基于Django框架的Recon2代谢模型的检索与应用系统开发

作 者 姓 名： 米昊炀

指 导 教 师： 巩天星，Peter A.J Hilbers

单 位 名 称： 中荷生物医学与信息工程学院

专 业 名 称： 生物医学工程

东 北 大 学

2018年6月

**The Development of an Index and Application System upon Metabolic Model Recon2 Based on Django Framework**

by Haoyang Mi

Supervisor: Tianxing Gong, Peter A.J Hilbers

Northeastern University

June 2018

# **毕业设计（论文）任务书Tasks of Final Project**

|  |
| --- |
| **毕业设计（论文）题目：基于Django框架的Recon2代谢模型的检索与应用系统开发**  **Title of Final Project：The Development of an Index and Application System upon Metabolic Model Recon2 Based on Django Framework** |
| **设计(论文)的基本内容：**  **Basic contents of Final Project：**  (1) Import the .json format data to database and connect to Django.  (2) Develop administration and user page to manage metabolic data.  (3) Visualize and analyze complex network for any number of reactions.  (4) Achieve the online simulation and analysis of metabolic model |
| **毕业设计（论文）专题部分：**  **Special topic of Final Project（If have）**  **题目：**  **Title：**  **设计或论文专题的基本内容：**  **Basic contents of Final Project or special topic：** |
| **学生接受毕业设计（论文）题目日期**  **Date of this student got the title of Final Project**  **第　　周 week \_\_\_1\_\_\_**  **指导教师签字：**  **Signature of Supervisor:**  **年　　月　　日**  **Date** |

基于Django框架的Recon2代谢模型的检索与应用系统开发

摘要Abstract（In Chinese）

新陈代谢是机体生命活动的基本特征。系统生物学的发展使代谢网络的研究得到重视，并促进化学反应的分析趋向专门化，而计算机科学为解码人体复杂代谢提供了技术手段。基于此，人类全基因组尺度代谢模型Recon应运而生。Recon描述了人体内近乎所有的遗传代谢信息，为模拟分析提供了必要工具。然而，数据的复杂性使得目前没有针对Recon模型的单一数据库。且已有的模型检索系统功能较为单一，没有应用价值，故有关人类代谢网络的研究十分局限。

在本项目中，一种新型的人体代谢网络检索与应用系统被开发出来。我们选取Recon3D，一个最新的模拟人体全代谢网络模型，作为本系统的数据来源。通过可视化工具将数据库与Django框架相连接，创建数据访问接口。同时，管理与用户界面用来实现反馈、代谢数据的管理以及高级查找等检索功能。由于BiGG数据库提供的模型文件囊括了反应与代谢物的包含关系，以此为参考，在系统内实现了数据间的互联后，便可通过Javascript图形库设计匹配算法实现任意数量反应的可视化并生成力导向图。反应间的相互关系被确定后，则可通过悬停算法实现主要结点的突出与边的高亮。该算法有利于帮助分析代谢物的来源与去向，明确反应间的关系以及在代谢网络中的位置信息。

此外，通过导入COBRApy第三方库以及最优控制器，本系统还实现了代谢模型文件的上传以及在线模拟。用户可根据需要获取模型数据，并进行流平衡分析。

系统生物学的研究与人类健康关系紧密，通过引入人类代谢网络的模拟方法，有助于发现潜在的药物作用位点、预测药物作用并依此提出新型治疗方案。因此，对人类代谢模型数据的严格分类检索、清晰可视化以及在线分析有助于相关领域的进一步探索。

**关键词**：系统生物学；代谢网络；Recon3D；Django框架；Python开发；检索；可视化；COBRApy；在线模拟；流平衡分析

The Development of an Index and Application System upon Metabolic Model Recon2 Based on Django Framework

Abstract

Metabolism is the fundamental of vital movement. As the development of system biology, the research on human metabolism and the compartmentally analysis on biochemical reactions are gaining attentions. Simultaneously, computer science helps to decode the complex metabolic networks. Consequently, the human metabolism model Recon is created. Recon describes majority of metabolism information and contributes to simulation and modeling. However, there is no database solely available for Recon because of the data complexity, besides, the function of current index database is relatively simple and with no application value, so the research on human metabolic network(HMN) is very limited.

In this project, a brand-new index and application system of HMN is developed. Recon 3D, an updated model for metabolic simulation, is chosen as the data source. visualization tool connects database with Django framework and creates interface for data access. Meanwhile, administration and user page is designed for realize the feedback, data modify and advanced search function. Because the model file from BiGG database reflects the reactions-metabolites relationship, after achieving the interconnection, force directed graphs for any numbers of reactions visualization would be derived based on Javascript graph library and matching algorithm. When confirmed the relationship, the mouseover algorithm is processed to highlight major nodes and links. This helps to analyze the source and destination of metabolites, explicit reactions interrelation and their locations within networks.

