



# Biopedia

Online Bioinformatics Data Center

Developed By **Nerdy Buddy**

# Biopedia - 需求文档

Nerdy Buddy

组长：宋佳铭

组员：袁源，肖迪，杨煜，滕爽

2014 年 12 月 3 日

## 目录

<b>1 概述</b>	<b>4</b>
1.1 项目背景	4
1.1.1 数据来源	4
1.1.2 来源数据处理程度	4
1.2 项目整体目标	4
<b>2 功能需求</b>	<b>5</b>
2.1 用户系统需求	5
2.1.1 用户注册	5
2.1.2 用户登录	7
2.1.3 用户登出	8
2.1.4 查看用户资料	8
2.2 展示功能需求	9
2.2.1 展示数据	9
2.2.2 展示内容	9
2.2.3 主页导航和模块	12
2.2.4 项目查询	13
2.3 检索功能需求	15
2.3.1 概述	15
2.3.2 按照项目属性进行筛选的功能	15
2.3.3 按照项目名字进行筛选	15
2.4 管理功能需求	15
2.4.1 概述	15
2.4.2 用户分类	18
2.5 自定义工作流的数据分析功能	20
2.5.1 概述	20
2.5.2 功能	20
2.6 数据库实现	20

---

<b>3</b>	<b>非功能需求</b>	<b>21</b>
3.1	良好的导航风格 . . . . .	21
3.2	友好的用户界面 . . . . .	21
3.3	兼容性 . . . . .	21
3.4	良好的代码风格 . . . . .	21
3.5	性能需求 . . . . .	21
<b>4</b>	<b>实现需求</b>	<b>21</b>
4.1	软件体系结构 . . . . .	21
4.2	基本框架 . . . . .	22
4.3	网站部署环境 . . . . .	22
4.4	软件开发环境 . . . . .	22
4.5	项目管理需求 . . . . .	22
<b>5</b>	<b>文档需求</b>	<b>23</b>
5.1	设计文档 . . . . .	23
5.2	使用说明 . . . . .	23
5.3	测试文档 . . . . .	23

## 1 概述

### 1.1 项目背景

新技术的发展,特别是DNA测序技术的突破,带动了生命科学研究的突飞猛进。

在生命科学以及相关领域如医学研究,药物研发,临床诊断,生态环境研究,农林渔牧等行业每天都有大量的各类数据产生,使生命科学的研究越来越依赖于数据分析。基因组学的研究就是对测序数据和其他各类生物医学数据进行分析、融合、理解,从中发现生命的基本规律。

虽然生物信息领域的基因组学数据量越来越大,但是国内尚未有统一展示生物信息数据及分析结果的平台。因此,我们将建设一个能基于已经采集的口腔微生物宏基因组测序数据库,以及已经得到的分析结果数据库,建立开放的生物信息宏基因组数据展示平台。

#### 1.1.1 数据来源

本平台的主要数据来源为舌苔样本数据库。信息科学与技术国家实验室收集了189个人的舌苔样本以及每一个人的健康状况指数(例如是否患有胃炎和胃病等)。

#### 1.1.2 来源数据处理程度

样本的宏基因组的测序: 已经完成。

数据的生物信息学分析结果: 已经完成。

将会按照特定格式存入数据库,提供给开发小组。

### 1.2 项目整体目标

- 开发一个WEB平台, 包括用户系统、项目系统、管理员系统。
- 在该平台上实现项目数据的展示功能、查询功能、检索功能、管理功能、 workflow 功能这五大主要功能。
- 前端美观, 交互界面用户友好。
- 后端稳定, 代码风格良好, 易于拓展和维护。

- 在已有功能的基础上，添加添加并查看“用户最喜欢项目”、查看“用户创建的项目”等创新、人性化的附加功能。

## 2 功能需求

### 2.1 用户系统需求

#### 2.1.1 用户注册

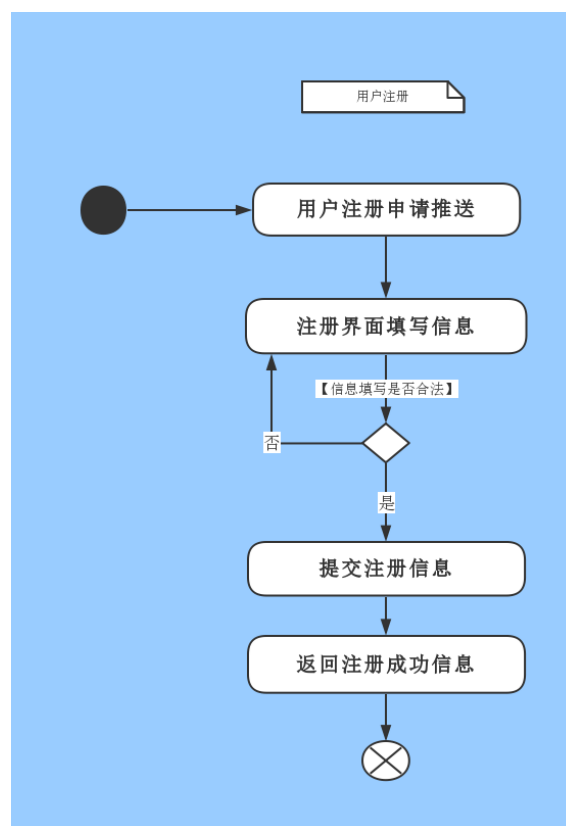


图 1: 用户注册

活动图 见图1

**注册页面填写信息** 用户需要填写的注册信息以表格形式列出

**提交注册信息** 将填写合法的用户信息写入数据库。当信息填写错误的时候返回错误信息：  
如电子邮件已经被注册、密码长度不符合要求、注册时两次密码不匹配等错误信息。

数据元素	描述	格式
Username	用户登录所需要的用户名	字符串，长度不超过20字符
Password	用户名对应的密码	字符串，长度不超过30字符
Email	用户的电子邮件	字符串，长度不超过30字符
Firstname	用户的名	字符串，长度不超过20
Lastname	用户的姓	字符串，长度不超过20

