**《数据库实训》 课程报告（第一组）**

**(2020-2021 学年第2 学期)**

**题目 投资组合评比器**

**提交日期： 2021 年 06 月 17 日**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **学 院** | **软件** | | **专业班级** | **软件工程** | |
| **课程名称** | **数据库实训** | | **任课教师** | **曾兵** | |
| **人数** | **3** | |  |  | |
| **学 号** | | **学生姓名** | | | **分 数** |
| **201936380215** | | **陈卓文** | | |  |
| **201930383212** | | **熊腾浩** | | |  |
| **201930400391** | | **庄昊敏** | | |  |
|  | |  | | |  |
|  | |  | | |  |
|  | |  | | |  |
|  | |  | | |  |
| **教师评语：**  **教师签名：** | | | | | |

目录

[需求分析 4](#_Toc74939863)

[概要设计 4](#_Toc74939864)

[E-R 模型 4](#_Toc74939865)

[总体框架图 4](#_Toc74939866)

[数据获取 5](#_Toc74939867)

[投资组合评比 5](#_Toc74939868)

[可视化 5](#_Toc74939869)

[数据源可扩展 5](#_Toc74939870)

[详细设计 6](#_Toc74939871)

[后端模块 6](#_Toc74939872)

[数据库接口 6](#_Toc74939873)

[基金网站爬虫模块 7](#_Toc74939874)

[前端模块 8](#_Toc74939875)

[基金展示模块 8](#_Toc74939876)

[基金网站录入模块 9](#_Toc74939877)

[基金 url 管理模块 9](#_Toc74939878)

[爬虫脚本管理模块 10](#_Toc74939879)

[软件测试 11](#_Toc74939880)

[爬虫测试 11](#_Toc74939881)

[可视化测试 13](#_Toc74939882)

[数据排序测试 15](#_Toc74939883)

[扩展数据源测试 16](#_Toc74939884)

[软件演示 18](#_Toc74939885)

[软件使用教程 21](#_Toc74939886)

[安装教程 21](#_Toc74939887)

[window下安装 21](#_Toc74939888)

[从源码构建 21](#_Toc74939889)

[软件功能使用教程 21](#_Toc74939890)

[功能一：爬取数据 21](#_Toc74939891)

[功能二：可视化基金投资组合 22](#_Toc74939892)

[功能三：按照某项指标对各个组合进行评比 24](#_Toc74939893)

[功能四：增加其他数据源 24](#_Toc74939894)

[小组讨论纪要 30](#_Toc74939895)

[第一次讨论 30](#_Toc74939896)

[第二次讨论 30](#_Toc74939897)

[第三次讨论 30](#_Toc74939898)

[第四次讨论 30](#_Toc74939899)

[第五次讨论 31](#_Toc74939900)

[第六次讨论 31](#_Toc74939901)

[个人总结 32](#_Toc74939902)

[陈卓文 32](#_Toc74939903)

[熊腾浩 32](#_Toc74939904)

[庄昊敏 32](#_Toc74939905)

[项目进度安排 33](#_Toc74939906)

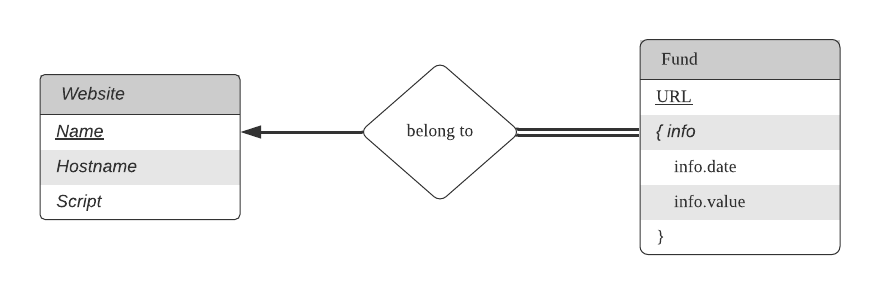
# 需求分析

本次实训任务需求如下：

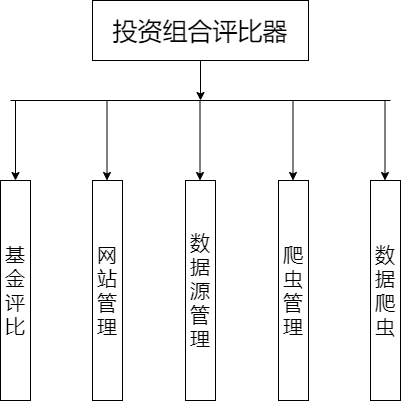
1. 爬取且慢和蛋卷基金上的投资组合数据
2. 把投资组合放在一起进行对比
3. 图形化展示，可以由用户选择多个投资组合, 放在同一页面展示
4. 数据源可扩展，可以添加且慢和蛋卷平台新的数据源和其他平台的数据源

# 概要设计

## E-R 模型



## 总体框架图



## 数据获取

可以爬取任务要求中指定的数据源（且慢和蛋卷基金网站上的指定投资组合）

## 投资组合评比

可以任意指定时间段，并基于起始时间计算基金收益

可以对比不同的指标，包括涨幅/年化收益率, 最大回撤, 夏普比, 年化波动率

可以根据不同的指标对基金组合进行多字段排序

## 可视化

在同一视图内展示需对比的基金的收益曲线

可以暂时屏蔽部分基金的收益曲线

可以对选定基金收益曲线进行高亮展示

可视化网站、数据来源的增删查改

可视化爬虫脚本编辑

## 数据源可扩展

可以添加蛋卷、且慢的其它数据源

可以通过添加自定义脚本来增加对其他网站的支持

# 详细设计

## 后端模块

### 数据库接口

#### 功能概述

提供便捷的接口与SQLite3操作

#### 关键逻辑与算法说明

使用nodejs模块better-sqlite3，操作SQLite文件，预编译SQL语句并执行

#### 访问的表或其他数据库实体

website 基金网站名称、域名、爬虫脚本

url 包含基金地址，hostname 和数据最后更新时间

fund 每个基金的每日收益

#### 提供调用接口说明

preload.ts 将前端所需的所有后端API暴露至全局变量

api\_website.select\_all() 返回所有基金网站

api\_website.select\_host(hostname: string) 返回指定 hostname 的基金网站名称

api\_website.insert(desc: WebsiteDesc) 插入输入的基金网站数据

api\_website.update(name: string, desc: WebsiteDesc) 用 desc 更新指定 name 的基金网站信息

api\_website.remove(name: string) 删除指定的基金网站

api\_website.get\_script\_name(name: string) 通过名字获取指定网站的爬虫脚本

api\_website.get\_script\_hostname(hostname: string) 通过 hostname 获取指定网站的爬虫脚本

api\_website.save\_script(name: string, code: string) 通过名字保存指定网站的爬虫脚本

api\_url.select\_all() 返回所有数据库中基金的 url

api\_url.select(url: string) 返回指定 url 的基金的 url 和 hostname

api\_url.insert(url: string) 检查数据库中是否存在指定的 url ，若不存在插入该 url

api\_url.remove(url: string) 删除指定 url

api\_fund.select(url: string) 返回指定 url 的数据

api\_fund.insert(url: string, data: FundInfo) 清空并插入指定 url 的数据

api\_fund.clear(url: string) 删除指定 url 的数据

api\_calc.calc(fund: FundInfo) 返回通过输入数据所计算得出的年化收益率，最大回撤，夏普率和年化波动率

#### 数据类型说明

DailyInfo: 包含两个 namber 类型数据，date 和 value

WebsiteDesc: 包含两个 string 类型数据，name 和 hostname

URLDesc: 包含两个 string 类型数据，url 和 hostname，和一个 number 类型数据，last\_update

FundInfo：DailyInfo 的数组

### 基金网站爬虫模块

#### 功能概述

爬取基金数据并进行处理：

1. 当不存在该基金所属网站的数据爬取脚本时

返回错误提示

2. 当存在该基金所属网站的数据爬取脚本时

对基金的数据进行爬取，并对数据库进行更新

#### 关键逻辑与算法说明

动态加载爬虫脚本，使用 Nodejs 的 VM 模块执行。

使用 axios 模块发起网络请求，获取网站基金数据。

受限于 electron 多进程设计以及 https 的跨站访问限制，需要将爬虫脚本的运行放在主进程中，但发起爬取请求的是渲染器进程，需要使用 electron 内置的进程间通讯发起请求与返回执行结果。