Additionally，COBRApy package and optimization solver are imported to handle with the metabolic model uploading and online analysis. Users can obtain metabolic data on request and conduct flux balance analysis.

System biology tightly associated with human healthcare. Introducing the HMN modeling methods would contribute to identify potential drug targets and test drug effects, upon which novel treatment is proposed. Therefore, the strict index, clear visualization and online analysis is conducive to explore relative research areas.

**Key words**：System Biology; Metabolic Network; Recon3D; Django Framework; Python Development; Index; Visualization; COBRApy; Online Analysis; Flux Balance Analysis

目录Index

**毕业设计（论文）任务书Tasks of Final Project** I

摘要Abstract（In Chinese） II

Abstract III

目录Index IV

第一章 ：绪论 Chapter I: Introduction 1

1.1 Computational modeling and simulation in system biology 1

1.2 Genome-scale metabolic model 2

1.3 The overall research status on human metabolic network pathways 3

1.4 The significance of developing index and application system for human metabolic model 4

第二章 ：工具与方法 Chapter II: Tools and Methodology 5

2.1 Operating system and development environment 5

2.1.1 Operating system selection 5

2.1.2 Development environment 5

2.2 Database selection and visualization tool 6

2.2.1 Database selection 6

2.2.2 Introduction to Navicat 6

2.2.3 Introduction to JavaScript Object Notation(JSON) format 7

2.3 Webpage development technology 7

2.3.1 Introduction to HTML5 7

2.3.2 Introduction to CSS 8

2.3.3 Introduction to JavaScript graph library 8

2.4 Analytic tool 9

2.4.1 System Biology Markup Language(SBML) 9

2.4.2 Constrained-Based Reconstruction and Analysis for Python(COBRApy) 10

2.4.3 Introduction to optimization solver 10

第三章 ：需求分析与设计 Chapter III: Demand Analysis and Design 12

3.1 Administration page design and intended functions 12

3.1.1 Authentication and login 12

3.1.2 Management of metabolic model data 12

3.1.3 Fundamental search engine and advance searching 13

3.1.4 Establishment of mutual relationship 13

3.1.5 Record operations history 14

3.1.6 Record user feedbacks 14

3.2 User page design and intended functions 15

3.2.1 Homepage and interfaces set 15

3.2.2 Metabolic data compartments 16

3.2.1 Global search 16

3.2.1 Online analysis 16

3.3 Database design 16

第四章 ：网站实现过程 Chapter IV: Website Realization Process 17

第五章 ：分析与总结 Chapter V: Analysis and Conclusion 18

参考文献References 19

致谢Acknowledgements 20

# 

# ：绪论 Chapter I: Introduction

The study mainly orients to the development of web-based system framework and related functions of model data application. With the reference of master thesis “Recon 2.2: from reconstruction to model of human metabolism”, I mainly worked with analyzing the structure and content of Recon model in order to design a system better fit the characteristic of Recon and utilize it. Following the guidance of Neil’s paper, I figured out the layout of metabolic model by evaluating Recon2.2 in SBML format. By comparison, the principle of how .json format file describe metabolic information is clarified. JSON is then considered acceptable for my project and thus an update has been made and the data source is switched to Recon3D.

## Computational modeling and simulation in system biology

System biology is an interdisciplinary area of study that focuses on complex interactions within biological systems. The major goal is to ensembles different scales of information, upon which to understand the how biological networks function as an integrity. Obviously, unlike molecular biology, which only concerns the behavior of individual gene and proteins, system biology concentrate on emergent properties(mainly on genetic level) and how organisms functioning as a system. Just like experimental methods to traditional biology, computational and mathematical modeling are core towards system biology research. In 1952,

Figure 1.1 The overview of signal transduction pathways

one of first numerical simulation in cell biology was published[1]. In 1960, Denis Noble developed the first computer model of the heart pacemaker[2].

Subsequently, several approaches have been used to study complex molecular systems, the computing efficiency exploded making large volume of biological data accessible and manageable. In 1997, a Japanese research group published the first quantitative model of the metabolism of a hypothetical cell. Around 2000, the well-known Institutes of System Biology have been founded both in Seattle and Japan. Meanwhile, several genomics projects have been put forwards, among which the Human Genome Project[3], led to novel, collaborative ways of working on problems in the biological fields of genetics. So far, system biology has gradually emerged as a relatively independent interdiscipline that combines computational and bioinformatics methods to explore cellular signaling networks, metabolic networks and genome-scale network domains.