表 1: 用户填写的注册信息

**返回注册成功信息** 注册成功返回注册成功信息界面。

### 2.1.2 用户登录

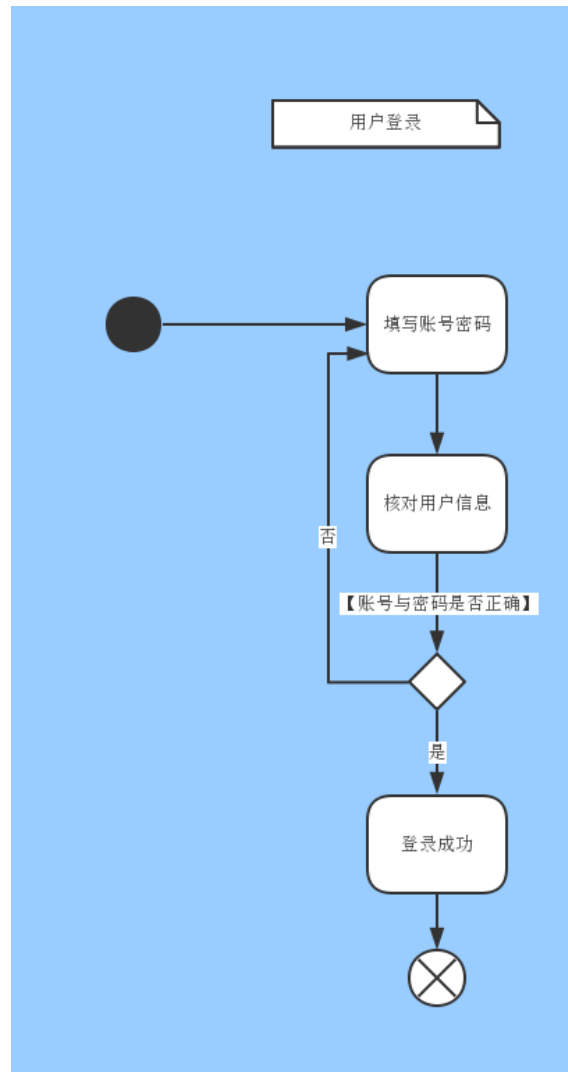


图 2: 用户登录

活动图 见2

**填写账号和密码** 填写用户ID与相应密码，按确认键将信息提交

**核对用户信息** 将提交的信息与数据库中的用户信息进行比对，确认账号与密码是否正确。

**登录成功** 用户进入系统，进行下一步操作



### 2.1.3 用户登出

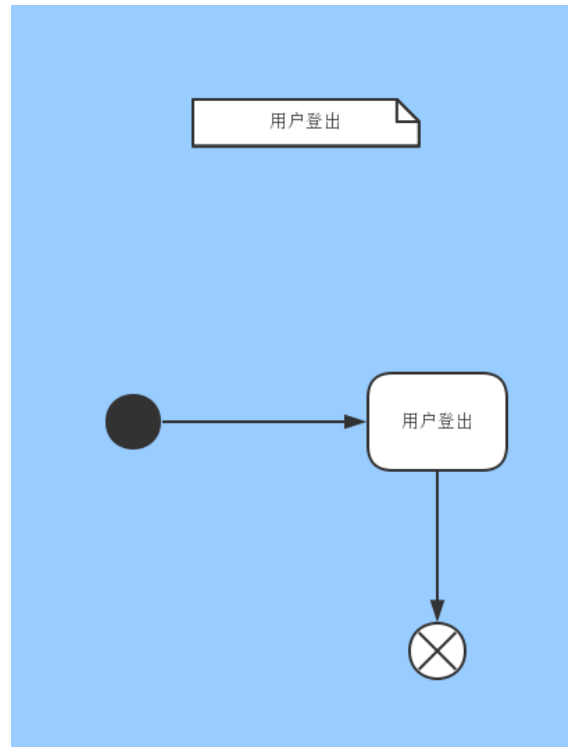


图 3: 用户登出

活动图 见3

### 2.1.4 查看用户资料

用户可以通过点击相应位置查看用户资料。

## 2.2 展示功能需求

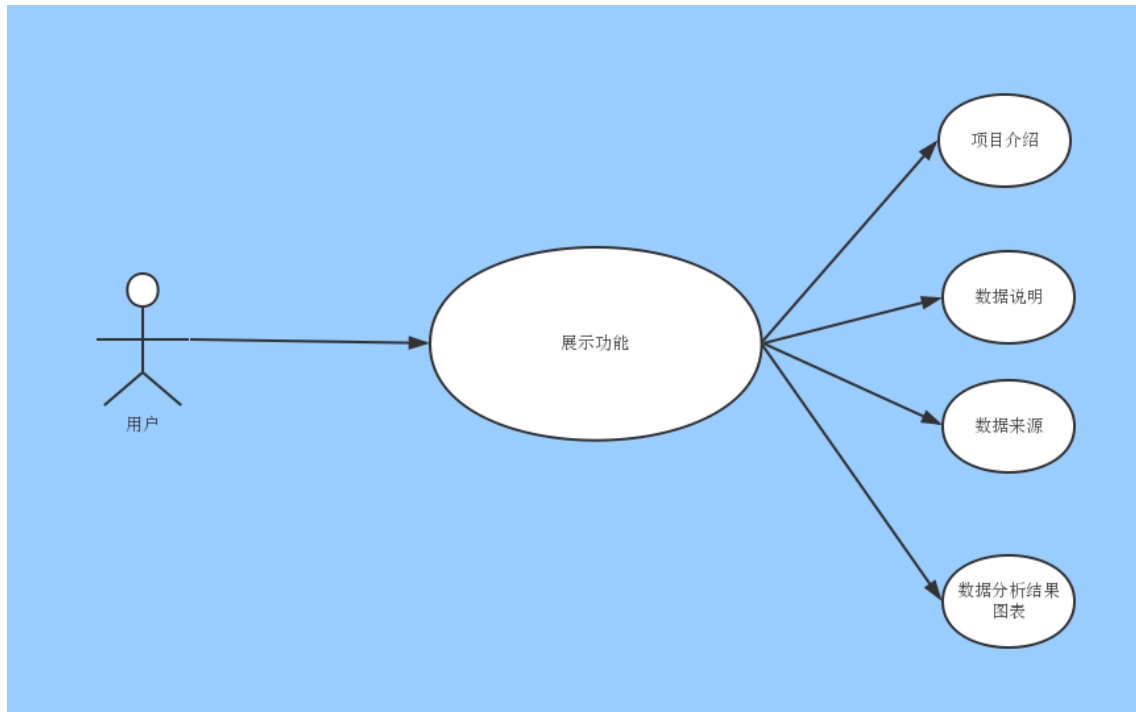


图 4: 展示基本功能

### 2.2.1 展示数据

以口腔宏基因组为例的project及样本数据展示网页。

### 2.2.2 展示内容

展示内容分为四大方面：包括项目介绍，数据说明，数据来源，数据分析结果图标等。