#### 访问的表或其他数据库实体

website 基金网站名称、域名、爬虫脚本

url 包含基金地址，hostname 和数据最后更新时间

fund 每个基金的每日收益

#### 提供调用接口说明

main-process-api.register() 将 exec\_script 函数注册至进程间通讯频道

main-process-api.exec\_script(url: string) 在主进程中创建 VM 执行爬虫脚本

api\_fund.exec(url: string) 向主进程发起指定 url 数据爬取请求

api\_fund.exec\_all() 向主进程发起所有存在脚本的 url 数据爬取请求

api\_fund.set\_exec\_callback(callback: (resp) => void) 设置爬取完成后执行的回调函数

api\_fund.rm\_exec\_callback() 移除回调函数

## 前端模块

### 基金展示模块

#### 功能概述

可视化展示所有基金的收益率：

1. 使用折线图可视化展示所有基金的收益曲线

2. 当用户鼠标覆盖某条基金曲线时，高亮其收益曲线

3. 当用户点击某基金时，隐藏/显示其收益曲线

4. 当用户在图表上选择时间段时，放大该时间段的收益曲线

5. 当用户在选择框中选择时间段时，将数据重置为该时间段的收益曲线，并使所有基金的起点相同

利用表格对比不同基金的数据：

1. 表格包含基金的年化收益率，最大回撤，夏普率和年化波动率

2. 当用户选择了一个或多个指标时，对图表进行数据排序

#### 关键逻辑与算法说明

根据所有获取的 url 从数据库中获取数据并画图

#### 访问的表或其他数据库实体

url 包含基金地址，hostname 和数据最后更新时间

Fund 各个 url 对应基金的每日收益

### 基金网站录入模块

#### 功能概述

对基金网站进行增删改查：

1. 页面展示已有的基金网站，包含 name 和 hostname

2. 用户点击删除，删除选定的基金网站

3. 用户点击增加，弹出表单，引导用户输入新的基金网站并保存

4. 用户点击修改，弹出表单，引导用户修改指定的基金网站并更新