## Genome-scale metabolic model

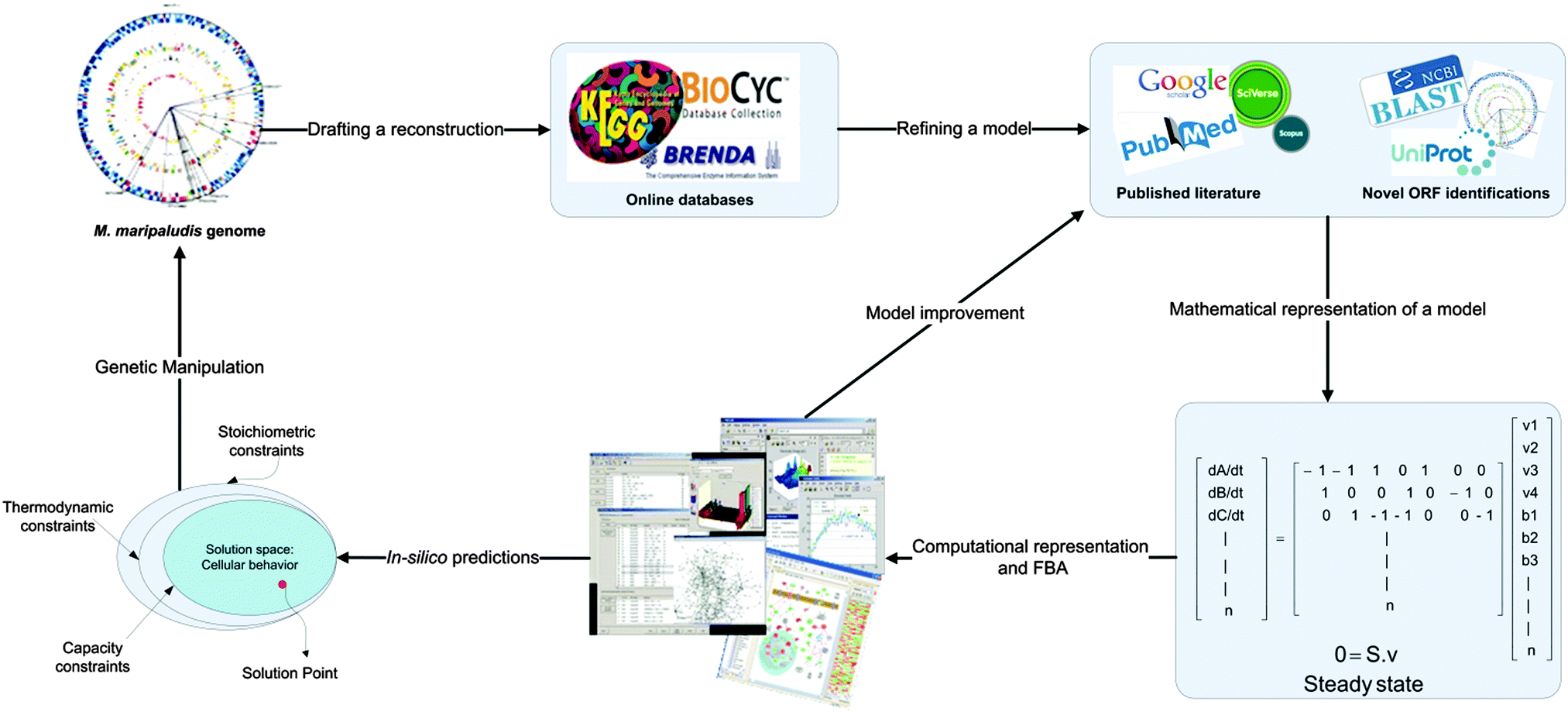
Metabolism is defined as a comprehensive cluster of biochemical reactions to satisfy the fundamental need of living organisms. To quantitative analyze metabolic process, bioengineers introduced mathematical and modeling thoughts to simulate biological networks. To describe metabolic pathways as accurate as possible, researchers reconstructed genome-scale metabolic netork of chemical reactions from enzyme-reaction interrelations. Basically, a well-designed model should contains three compartments: genes, metabolites, and reactions. The mutual related property becomes the basis for network reconstruction, together with the curation and

Figure 1.2 An example process of model reconstruction for a specific organism

refinement procedure a model for simulation purpose is derived.

In 1999, Palsson’s group published the first genome-scale metabolic network model that can be used to simulate the phenotype of metabolism[4]. Afterwards, a bunch of models have been constructed and the research on computational biology is increasing sharply. Many classical models have been identified. Meanwhile, multiple databases are established to store biological data, like KEGG, BiGG and BioModels. All of these advancements reflect the prosperous of computational and system biology.

## The overall research status on human metabolic network pathways

The construction of human metabolic model is based on the former experience of developing micro-organisms models. Plenty of useful databases, which are mutually supplementary, have been built for search in genome, metabolites or reactions. In 2007, the first model of human genome-scale metabolic networks models-Recon1 was finished and published based on KEGG database, and later was thermodynamically curated by LE Quek[5]. Almost at the same time, another high-quality human metabolic network was reconstructed by a research group from the University of Edinburgh. This model, as referred to as Edinburgh Human Metabolic Networks(EHMN), though comparable to Recon1 regarding the scale, is innovative since it reorganized 70 human-specific metabolic pathways according to their functional relationship[6].

