\_relation

# References

1. Ashburner, Michael, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis et al. "Gene Ontology: tool for the unification of biology." Nature genetics 25, no. 1 (2000): 25-29.<http://www.nature.com/ng/journal/v25/n1/full/ng0500_25.html>
2. …

# Appendix A: Definitions and Abbreviations

|  |  |
| --- | --- |
| GCT | Gene Curation Tool |
| cROP | “Common Reference Ontology for Plants”, a set of ontologies concerning plants. This was the name of the project before it was changed to Planteome by Dr. Pankaj. |
| Ontology | Ontologies have long been used in an attempt to describe all entities within an area of reality and all relationships between those entities. An ontology comprises a set of well-defined terms with well-defined relationships. The structure itself reflects the current representation of biological knowledge as well as serving as a guide for organizing new data. Data can be annotated to varying levels depending on the amount and completeness of available information. This flexibility also allows users to narrow or widen the focus of queries. Ultimately, an ontology can be a vital tool enabling researchers to turn data into knowledge. Ultimately, an ontology can be a vital tool enabling researchers to turn data into knowledge. |
| Annotation | Information about a gene that is attached to these vocabularies (concepts) in ontologies and used to describe their relationships. Often contain an evidence code and literature associated with it to back up this newly found information about a gene according to Dr. Pankaj. |
| Gene Ontology | The Gene Ontology refer to vocabulary applied to all gene and protein roles in cells. Which including three main parts: the biological process (p), molecular function (f) and cellular component (c). [1] |
| AmiGO | A web tool for accessing the Gene Ontology project’s data (including browsing genes and their corresponding annotations). |
| Genomes | The complete genetic material of an organism [11]. |
| MapReduce | A programming model that where you split up data across processes, so that they can be ran independently in parallel [6]. |
| NoSQL | “Not Only SQL”, steer away from your traditional relational database model for better performance on flat data [6]. |
| Phenotypes | Observational characteristics of an organism [11]. |
| Protein | Large [biomolecules](https://en.wikipedia.org/wiki/Biomolecule), or [macromolecules](https://en.wikipedia.org/wiki/Macromolecule), consisting of one or more long chains of [amino acid](https://en.wikipedia.org/wiki/Amino_acid) [residues](https://en.wikipedia.org/wiki/Residue_(biochemistry))” [8]. |