#### 关键逻辑与算法说明

调用后台的 api\_website 相关模块网站进行增删改查

#### 访问的表或其他数据库实体

website 包含网站名称，hostname 和爬虫脚本

### 基金 url 管理模块

#### 功能概述

对指定基金 url 进行管理：

1. 页面展示已有的基金 url，包含 url 地址和数据最后爬取时间

2. 用户点击爬取全部，重新爬取全部基金的数据并更新

3. 用户点击新建 url，弹出表单引导用户输入 url，判断其是否合法并更新

4. 用户点击爬取图标，爬取指定基金的数据并更新

5. 用户点击删除数据图标，删除指定基金的数据

6. 用户点击删除基金图标，删除指定基金

发起数据爬取请求，显示数据爬取结果

#### 关键逻辑与算法说明

调用后台的 api\_fund，api\_url 相关基金进行增删改查，数据爬取

#### 访问的表或其他数据库实体

url 包含基金地址，hostname 和数据最后更新时间

fund 每个基金的每日收益

### 爬虫脚本管理模块

#### 功能概述

对爬虫脚本进行管理：

1. 页面默认展示模板爬虫脚本

2. 用户点击展示模板爬虫脚本图标，展示模板爬虫脚本

3. 用户点击选择修改的爬虫脚本图标，提供已有爬虫脚本列表

Ⅰ. 若当前网站已有脚本，在用户选定后展示该爬虫脚本

Ⅱ. 若当前网站暂无脚本，在用户选定后使用模板脚本初始化

4. 用户修改后点击保存图标，更新该爬虫脚本

#### 关键逻辑与算法说明

读取后台的爬虫脚本代码，并对其进行修改

#### 访问的表或其他数据库实体

website 包含网站名称，hostname 和爬虫脚本

# 软件测试

对软件进行黑盒测试, 它是通过测试来检测每个功能是否都能正常使用。

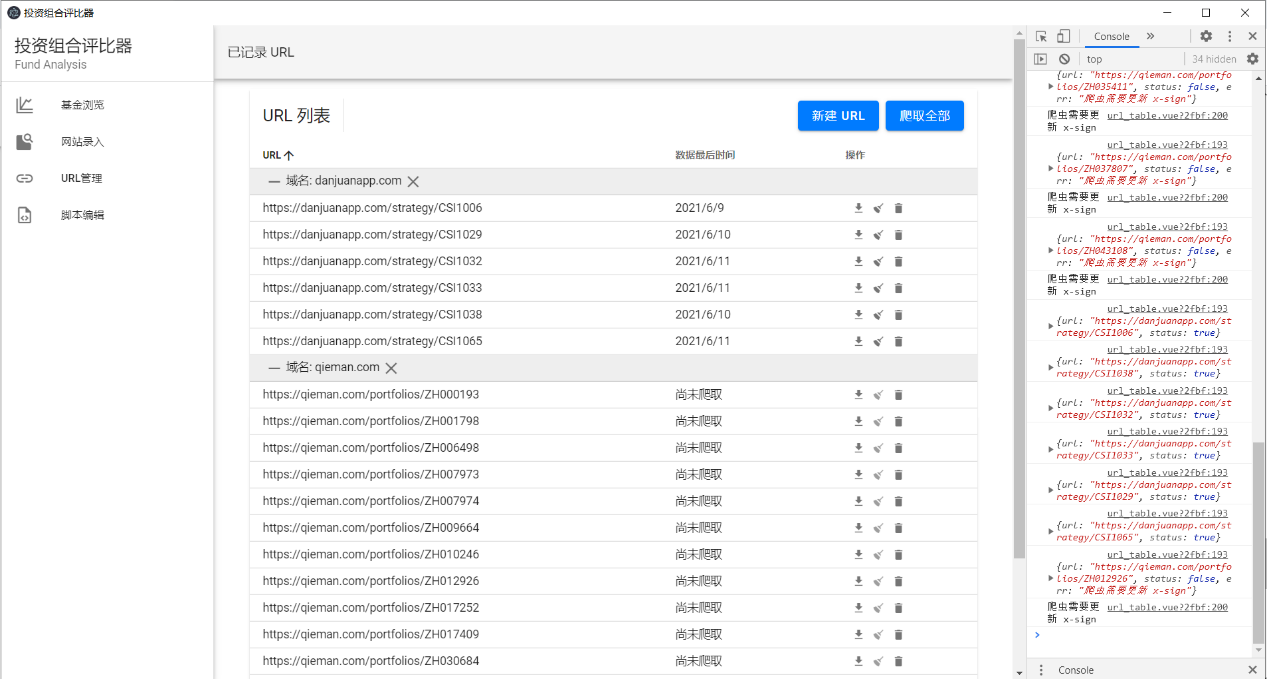
测试内容将分为四个部分：

1. 测试爬虫运行情况
2. 测试组合可视化渲染结果
3. 测试组合依据各个指标排序结果
4. 测试数据源扩展功能

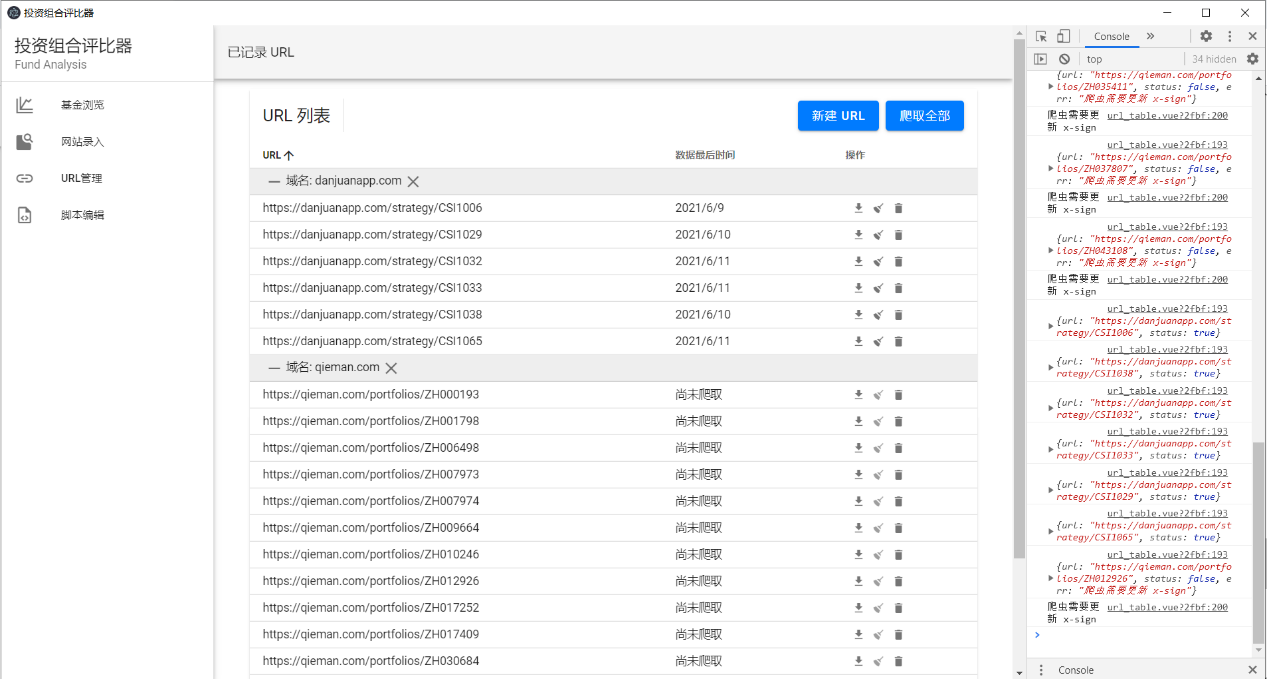