In 2016, the ever-most predictive model update Recon2.2 have been published. With extensive manual curation, this model reaches the size of 5324 metabolites, 7785 reactions and 1675 genes[7]. Recently, the model for human metabolic pathways has been further perfected. With more genes, metabolites and reactions been depicted in the model for the first time, the Recon3D, a most updated model for homo sapiens which was published in 2018, provides a more powerful source for human metabolic pathways, even the structure of gene variation.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Human Metabolic Model | | | | | | |
| Recon3D | Recon2.2 | Recon2.04 | Recon2 | EHMN | Quek | Recon1 |
| Metabolites | 5835 | 5324 | 5063 | 5063 | —— | 4962 | 2766 |
| Reactions | 10600 | 7785 | 7440 | 7440 | 2823 | 7785 | 3741 |
| Genes | 2248 | 1675 | 2140 | 2191 | 2322 | 1675 | 1905 |

Table 1.1 Reconstruction scales for multiple human metabolic models

## The significance of developing index and application system for human metabolic model

This project is initially inspired by the development history of human metabolic models. Within few years, multiple models and updates have been made upon human metabolic pathways research. However, the models were designed oriented from different perspectives, the simulation results may also vary to some degrees. To unify the further study on related areas, a standard model should be confirmed.

The project is also by part inspired by my personal learning experience on genome-scale models. The prerequisite work for metabolic pathways is quite complicated. First the developing environment should be configured. Analysis package is ought to be imported in order to use analytical functions. Furthermore, optimization solver should be set to perform mathematical simulation. Besides, the function of current existing databases for metabolism study are imperfect. For example, though BiGG database possesses the updated model for human metabolic networks, the visualization for core reactions is not available. Though KEGG database provides complete pathway maps for biochemical reactions and are accessible for download as well as clear reference for metabolic information, it only functions for index. And even though BioModels database includes comprehensive models from E.coli to human beings and provide extremely detailed relevant information, SBML is the only format available for metabolic modeling, and consequently visualization is not possible.

To significantly reduced the preparing work for researchers, in this study, I designed a web-based index and application system(IAS) for Recon model. Recon3D is considered as the ‘standard’ model. Indexation is fully achieved to provide information search function for bioengineers. Besides, constrained-analysis package in Python and optimization solver are imported to project through backend. In that case bioengineers can complete analysis without referencing multiple databases and wasting their time on environment configuration.

Figure 1.3 Layout for IAS structure

# ：工具与方法 Chapter II: Tools and Methodology

## Operating system and development environment

### Operating system selection

There are currently many operating systems available for development, including Unix, Linux, Windows and macOS. Linux is an Unix-alike system, thus has multiple similar functions and characteristics to Unix and is compatible to Unix[7]. However, Linux and Unix are not widely used and provide far less software when compared to macOS and Windows. Besides, the Unix and Linux are relatively complicate to use and are immature in some degrees. Based on comprehensive consideration, macOS and Windows are ideal for my project.

Currently, macOS is increasingly welcomed by programmers for its Linux-like characteristics. Besides, Python execution environment is pre-installed to macOS, which liberated developers from configuration.

Finally, macOS is selected as the primary development environment. The specific version is: macOS High Sierra, 10.13.4.

### Development environment

2.1.2.1 Introduction to Python language

Python is an advanced language for general purpose programming. Since first been published in 1991, Python is gaining popularity among developers for its clarity and code readability. Python features a dynamic type system and automatic memory management, besides, it supports multiple programming paradigms including object-oriented, imperative, functional and procedural and has a large and comprehensive standard library. Consequently, it is relatively easy to use for amateurs[8].

Though Python 2.X is till the current mainstream version, however, multiple mainstream third-party frameworks claimed they will no longer provide support for Python 2.X. Considering the later system maintenance issue, I chose Python 3.6.4 as the primary developing language, which is also a comparatively updated and robust version among Python 3.X.

2.1.2.2 Introduction to Django framework

Multiple web application frameworks have been designed to support the development of dynamic websites, web application software and internet services. Traditional frameworks is lightweight class, which means they only provides core functions. However, Django provides a one-stop service on URL, ORM and even authentication dispatch[9]. Furthermore, the admin

|  |  |
| --- | --- |
|  | Web frameworks |
|  | Classified by Language |
| C++ | CppCMS · Wt |
| Java | AppFuse · Flexive · Grails · GWT · ICEfaces · ItsNat · Jspx · OpenXava |
| Javascript | Ample SDK · Angular/Angular JS · Backbone.js · Closure · jQuery |
| PHP | CakePHP · CodeIgniter · Fat-Free · FuelPHP · Flow · Gyroscope · Horde |
| Python | BlueBream · CherryPy · Django · Flask · Grok · Pyjs · Pyramid · web2py |

characteristic of Django make it extremely easy for backend development.