**项目介绍** 对需要进行数据展示的项目进行介绍，用户可以在这里添加所有关于该项目的描述。

**数据说明** 对项目中展示的数据本身进行说明，包括数据的名称、属性等。

**数据来源** 对项目中展示的数据的来源进行说明，有助于用户对于数据的深入理解。

**数据分析结果图表等** 我们需要对数据用图表形式进行数据分析结果的展示，共需要六种图表：

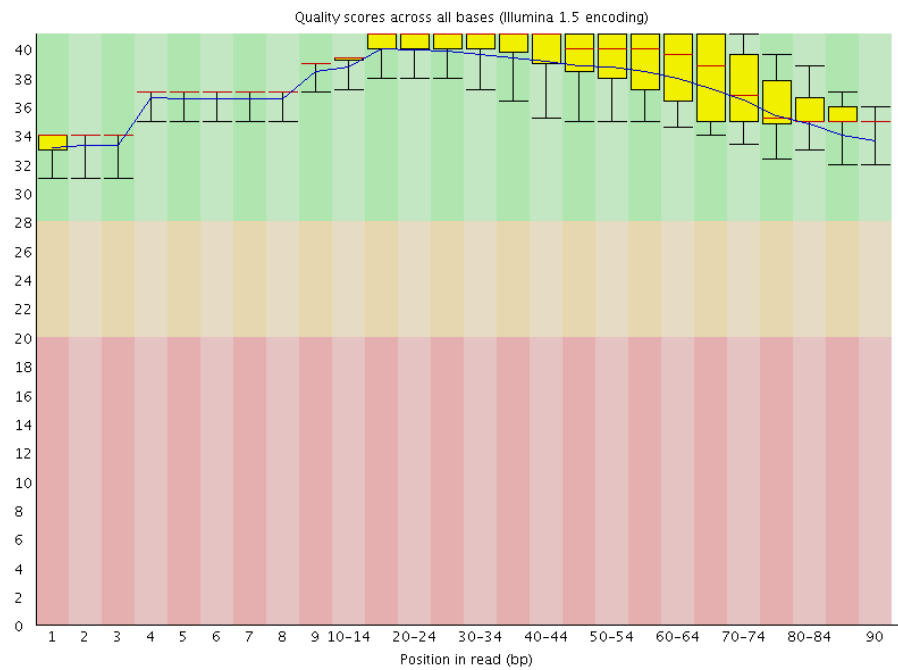


图 5: 第一类图表

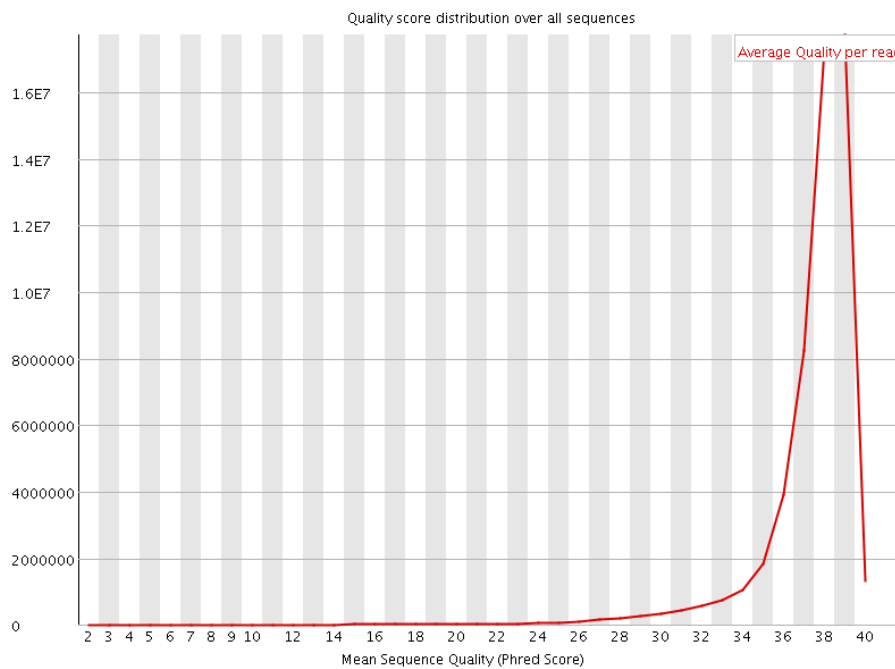


图 6: 第二类图表

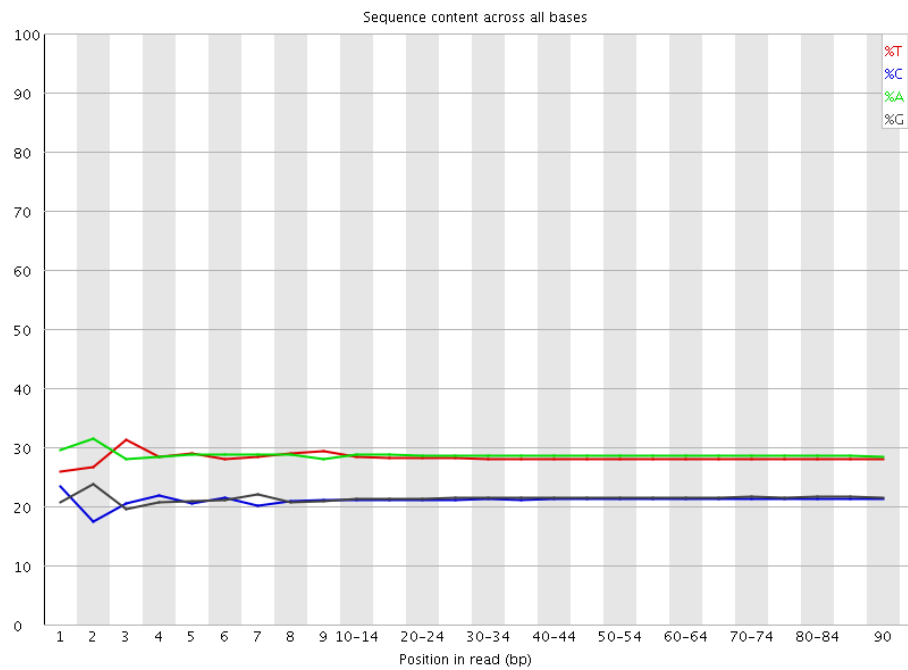


图 7: 第三类图表

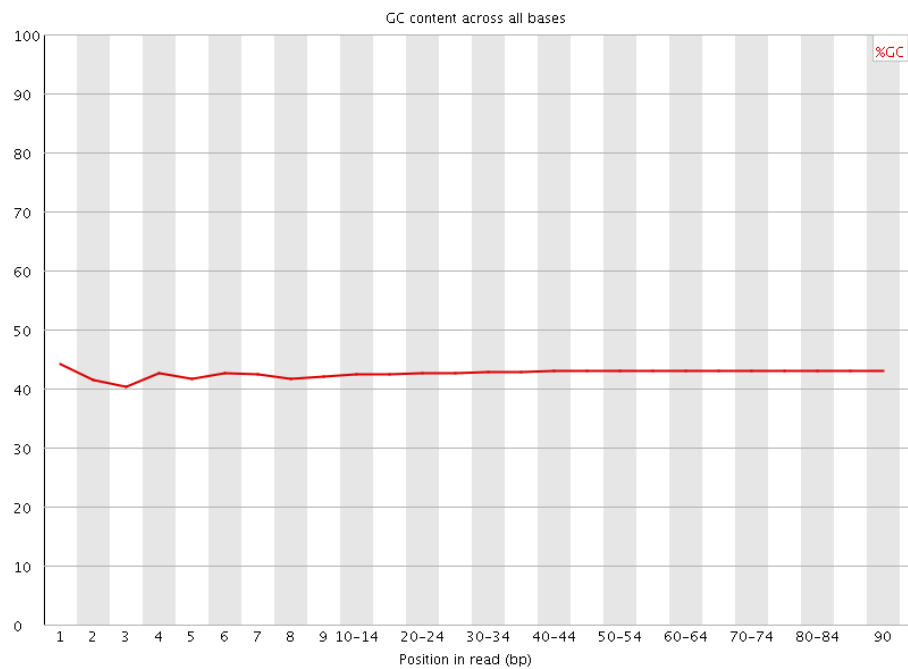


图 8: 第四类图表

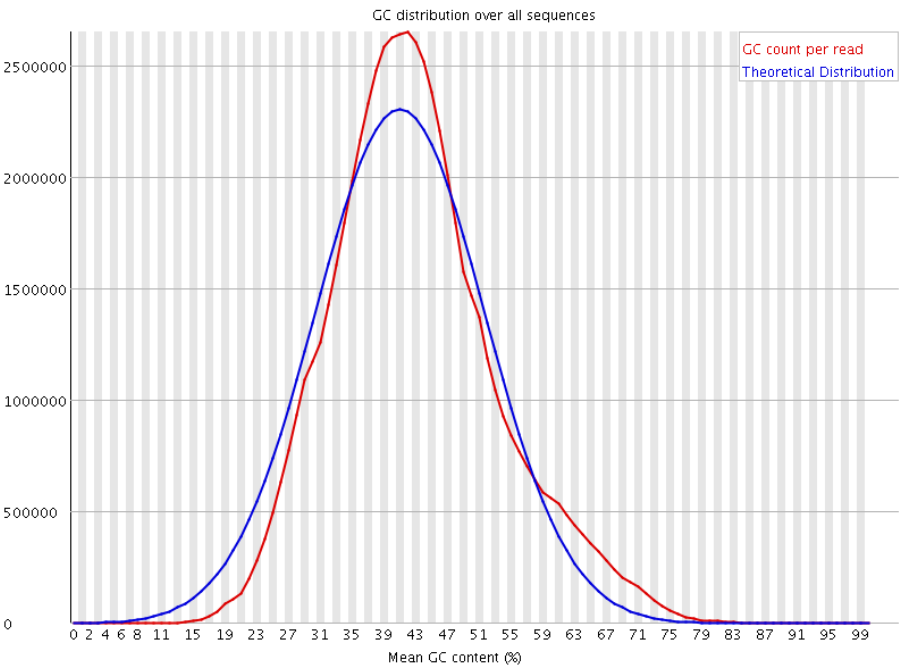


图 9: 第五类图表

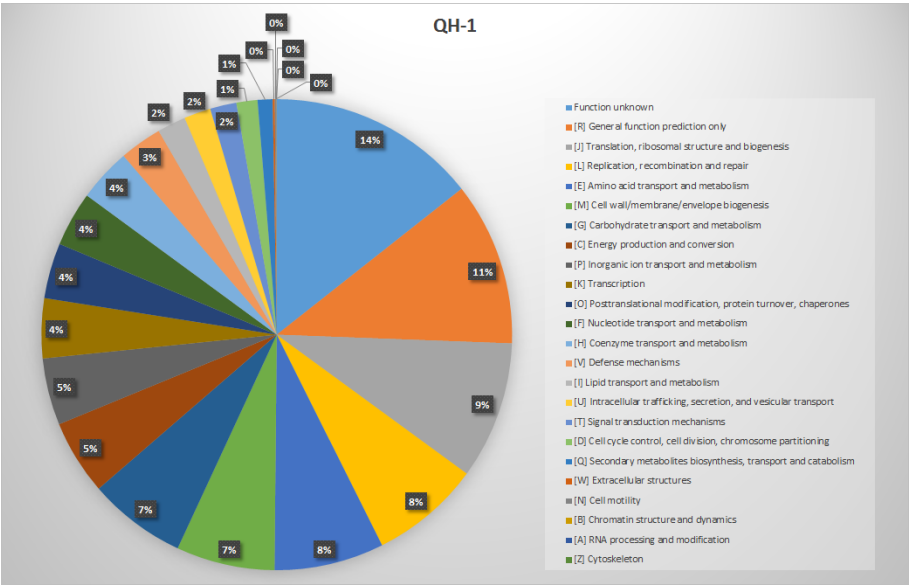


图 10: 第六类图表

2.2.3 主页导航和模块