## 爬虫测试

用例一，对任务书中指定数据源数据进行正常爬取

按下爬取全部按钮，测试结果，蛋卷基金可以正确爬取，系统操作台显示各个数据状态均为“true”，表明数据正常爬取



用例二，数据爬取失败场景

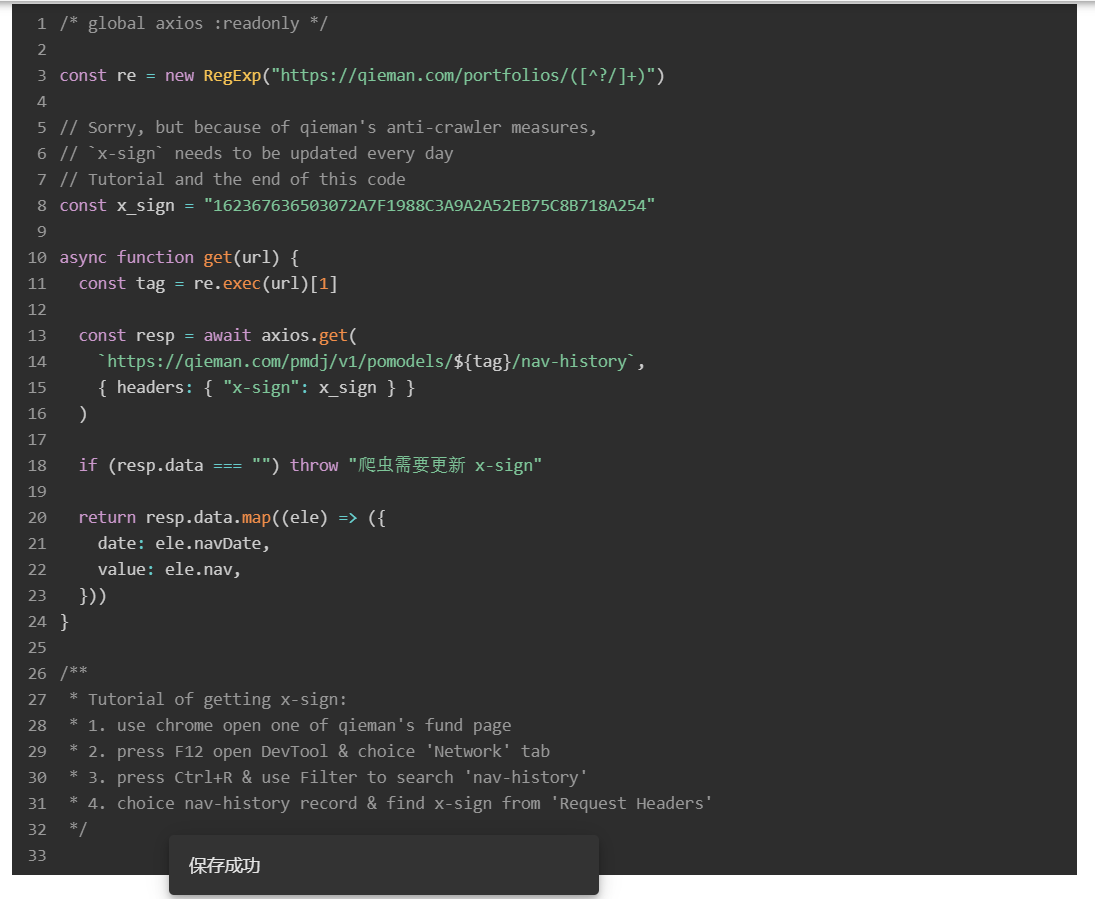


且慢数据爬取失败，系统操作台显示数据库中且慢状态为“false”，且界面有相应的报错提示，需要更新x-sign

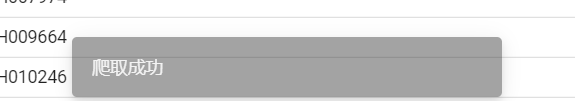


用例三，对失败场景修正

获取当日（2021-06-14）的x-sign后进入脚本编辑界面，更新爬虫

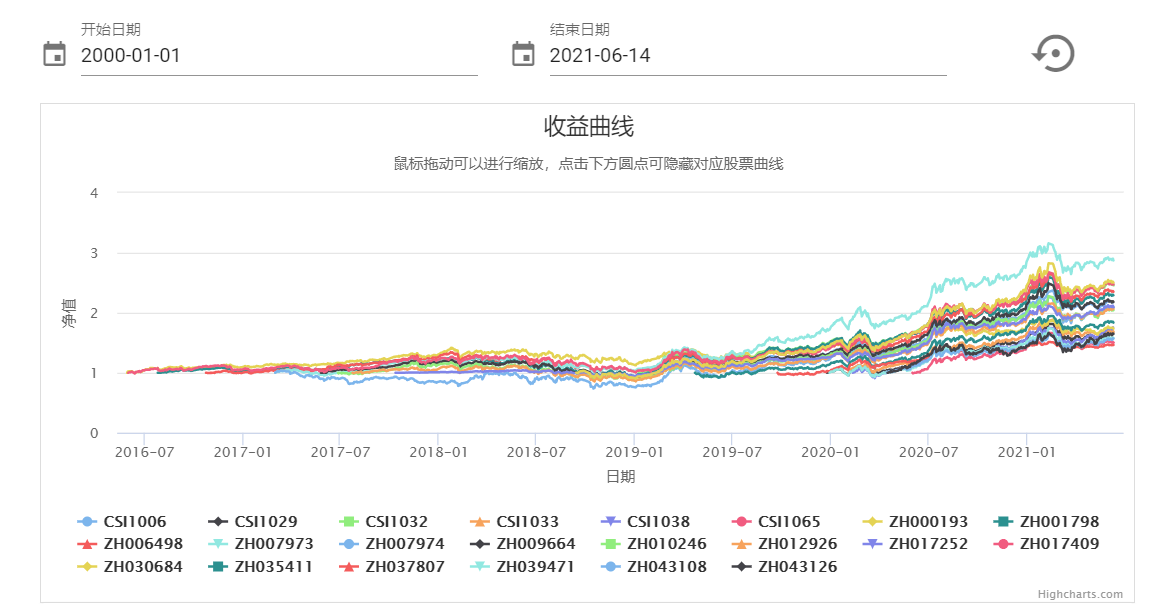


更新成功后，再次爬取数据源信息，下部消息提示爬取成功



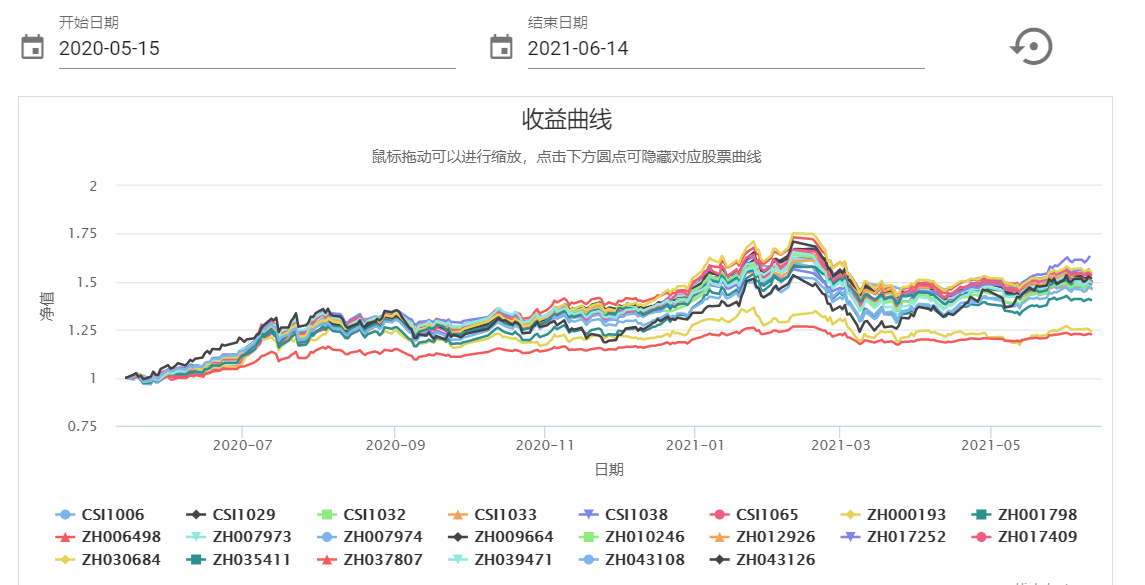
## 可视化测试

用例一，打开基金浏览界面，查看收益曲线总体绘制情况，各曲线绘制成功



用例二，查看特定时间段数据

点击上方选区框，选择起始和截至时间，将时间设置为2020-05-15到2021-06-14，收益曲线可以正常绘制



用例三，观察指定的一个组合

将鼠标悬停在曲线上，或者下方图例的标签上，可获得单独组合的高亮显示



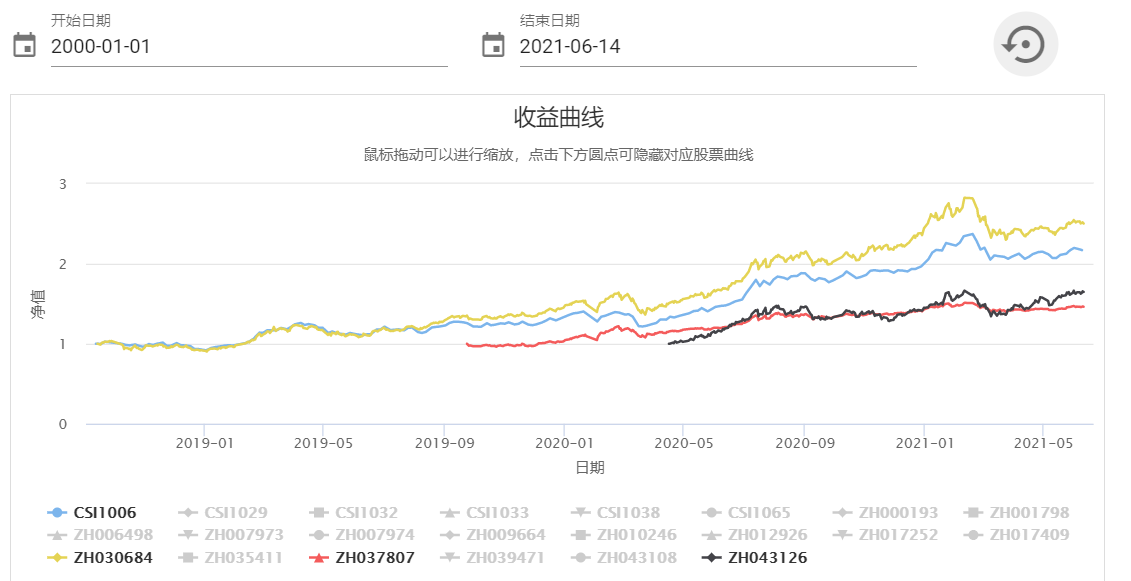
用例四，观察特定的数个组合

在下方图例部分，点击使得不需要的组合数据变暗，可视化界面可以显示图例中亮的组合信息。



用例五，返回初始化绘图时间，使得所有组合显示完整曲线

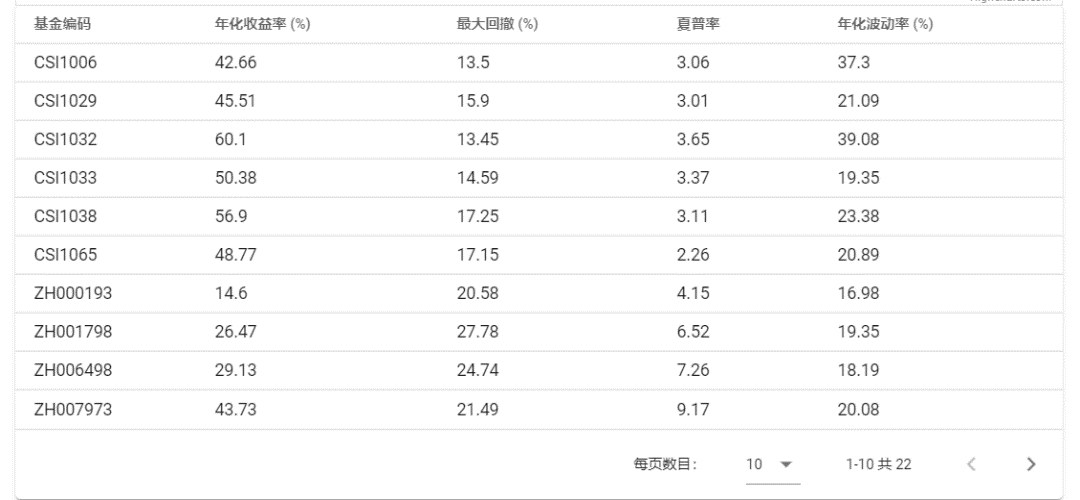
点击初始化按钮，各个数据起始点不相同，但仍然可以正常显示



## 数据排序测试

用例一，查看各项指标

按照预计结果显示



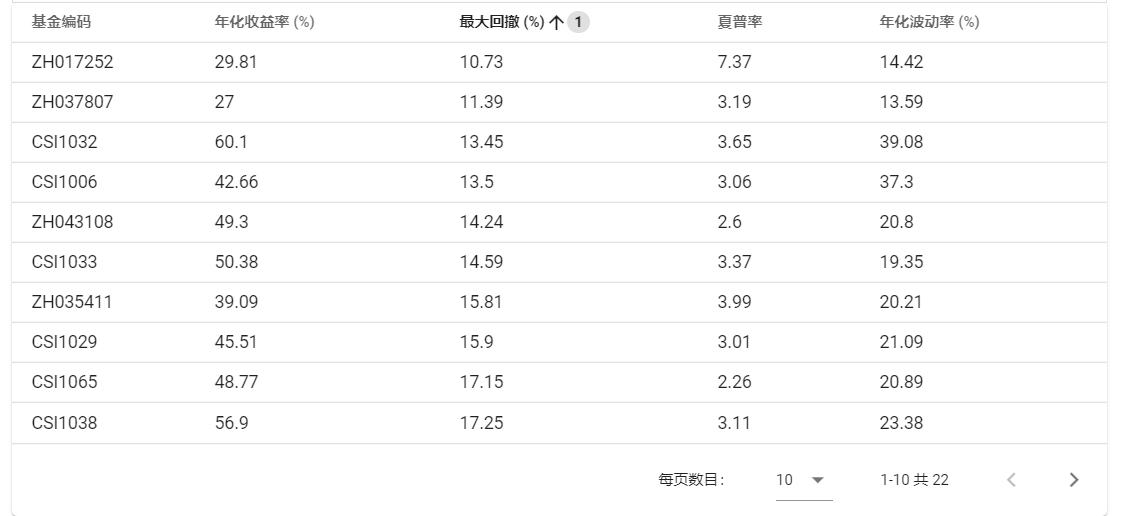
用例二，按照年化收益率降序排序

按照预计结果显示



用例三，按照最大回撤按照升序排序

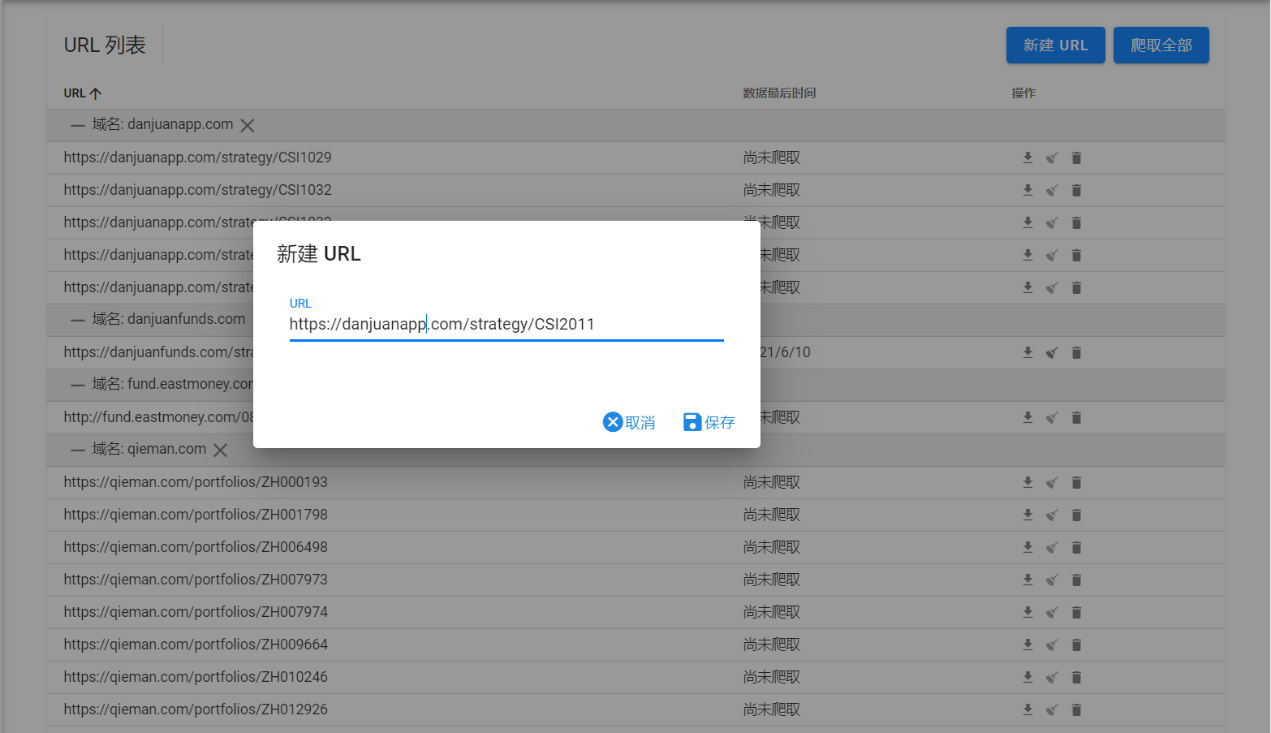
按照预计结果显示