Table 2.1 Web framework category

Based on the demand analysis of my project, Django is selected as the web framework for developing. And the selected version is: Django 2.02.

2.1.2.1 Introduction to Pycharm

There are currently plenty of integrated development environment(IDE) for python, among which Pycharm is considered as the most powerful one. Pycharm has very comprehensive features, including code completion, python refactor and integrated Python debugger. Besides, Pycharm has embedded project interpreter and support for web frameworks including Django so that developers are able to freely install necessary packages[10]. Consequently, Pycharm is the ideal IDE for web development and python programming.

Selected version in this project: 2017.3.3(Professional Edition).

## Database selection and visualization tool

### Database selection

Django select SQLite as the database by default, however, SQLite behaves relatively bad when handle a large volume of data. MySQL, as the current mainstream open-source database, have been used for an increasingly number of large-scale websites and applications[11]. Besides, MySQL is an user-oriented database. Apart from the traditional way to operate database by terminal, MySQL could also be administrated by MySQL Workbench, which is a graphical tool for database design and modeling. Consequently, MySQL database is the ideal software for amateur backend developers.

Selected version in this project: 5.7.21 MySQL Community Server(GPL).

### Introduction to Navicat

When handling a huge amount of data, an ideal graphical tool for data management is necessary. In my project, Navicat visualization software is selected as the supplementary tool to database construction. Navicat is created by PremiumSoft Cyber Tech Ltd. In 2002, and is suitable to multiple operating systems. Navicat has following functions:

1. Code snippet,
2. Visual query-builder,
3. SSH and HTTP tunneling,
4. Import and Export and backup of data(support for multiple file format including Excel, JSON, text and XML and DBase),

and so on. Because we need to import Recon model to our database, thus Navicat would greatly facilitate my work on database construction.

Selected version in this project: 11.1.13-Enterprised.

### Introduction to JavaScript Object Notation(JSON) format

JavaScript Object Notation(JSON) file, is an open standard file format that uses human-readable text to transmit data objects consisting of attribute-value pairs and array data types. System biology models normally describes thousands of data thus an easy loaded document file is preferred in my project.

Currently, metabolic models in BiGG database are either stored in JSON or SBML format. Though SBML format is most widely used for describing computational models, this XML-based format is relatively hard to parse and is not supported for data import via Navicat.

Therefore, Recon3D model in JSON format is considered as the data source in my project.

## Webpage development technology

### Introduction to HTML5

HTML5(HyperText Mark-up Language), is the most widely-applied language for website development, and is also the primary language that constitutes webpage text. HTML5 was created to conveniently combine the text/graph stored in one computer to text/graph stored in another and becomes an integrity. HTML5 text is a descriptional text that made up with HTML5 commands, which explained text, graphs, animation sounds, charts and links. The two major components of HTML5 are head and body. ‘Head’ describes the information needed by browsers, and ‘body’ describes those information in detail.

Besides, HTML5 is a universal language for website development, an easy full mark-up language. It allows developers build complex pages that combines text and graphs. These pages could be viewed by anyone in anywhere, regardless of the type of browsers or computers.

### Introduction to CSS

CSS(Cascading Style Sheets), is a style sheet language used for describing the presentation of a document written in mark-up language. Generally speaking, it is not likely that a website is designed only to present raw text or graph without any layouts and effects. For most cases, we should further make the website more beautiful and pleasant to use. CSS is just what we need, it is designed to enable the division of presentation and content including colors, fonts and layouts. This division can boost the overall accessibility, provide more flexibility and control in the specification of presentation characteristics, enable multiple web pages to share formatting by specifying the relevant CSS file, and reduce complexity and repetition in the structural content.

Figure 2.1 A simple example of the structure of CSS file on Google search

Basically, CSS language helps to compartmentalize information within a webpage and enhanced readability of contents. Consequently, introduce proper CSS file to project has gradually becomes an integrate part in webpage development.

### Introduction to JavaScript graph library

In web development, sometimes developers need to graphing on web page. In that case, traditional developing language is incompetence in doing this job. In my project, the visualization of reactions and the graphing of complex networks are also depend on the dynamic open-source figure construction tool.

To do this, JavaScript graph library is introduced to my project. JavaScript(JS) is a high-level, interpreted programming language. Together with the HTML and CSS, JS is considered one of the core technologies of web development. JavaScript enables interactive web pages and thus is an essential part of web application. Unlike static text language such as HTML and CSS, JS adds dynamic characteristic to web contents and is used as the core algorithm development language in my project.