我们需要在网页中呈现一个简洁清晰、展示信息高效的导航栏。在导航栏中集成了

“主页”、“项目”以及“登录”、“注册”、“显示用户名”等模块。处于项目，数据页面时，导航栏会显示出当前页面的路径，通过点击导航栏的相应位置，自由返回不同上级页面。

#### 2.2.4 项目查询

在这个页面，我们可以对项目的基本信息和详情进行查询。

**项目基本信息** 我们将在项目基本信息中展示以下属性：

Projects, Environment, Site, Sequence of Type, Project ID, Number of Total Sequences, Number of ORFs, Number of Samples, Read Length, Platform, Create Date, Update Date等。

Meta Genome Projects											
Projects	Environment (Water/Air)	Site	Sequence Type	Project ID	# of Total Sequences	# ORFs	# Samples	Read Length	Platform	Create Date	Update Date
	All ▾	All ▾	All ▾	All ▾	All ▾	All ▾	All ▾	All ▾	All ▾	All ▾	All ▾
MetaTongue	Human	Oral	WGS	MGP-0001					Illumina Hiseq 2000	2014/08/12	
MetaIntestinal	Water			MGP-0002							
Meta-LakeTai				MGP-0003							
Meta...				MGP-0004							

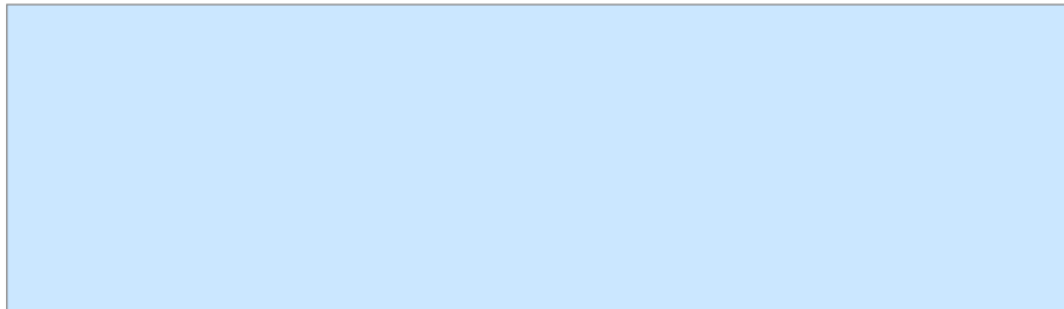
图 11: 项目属性展示

点击项目名字进入项目详情,

**项目详情** 我们将在这个页面，分成项目简介、项目数据、鸣谢、联系方式等部分对项目展示项目详情，并可以从该页面进入数据展示页面查看数据展示的图表。

**Project Overview****Meta Genome Projects**

Name: MetaTongue  
Project ID: MGP-0001  
Create Date: 2014.06.01  
Update Date: 2014.08.20