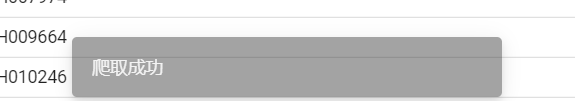
## 扩展数据源测试

用例一，新增蛋卷URL

新增组合CSI2011，将<https://danjuanapp.com/strategy/CSI2011>添加至管理界面



数据成功爬取，并可以可视化收益曲线和计算各项指标





用例二，扩展其他网站数据源

添加天天基金的投资组合数据源（<https://fund.eastmoney.com/000794.html>）加入评比器

进入网站录入界面将域名fund.eastmoney.com录入

进入URL管理界面将URL<https://fund.eastmoney.com/000794.html>录入

进入脚本编辑界面，基于模板修改爬虫，适配天天基金网，编辑完成后保存

再次进入URL管理界面点击爬取数据



进入基金浏览界面，数据可以被成功显示



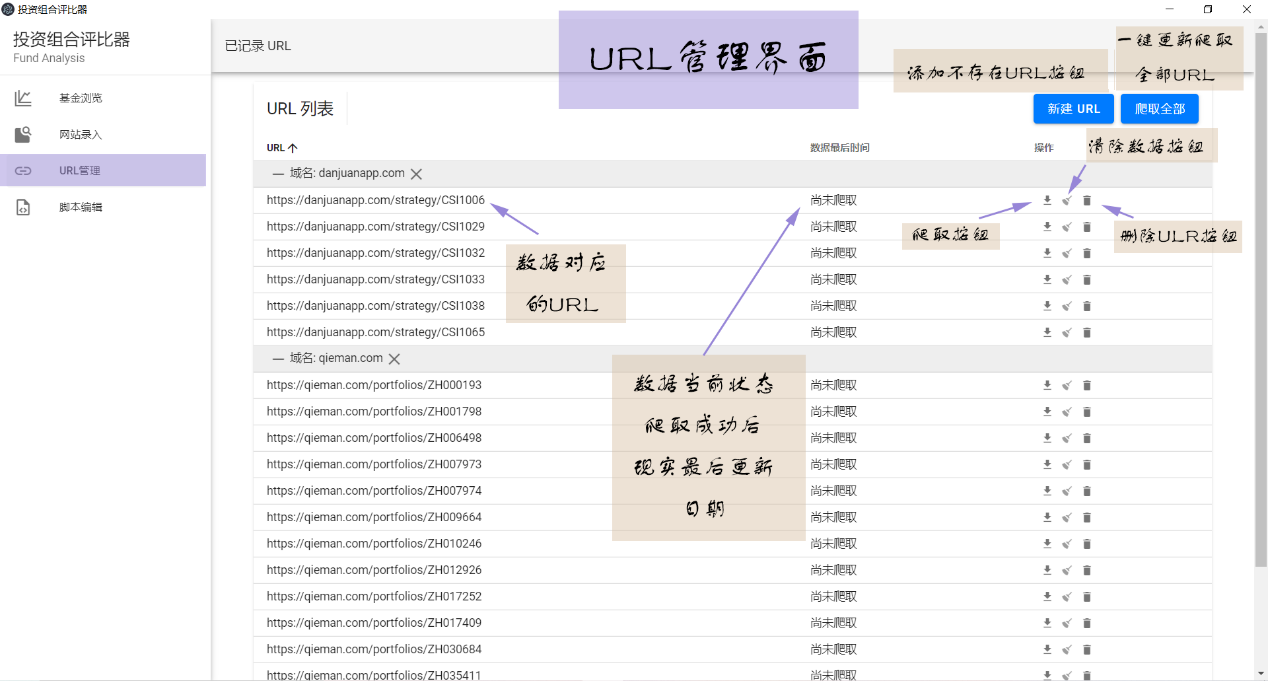
# 软件演示



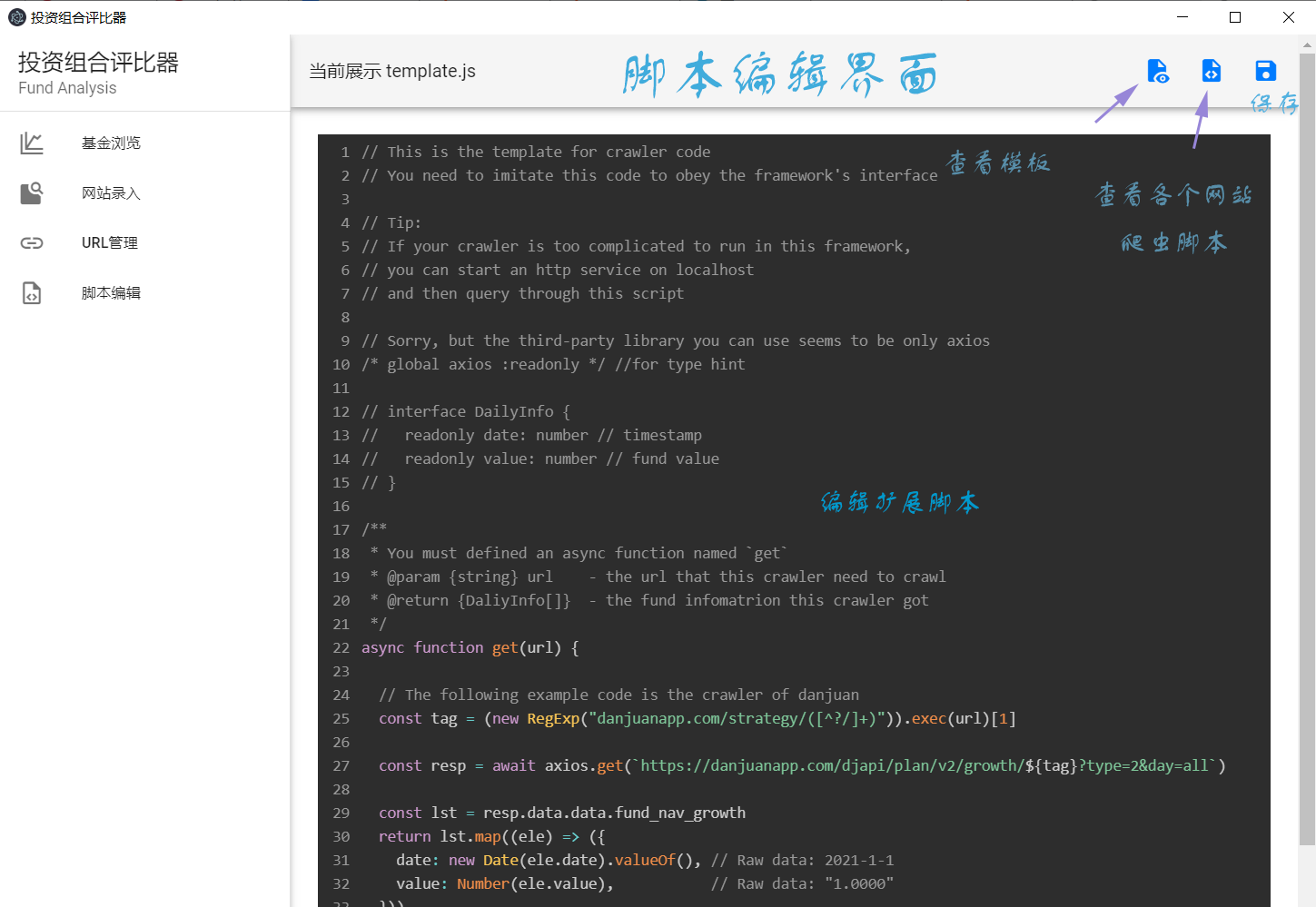
主界面功能展示



网站录入界面



URL管理界面



脚本编辑界面



基金浏览界面

# 软件使用教程

## 安装教程

### window下安装



运行可执行文件即可一键安装

### 从源码构建

步骤一、获取源码

步骤二、安装环境

访问 electronjs.org 查看最新稳定版适配的 NodeJS 版本并安装

在源代码目录下执行 npm install

步骤三、编译源码

在源代码目录下执行 npm run electron:build

步骤四、安装

等待命令执行完成后，即可在dist\_electron 中找到可执行文件，运行即可安装

## 软件功能使用教程

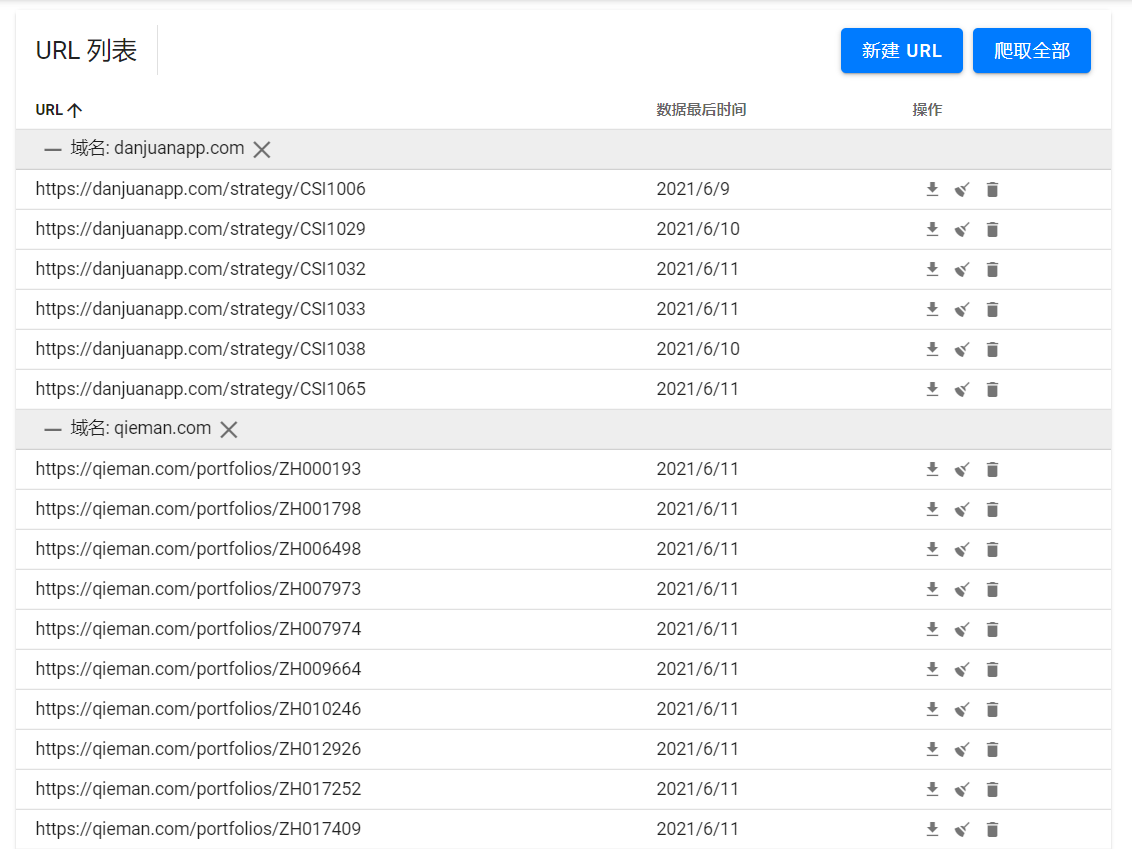
### 功能一：爬取数据

打开URL管理



点击爬取全部即可将数据存入数据库

若只想爬取单独数据可以点击对应URL右边的进行爬取



### 功能二：可视化基金投资组合

功能二需要在数据库中有数据下情况下才能使用，如果数据库中没有数据，请先使用功能一教程爬取数据后，再使用功能二。