There are currently multiple JS graph library available for developers. They have similar functions and usage, however, they could still be divided as two types. First is graph operation-oriented and the other is data-oriented. The core of algorithms in my project is intelligently manage metabolic data within database, so data-driven library is preferred. D3.js(Data-Driven Document) library, the most popular front-end visualization tool, is selected as the JS library for my project. When compared with other JS libraries, d3.js provide with the high degree of freedom and enhanced the interaction between users and data.

Figure 2.2 D3.js samples on Google search

## Analytic tool

### System Biology Markup Language(SBML)

The System Biology Markup Language(SBML), is a representation format based on XML, for communicating and storing computational models of biological pathways. SBML describes multiple different scales of biological processes, including metabolic networks, cell signaling pathways, regulatory networks and indicates all the necessary information for simulation.

A typical SBML file includes “model id”, “listOfCompartments”, “llistOfSpecies”, “listOfReactions”, “listOfParameters” sometimes also “listOfGenes” if gene analysis is needed. “listOfUnit” may sometimes been included if various kind of value in the model is requested. Besides, “listOfFluxBounds” and “listOfObjectives” will existed in “fbc extends” as the symbol if the model is available for flux balanced or flux variability analysis. Figure 2.3 explicates the

structure of an example file in SBML format.

Figure 2.3 The structure of Recon3D model in SBML format

### Constrained-Based Reconstruction and Analysis for Python(COBRApy)

Constrained-based reconstruction and analysis(COBRA) is a method to reconstruct and analyze genome-scale models of metabolic networks. To effectively utilize the data, multiple computational methods and software are associated with metabolic models. COBRA has two interfaces available, which are MATLAB toolbox and Python package. In my project, COBRApy is selected as the analytic tool considering third-party package support is embedded in Django framework.

COBRApy is designed exclusively for metabolic model using Python language. Bioengineers could simply import the package to project and use pre-defined functions to analyze models. COBRApy provides accessible COBRA methods that are commonly used such as flux balance analysis, flux variability analysis and gene deletion analysis. Through COBRApy, one can create and manage metabolic models, access popular solvers for mathematical calculation and model analyzing as well as inspect models and draw conclusions on gene-related operations. In our project, embedded COBRApy do not allow users to modify models, but provide them with interface to conduct simulations and modeling.

### Introduction to optimization solver

Given a linear system converse a set of inputs to output values, described by a mathematical function f, optimization refers to the generation and selection of a best solution from a certain set of available alternatives,[] by systematically selecting input values from within an allowed set, computing the value of the function, and recording the best value found during the process. Many real-world and theoretical problems may be modeled in this general framework. The optimization theory could be applied to solve biological problems. The detailed mathematical principle on how to work out the best solution to an equation will be explained in the chapter four, for simplicity, here we will only discuss the optimizer to use.

Mathematically speaking, most optimization problems are possible to be write in the generic form:

minimize

subject to

Where and are functions of the design vector

As a matter of fact, COBRApy does not implement any algorithms to find solutions to optimization problems, however, it do creates a biologically motivated framework to these techniques to simplify how metabolic pathways work without focusing on how that produces to a better solution. The math is actually done by tools, which are called, optimization solver. COBRApy has available internal interfaces for multiple solvers such as glpk, Gurobi and cplex. Those solvers are free to use for academic purpose. Considering the software compatibility issue, cplex is selected as the optimizer for mathematical simulation.

# ：需求分析与设计 Chapter III: Demand Analysis and Design

## Administration page design and intended functions

Administration page is created as the backend for developers to better manage data obtained from the database. Generally speaking, the emphasis on administration page focuses on displaying how the data is structured without applying complex contents and CSS.

Administration page will only provide interface to administrators, other unauthorized access should be prohibited. Through this page, administrators could easily manage all the data stored in the database, for example, to decide which data to be presented via front-end while which are not and respond to user request. The significant point is, a well-designed backend is the guarantee to a robust website. It not only provide a fast way to figure out the core of front end, but also simplify the process to access data and empowers administrators to operate database directly.

### Authentication and login

Because administration page only provides access to administrators, the authentication is required to keep unauthorized users away. Administrators will be assigned with a pair of user name and password individually through which to successfully login. For security reasons, every administrator’s username-password pair is recorded in the backend. In that case, when someone forget his/her credentials, others could simply refer to the list and remind him/her. Besides, the users list should also records the contact information and functions as a convenient platform for communication. When emergencies occur, relevant responsible persons can be informed in time and problems or further loss would be prevented.

### Management of metabolic model data

The fundamental function of administration page is the management of data. Normally, ‘management’ refers to add, delete, modify and search. For most projects, ‘add’ , ‘delete’ and ‘modify’ are mandatory, however, in my project, they are relatively useless. The reason is, metabolic models are strictly pre-defined and tested, which means any random modify to data value will leads to the malfunction of simulation tools. Nevertheless, for version update purpose and security issues, they are still worth developing.