**Introduction****About Data**

Meta Genomes Data <span>Download</span>											
MDB - ID	Sample ID	Environment	Site	Sequence Type	Platform	Read Length	# of Raw Reads	# of Clean Reads	#Scaffold	#ORFs	Meta Data
		All ▾	All ▾	All ▾	All ▾	All ▾	All ▾	All ▾	All ▾	All ▾	All ▾
MDB_S_001	D087						60,000,000		50000	20000	

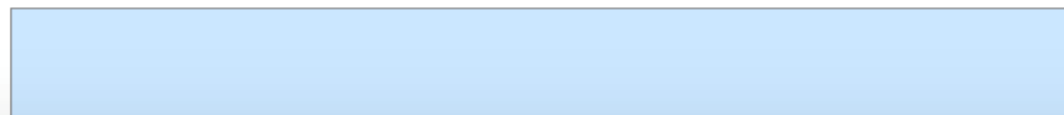
**Acknowledgement**

图 12: 预期展示效果

项目详情需要展示每个样本的统计图表 对project中每一个测序样本(about data里的每一行结果),点击ID之后进入的测序结果,用表格展示其测序质量信息;用曲线图、柱状图或饼状图展示DNA片段拼装、聚类后的分析结果。

## 2.3 检索功能需求

### 2.3.1 概述

我们将在数据检索页面实现两个主要功能，第一个是实现根据项目属性对所有项目的检索功能，第二个是实现根据所有样本属性对样本的检索功能。

### 2.3.2 按照项目属性进行筛选的功能

Meta Genome Projects

Projects	Environment (Water/Air)	Site	Sequence Type	Project ID	# of Total Sequences	# ORFs	# Samples	Read Length	Platform	Create Date	Update Date
	Water ▾	All ▾	All ▾	All ▾	> ▾	> ▾	All ▾	All ▾	All ▾	All ▾	All ▾
					5000	5000					

图 13: 项目筛选功能展示

**非数量型(枚举型)属性按属性名称筛选** 例如Environment属性是枚举型,可选择Water, Air等。

我们可以在筛选信息中选择Water或者Air来筛选对应各自环境的项目。

**数量型属性按数值区间筛选** 例如Number of ORFs 属性是数值型。我们可以根据用户输入的阈值,选择<= 或者>=, 来筛选出符合要求的项目。

**支持多项筛选规则同时作用,并可以隐藏/显示属性列** 用户可以同时选择多项筛选规则进行筛选, 并可以根据需求, 选择隐藏/显示属性列。

### 2.3.3 按照项目名字进行筛选

## 2.4 管理功能需求

### 2.4.1 概述

实现用户信息管理功能,部分功能对所有访问人员可用,部分功能(例如下载功能)仅对登陆人员可用。用户类型表(见表2)及权限示例如下:



用户类型	下载/上传数据	用户账户增删改	后台数据库处理
匿名用户			
普通用户			
管理员用户			

表 2: 用户权限表

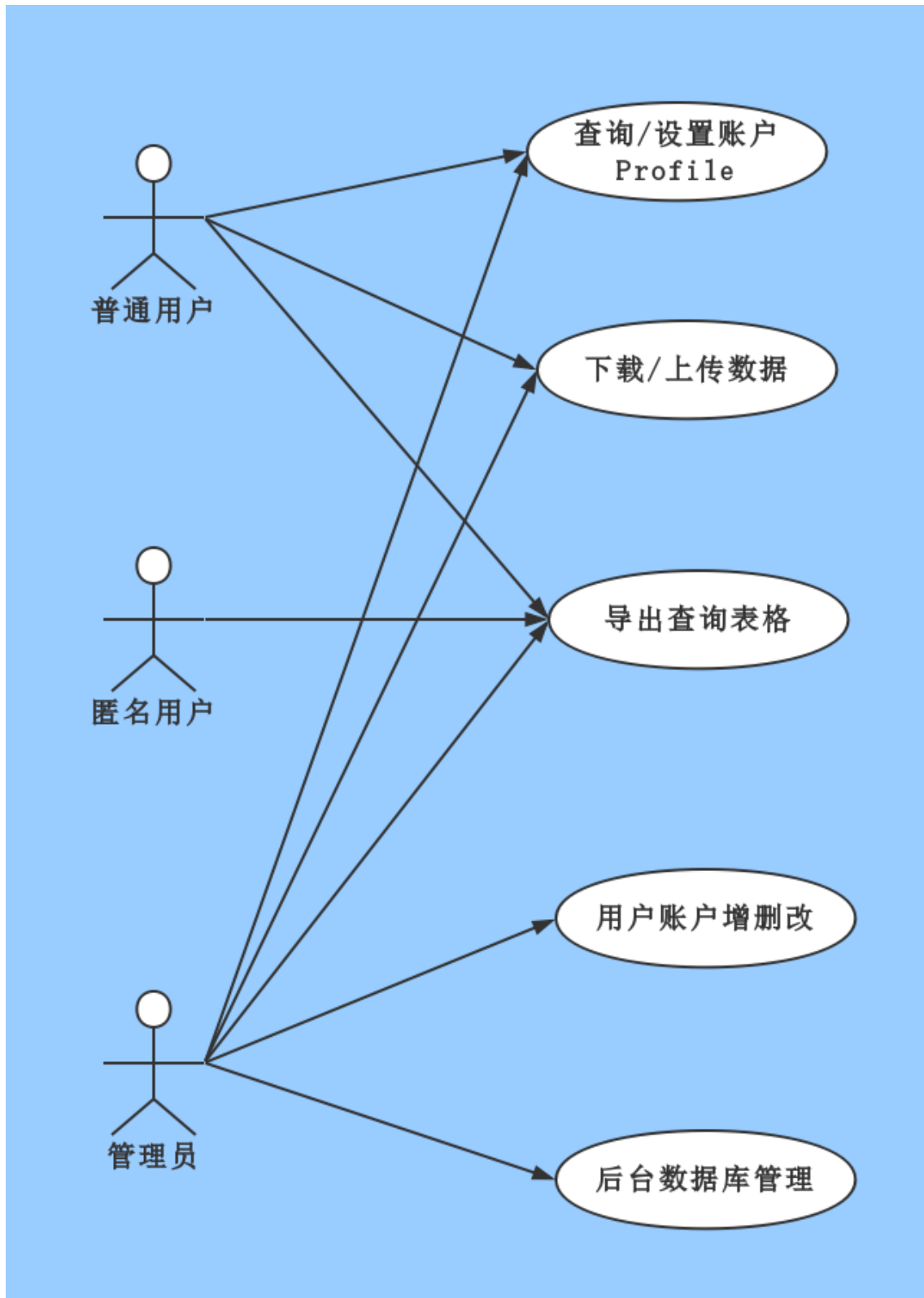


图 14: 用户权限示例

### 2.4.2 用户分类

**匿名用户** 匿名用户能在相关页面进行导出查询表格的操作。

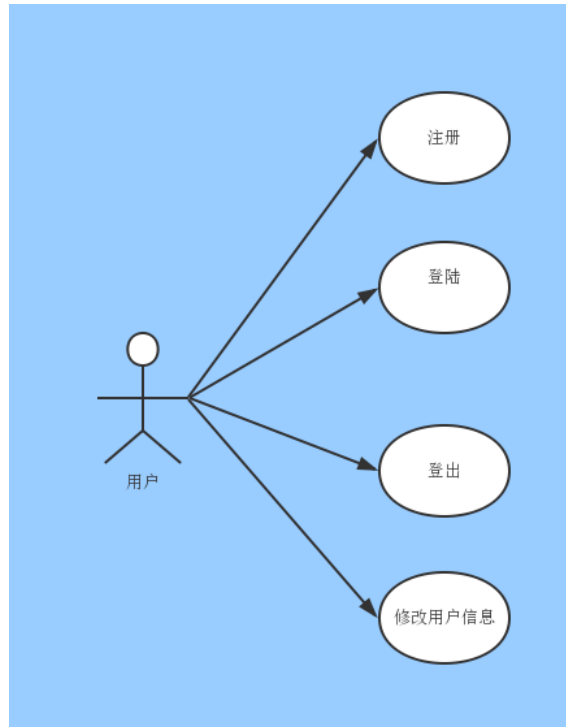


图 15: 普通用户基本权限

**普通用户** 普通用户能在相关页面进行查询/设置账户Profile、下载/上传数据、导出查询表格的操作。（见图??和16）

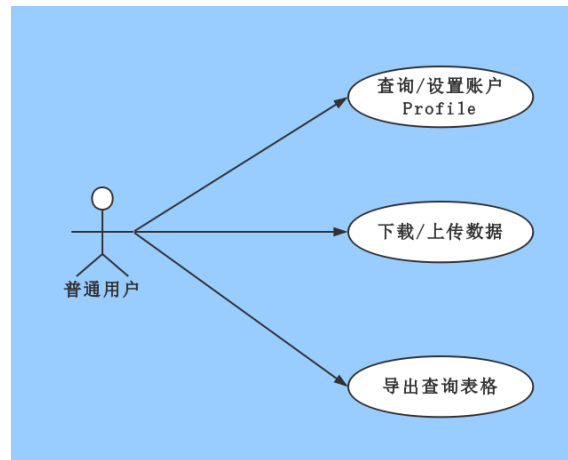


图 16: 普通用户数据权限

**管理员用户** 管理员能在相关页面进行查询/设置账户Profile、下载/上传数据、导出查询表格、用户账户增删改、后台数据库管理的操作。（见图17）

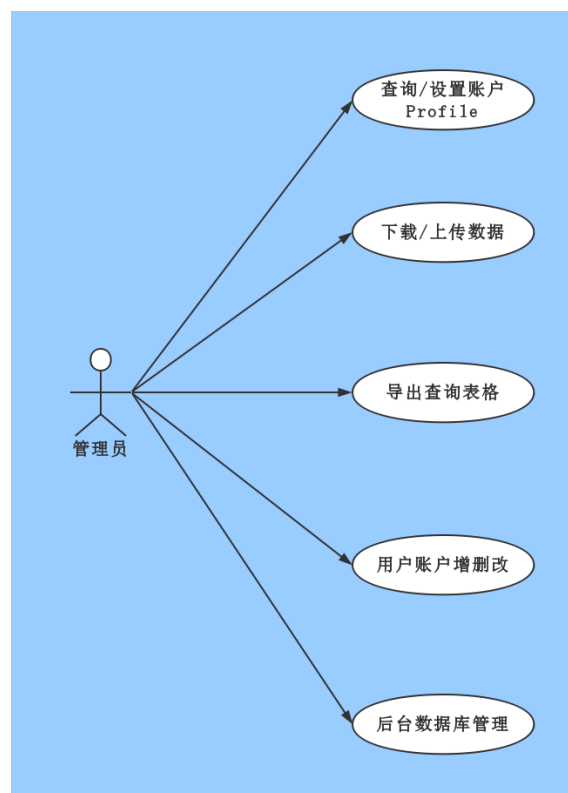


图 17: 管理员用户额外权限

## 2.5 自定义工作流的数据分析功能

### 2.5.1 概述

用户能自定义工作流，进行数据分析。

### 2.5.2 功能

**数据处理功能** 允许用户上传自己的数据，选择工作流的工作模块,提交数据进行分析,并展示和记录分析结果。

**自定义工作模块** 实现几个简单算法工作模块(如向量归一化、向量乘、向量加、向量点积等),根据用户给定的初始向量,实现用户自定义工作流,提交任务后,输出结果。用户也能非常方便地自己添加新工作模块。

**前端实现** 用户能上传数据，并通过拖拽来规定工作模块的执行顺序。执行完后能生成数据文件并下载。

**后端实现** 后端代码风格良好，用户能较为简单添加或者删除工作模块。

## 2.6 数据库实现

**后台数据库的可视化管理界面** 后台数据库实现无代码化，拥有可视化管理界面

**数据导入功能** 支持通过文件的形式的数据导入功能。

**数据导出功能** 支持数据通过json文件的形式的导出功能。

**数据修改功能** 支持对已导入数据进行修改的功能（或者重新上传）。

**原始数据文件的下载功能** 能在制定界面进行原始数据文件下载。

**界面语言可选（中英文切换）** 在所有界面都能进行中英文切换。

**访问计数器** 在页面添加访问计数器功能，可使网站访客直观地看到有本站有多少访客。

## 3 非功能需求

### 3.1 良好的导航风格

网页具有良好的导航风格，用户能非常简单清晰地到达想要到达的界面。

### 3.2 友好的用户界面

用户界面友好，交互方便，网站布局紧跟时代潮流。

### 3.3 兼容性

支持ie9/firefox/chrome/safari