第一步，打开右侧栏中的基金浏览界面

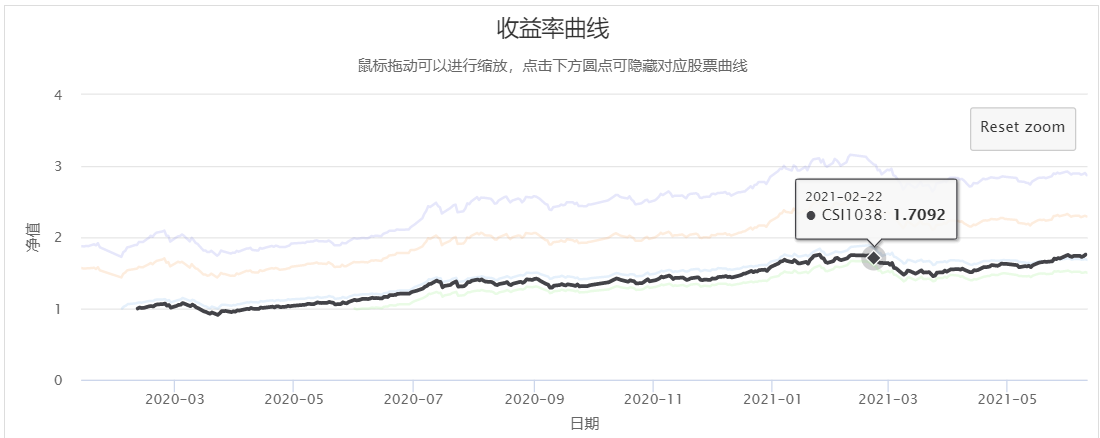


第二步，选择时间

选择起始时间和结束时间可以得到该时间段的数据



可选功能，高亮一个组合



将鼠标悬停在某个组合，或者图例上面即可观察某一个组合的走向



可选功能，观察特定几个组合

在图例上勾选，使不需要观察的组合变暗



### 功能三：按照某项指标对各个组合进行评比

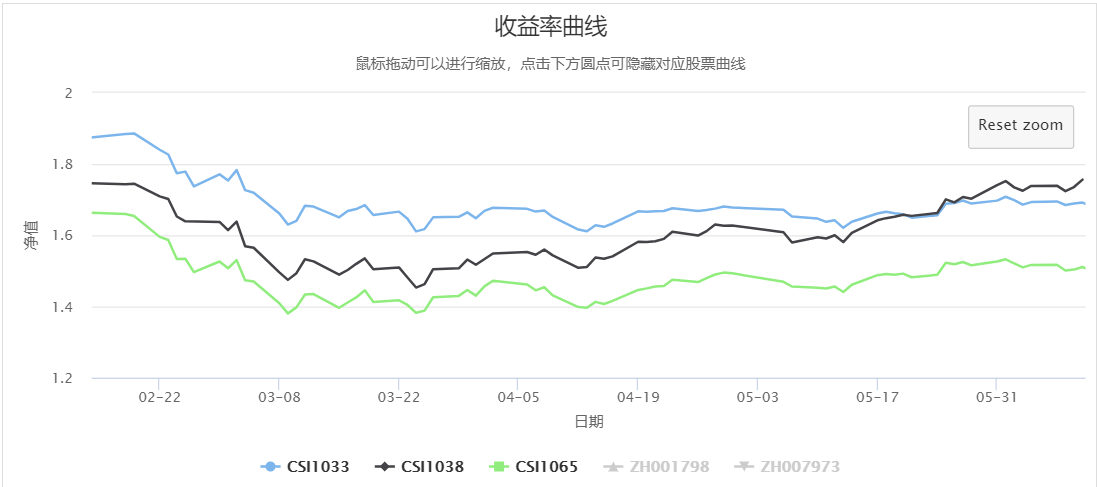


点击需要排序的文字，即可对该项指标按升序或降序排序。

可选功能，快速观察某一区间组合走向（查看快速下跌阶段日期分析原因，规避风险）



按下鼠标拖动使得选区区域覆盖所需观察时间段即可



按下图中 Reset zoom 按钮还原

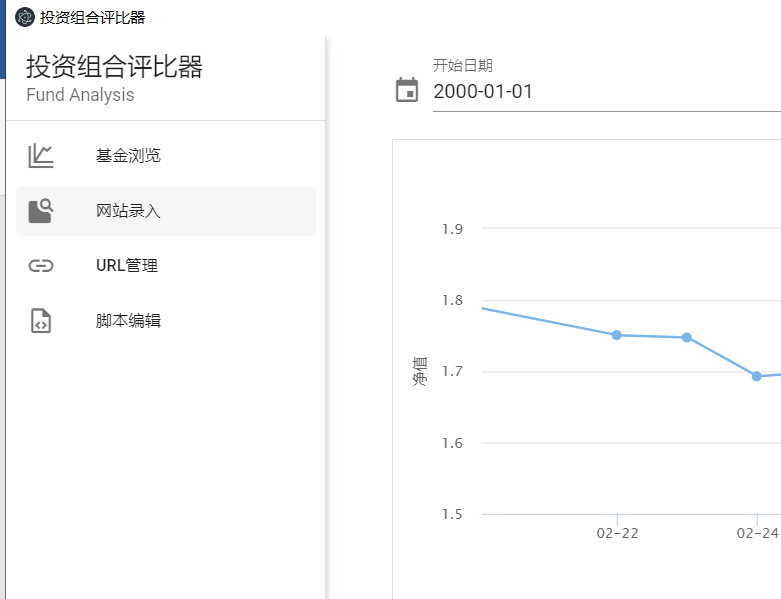
### 功能四：增加其他数据源

第一步，打开URL管理界面

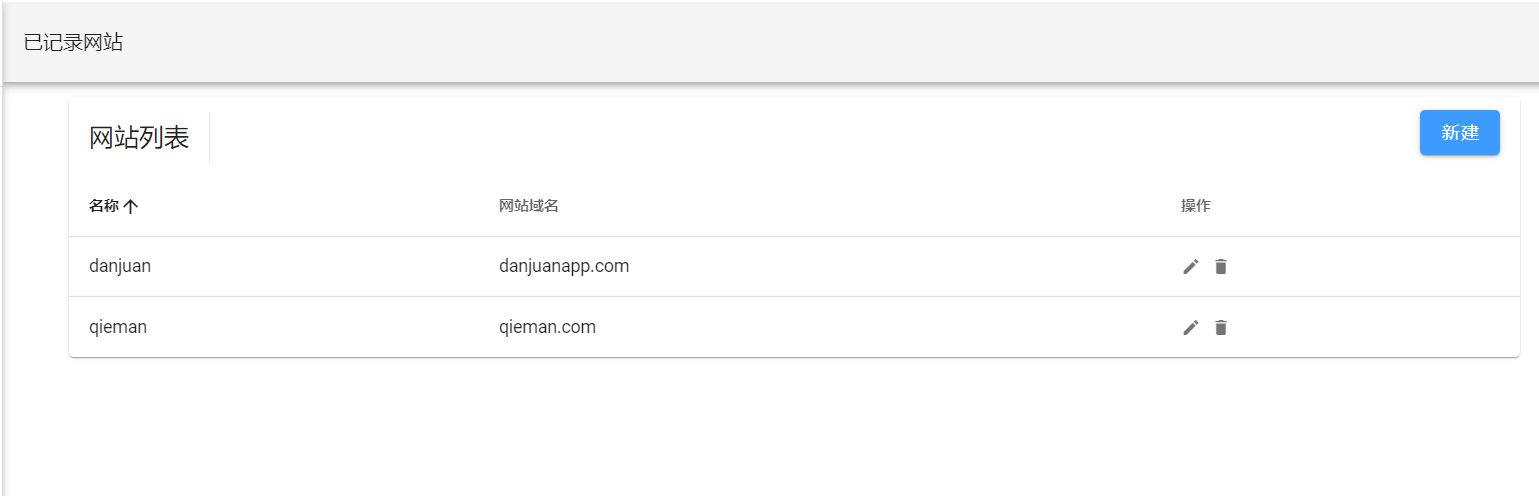


第二步，添加新数据源域名（若数据源在已添加域名下，则跳过此步）

点击网站录入，进入网站录入界面



打开界面后



点击按钮

进入网站录入界面，后输入网站的名字（自拟），网站的域名



现在以天天基金组合作为例子



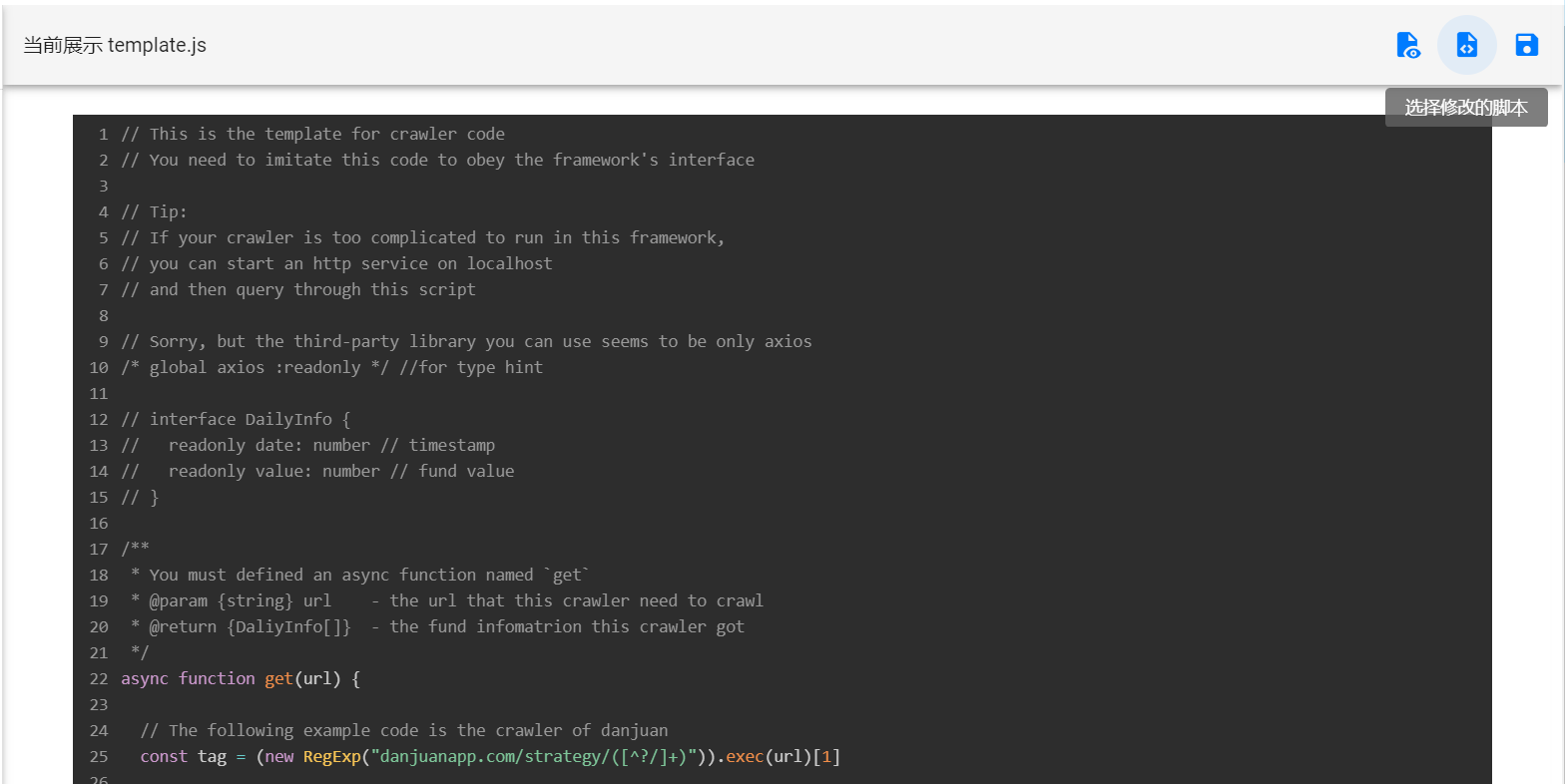
填写完成后按保存按钮

步骤三，修改域名下爬虫，适配新网站（若域名已存在则跳过此步骤）

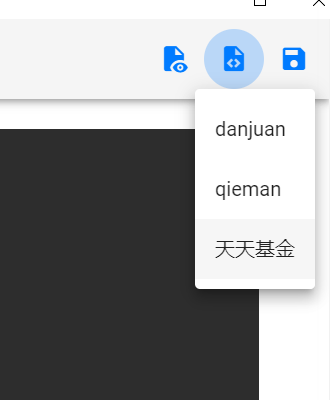
点击脚本编辑，进入编辑界面



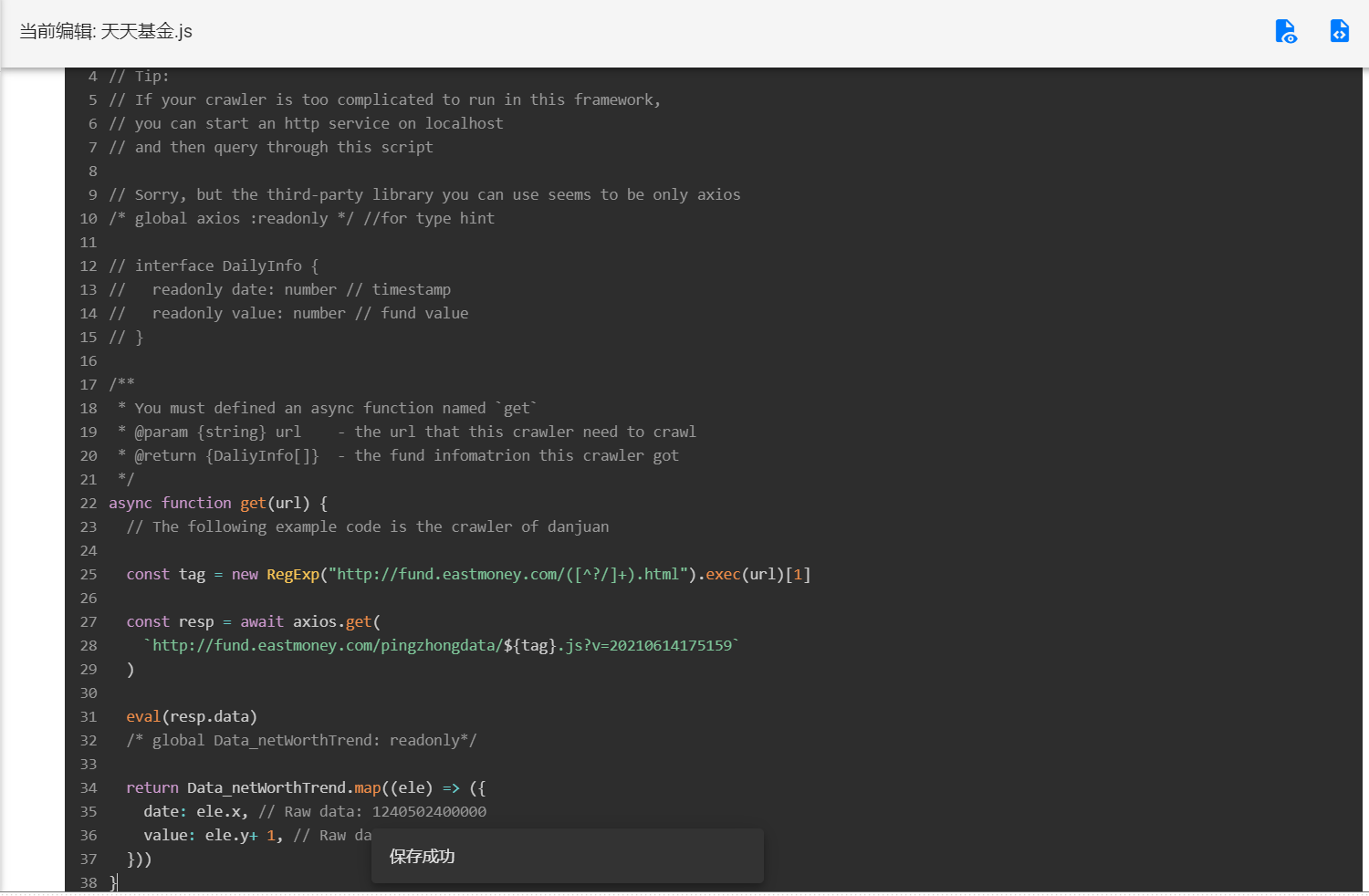
点击右上角的第二个按钮，选择修改的脚本



点击我们刚刚创建的天天基金



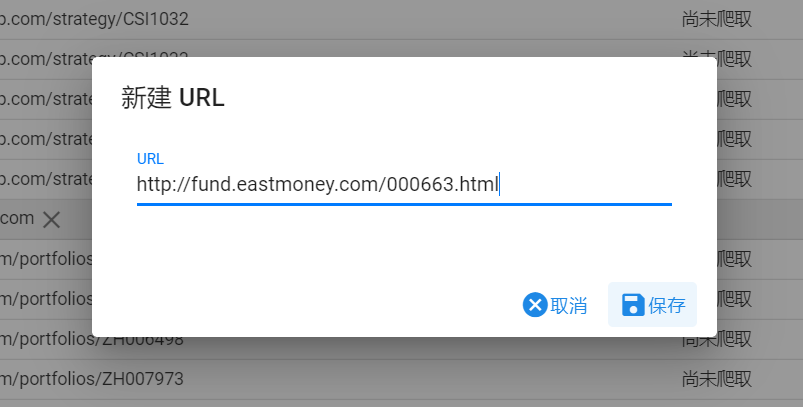
进行天天基金爬虫脚本的编写



编写完成后按右上角保存，或使用快捷键 Ctrl + S 进行保存

步骤四，添加URL

返回URL管理界面，点击填入URL



点击保存后，点击该URL右边的爬取数据