### Fundamental search engine and advance searching

‘Search’, obviously, is the core of backend functions. Administrators will deal with an incredible number of data within the metabolic model and thousands of user feedbacks a day. So a robust search engine and advanced searching function is definitely preferred. First, search should be achieved in each compartment. In the administration page, data will be divided into different compartments. Even though the volume of data will be reduced along with the increase of sections, the amount of data is still exceed the limit of manually indexation.

Search, is realized based on the extraction of key words. In my project, the selection of key words varies according to the type of data. As I claimed in chapter two, the three major components of metabolic data is metabolites, reactions and genes, each featured by different characteristics. As for metabolites, the potential key words for searching should be id, name and formula. As for reactions, the potential key words for searching should be id, name and metabolites involved in the corresponding reaction. As for genes, the potential key words for searching should be id and name. Apart from the constraint that the field for search is solely exist in corresponding type of data, the data stored in the field should also be unique, which means, elements that shares same strings or values are prohibited. Based on these two constrained, id and name are ideal fields for searching.

Besides the fundamental search using key words, the engine is ought to be intelligent to address the exceptions that may arise during practical operations. There are over ten thousands reactions that are described in the Recon3D model and the id for each reactions though different, but may sometimes very similar. For example, ‘10FTHF6GLUtm’ refers to ***5-glutamyl-10FTHF transport reaction*** that took place in mitochondria, while ‘10FTHF6GLUtl’ refers to ***5-glutamyl-10FTHF transport reaction*** that took place in lysosome. The point is, sometimes the ids are complicated and alike so that users are easily to get confused. In that case, fuzzy search is extremely in need. Fuzzy search liberates users from remembering key words for all the reactions and users only need to have some knowledges about what they are looking for and search engine will present all the possible search results for users. To accomplish this goal, search engine should be equipped with auto-complete function and wildcard search.

### Establishment of mutual relationship

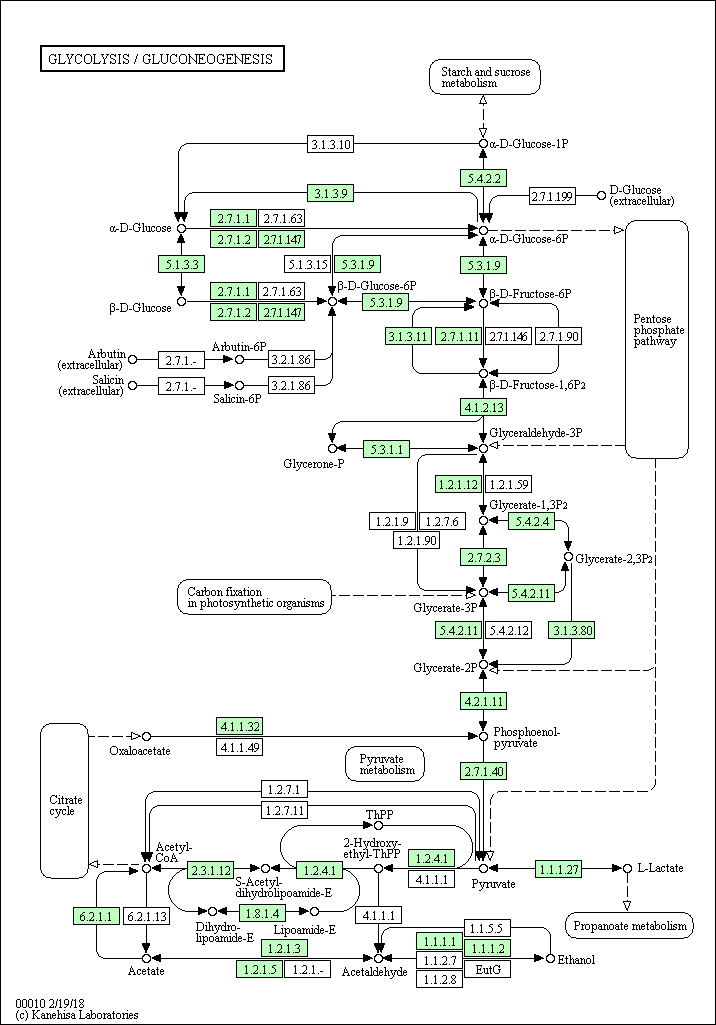
Recon3D model not only presents the metabolic data, but also describes the mutual relationship by indicating the constraints and regulations. The relationship between metabolites are reflected by participating in reactions. The relationship between genes and metabolites is not directly clarified, however, reactions are regulated by gene reaction rules, thus an indirect relationship between genes and metabolites are constructed, and it can be considered as an one-to-many relation.

Figure 3.1 Human glycolysis pathways indicates metabolite-reaction relationship on KEGG

Mutual-related correlations is derived for users to deal with actual needs. When an user is interested in a reaction, ***Xenobiotic transport***(id: 24NPHte), for instance, there is a large chance that the user may also interested in the one or more metabolites that involved in this reaction. Once the mutually connected relationship is established, different compartments will be linked and greatly save users time by referencing.