### 3.4 良好的代码风格

使用Flask的Web应用框架，辅以Jinja2 优化代码风格，使得后端代码简洁美观、易于维护。

### 3.5 性能需求

完成平台可以稳定运行，数据量较大时性能稳定。

## 4 实现需求

### 4.1 软件体系结构

使用MVC(Model-View-Controller)软件体系结构。

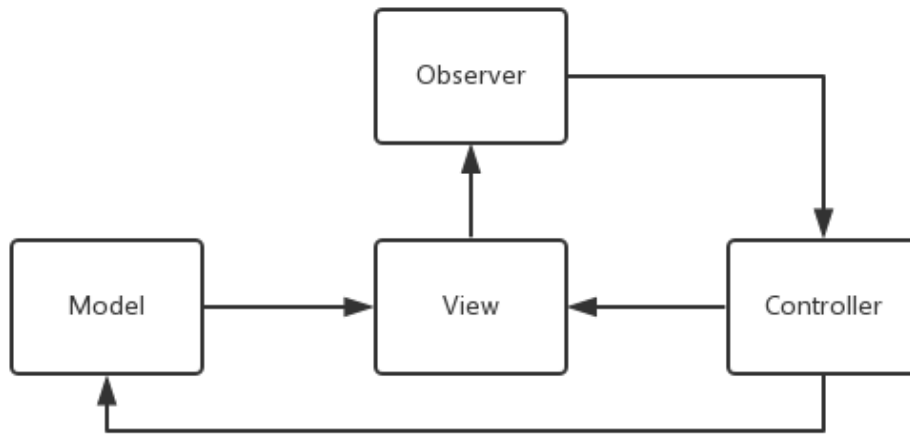


图 18: MVC概念图

## 4.2 基本框架

使用Flask的Web应用框架。

## 4.3 网站部署环境

运行于清华大学生物信息学部的高效服务器上,典型linux环境(Dell PowerEdge R270xd 的服务器,4核64GB内存,Ubuntu 12.04)

## 4.4 软件开发环境

采用Python/MongoDB框架。使用chart.js, d3.js, highcharts.js等画图软件进行图形绘制。

## 4.5 项目管理需求

使用JIRA进行项目管理,使用Stash进行源代码管理。

## 5 文档需求

所有文档使用CTex进行排版。UML图统一使用ProcessOn进行编辑。

### 5.1 设计文档

包括以下几个模块:

1. 体系结构设计。要求采用模块图系统结构，采用UML序列图描述系统行为。
2. 模块之间的主要接口设计
3. 数据库或数据存储结构设计
4. 设计模式。说明主要采用的设计模式及其设计实现方式。

### 5.2 使用说明

应该包括分别给用户和管理员的使用说明。

### 5.3 测试文档

应该包括以下几个模块:

1. 针对特定选取的模块，采用xUnit框架，实现单元测试，形成测试覆盖率分析报告。
2. 根据项目特定的要求，完成系统测试及性能测试，并形成测试报告。



## 图片列表

1	用户注册 . . . . .	5
2	用户登录 . . . . .	7
3	用户登出 . . . . .	8
4	展示基本功能 . . . . .	9
5	第一类图表 . . . . .	10
6	第二类图表 . . . . .	10
7	第三类图表 . . . . .	11
8	第四类图表 . . . . .	11
9	第五类图表 . . . . .	12
10	第六类图表 . . . . .	12
11	项目属性展示 . . . . .	13
12	预期展示效果 . . . . .	14
13	项目筛选功能展示 . . . . .	15
14	用户权限示例 . . . . .	17
15	普通用户基本权限 . . . . .	18
16	普通用户数据权限 . . . . .	19
17	管理员用户额外权限 . . . . .	19
18	MVC概念图 . . . . .	22

## 图表列表

1	用户填写的注册信息 . . . . .	6
2	用户权限表 . . . . .	16