显示爬取成功，再打开基金浏览界面即可查看组合数据



成功可视化收益曲线，并且可以看到各项数据

# 小组讨论纪要

## 第一次讨论

时间：2021-05-10

成员：陈卓文，熊腾浩，庄昊敏

主要内容：

1. 确定了小组成员为陈卓文，熊腾浩，庄昊敏，组长为陈卓文。

2. 确定了选题为投资组合评比器。

## 第二次讨论

时间：2021-05-12

成员：陈卓文，熊腾浩，庄昊敏

主要内容：

1. 确定了开发的技术栈。

2. 确定了具体分工：由陈卓文负责基金数据爬取和应用框架搭建，由庄昊敏负责爬取数据处理，由熊腾浩负责前端页面实现。

3. 确定了第一次中期报告前需要完成的内容：完成基金数据爬取和数据处理。

## 第三次讨论

时间：2021-05-27

成员：陈卓文，熊腾浩，庄昊敏

主要内容：

1. 汇总了第一次中期报告前的任务完成情况。

2. 完成了第一次中期报告。

3. 确认了第二次中期报告前需要完成的内容：完成应用框架搭建和基础页面实现。

## 第四次讨论

时间：2021-06-09

成员：陈卓文，熊腾浩，庄昊敏

主要内容：

1. 汇总了第二次中期报告前的任务完成情况。

2. 完成了第二次中期报告。

## 第五次讨论

时间：2021-06-10

成员：熊腾浩，庄昊敏

主要内容：

1. 确认了前端页面和后端通信的接口。

2. 确定了前端页面的具体显示细节。

## 第六次讨论

时间：2021-06-12

成员：陈卓文，熊腾浩，庄昊敏

主要内容：

1. 汇总了全部任务完成情况。

2. 确定了所有文件的最后修改日期：2021-06-15。

3. 确定了最终展示时的分工。

# 个人总结

## 陈卓文

我从这次数据库开发实训中学到了很多。

首先我凭借这次的项目开发，了解了SQLite这一小型的、功能齐全的SQL数据库。

并且给予我一个将数据库课程中所学的知识应用到实际项目开发中的经验。

其次，我在编写应用界面的途中掌握了 HTML、JavaScript、以及Vue框架的用法。

凭借着这些高效的、先进的技术与框架，让我能够以最短的开发时间、最快的开发速度，构建出一个漂亮的、高效的应用界面。

这次实训，让我受益良多。