### Record operations history

Normally, the administration page is run and maintained by more than one administrators. Sometimes, different administrators are responsible to different compartments. Sometimes, they just shift duties on a regular basis. The cooperation of administrators makes the backend of a certain website run smoothly, and the operation history record is the guarantee to the flawless cooperation. Basically, the history records described all the operations done by each administrators and detail information, including the specific type(add, delete or modify, etc), operators and at which time.

The operation history record avoids potential conflicts and disputes. Besides, it also functions as a reliable backup for administrators, in case of carelessness. For example, if an administrator mistakenly modified a metabolic data, the detail of this operation will be recorded and the administrator who in charge of inspecting the history can easily recover the modified data to original value according to the record.

### Record user feedbacks

It is inevitable that multiple problems, technical issues or questions may occur during user-web interactions. In my project, reports of these issues are all claimed as ‘feedbacks’. Collecting user feedbacks is beneficial to improve the overall performance of project. Developers could fix bugs and system vulnerabilities according to user feedbacks. Enhance user experience is another motivation, users may regard the website design from distinct perspectives and thus may give constructive suggestions. Furthermore, users are not always familiar with the model and usage, so that detailed instruction should be delivered to users on request. Besides, there are still some other situations, like related institutes want to send academic cooperation inquires, functions provided can no longer meet users’ need and so on.

A generic form of feedback should contain at least four elements: name, contact information, title and main context. The corresponding part should also be developed in front-end, which will be discussed in following sections, and here we just need to aware that a table should be created to accept feedback data insertion and backend will provide interface to the frontend.

## User page design and intended functions

The development of user page is the core of whole project. Administration page is based on management, however, user page emphasized on application. Apart from the all the function mentioned in last section, user page will extend the scope of function even larger.

Generally speaking, user page is open access to public and no authentication process is needed. User page and administration page are seemingly two separated part, but actually they are not independent. In simple terms, front-end reflects how to present data, back-end decides what data to be presented.

There is no doubt that this page is user-oriented, consequently, the realization of functions is no longer the sole goal, appearance and visual effects are also crucial elements that cannot be neglected. It is worth mentioning that user page is not the simple combination of all the functions, but also the optimizing match based on user preference.

### Homepage and interfaces set

Homepage is the first page of a website displayed through search engine, and thus can be considered as the ‘face’ if we regard the whole website framework as ‘body’. Homepage is normally functions as the media for users to access other compartment. When combined the URL of homepage and extensions, users are able to reach out all the pages defined within the project. In most cases, visit webpage by manually type or combine the URL in search engine is never the best practice, and that is why interfaces set are to be created in homepage.

In my project, all interfaces to external links should be integrated in home page, for example, BiGG database, KEGG database, the homepage of Northeastern University and Eindhoven University of Technology. Besides, most internal links should also embedded in homepage. First, homepage should provide access to metabolic data compartments compartment. Second, feedback module should be presented in homepage. Last but not least, homepage is also the most direct way for users to learn about the website, therefore, some basic information including the introduction to project, model and team members are to be displayed as well.

### Metabolic data compartments

### Global search

### Online analysis

## Database design

# ：网站实现过程 Chapter IV: Website Realization Process

# ：分析与总结 Chapter V: Analysis and Conclusion

[参考文献](file:////Users/mihaoyang/Downloads/20080512515147/模版/自动化本科毕业设计格式说明、模板及示例v2.0/桌面/论文模板/文献格式(篇数大于30篇且含近三年五篇文献)%201专著格式:序号.编著者.书名【M】,出版地:出版社,年代,起止页码%202期刊论文格式:序号.作者.论文名称【J】,期刊名称,年度,卷（期）:起止页码%203学位论文格式:序号.作者.学位论文名称【D】,发表地:学位授予单位,年度%20例:1.张艺.铸造工艺【M】,北京:机械工业出版社,1994,4-5%202.张颖伟.一类大组合系统的容错控制【J】,东北大学学报,2000,52(4):351-355%203.周丽.挖掘机的优化【D】:沈阳:沈阳东北大学:2000)References

1. Hodgkin A L, Huxley A F. A quantitative description of membrane current and its application to conduction and excitation in nerve[J]. The Journal of physiology, 1952, 117(4): 500-544.
2. NOBLE D. Cardiac action and pacemaker potentials based on the Hodgkin-Huxley equations [J]. Nature, 1960, 188(4749): 495.
3. Watson J D. The human genome project: past, present, and future[J]. Science, 1990, 248(4951): 44-49.
4. Schilling C H, Schuster S, Palsson B O, et al. Metabolic pathway analysis: basic concepts and scientific applications in the post‐genomic era[J]. Biotechnology progress, 1999, 15(3): 296-303.

致谢Acknowledgements