## 熊腾浩

我从这次数据库开发实训中学到了很多。

我在本次小组开发中负责软件页面的编写，凭借本次项目开发，我接触了 electron 框架，并用该框架完成了桌面端应用程序的开发。

同时，在本次开发中我掌握了 TypeScript 这一编程语言，在此之前我都是使用 JavaScript 语言编写前端页面，TypeScript 语言较其更为严谨，纠正了我一部分不好的编码习惯。

感谢学校开设数据库实训课程，感谢曾兵老师在实训过程中的教导，感谢助教老师们的帮助，本次实训使我受益良多。

## 庄昊敏

从2021.05.07开始选题，直到06.15结束，为期一个多月的数据库实训让我领会到数据库在软件中的作用。这一次综合实训，也让我在一个完整的团队项目开发中进一步了解了软件工程在软件开发过程中的作用。

在投资组合评比器开发中，需要在网站上爬取数据，需要使用数据库存储数据，需要在软件前端进行页面展示。这些都是我在此之前不了解的技术。在实训开发的过程中，快速学习爬虫，数据库的增删改查以及前端的绘图展示，并运用到项目中，是对我的一个极大挑战。 在学习和运用的过程中，也出现了很多更加细节上的问题，比如对前端数据json格式的解析和使用，还有基金指标涉及到的计算等问题。问题解决的是令人头疼的，但问题最终得到解决又鼓励着我往下开展工作和解决更多的问题。

这次实训中，我身处一个非常优秀的团队，团队的领导者教会我如何进行团队协作开发，让我对详细地体会到开发软件的各个阶段，对各个阶段应该完成的任务和需要的准备工作有了新的认知。在实践的过程中去学习软件工程，让我明白软件过程的规定并不是写在书本上的死知识，而是对现实的开发起着指导作用，让开发团队有更加高效的协作和更加良好的工作环境。队里的每一个人都给予了我很多技术上的教导，让我在学习的道路上更加轻松，也给我提供很好的工作氛围，让我期待拥有更多的团队协作开发经历。

# 项目进度安排

项目分为四阶段

第一周：确定开发技术栈，确定程序框架，实现基础的爬虫脚本

第二周：建立起基础的应用框架，计划后续开发的详细内容

第三、四周：实现数据源及爬虫数据的可视化管理，实现基金收益曲线的展示，实现投资组合评比。

第五周：对程序进行测试并修复可能出现的bug，编写实训报告等内容