# **软件设计报告书**

**项目内容：**用爬虫爬取51job网站中python开发工程师的薪资并求北京地区工程师的平均薪资

班级：软件1802

姓名：张凯恒

学号：120181080602

日期：2020/6/23

1. **引言**

整个项目完成了以下功能：

使用requests和BeautifulSoup库搭建网络爬虫；

使用爬虫爬取网站上指定信息；

用数据库对爬取到的信息进行存储；

从数据库中提取信息求得平均薪资。

文件构成如下图所示

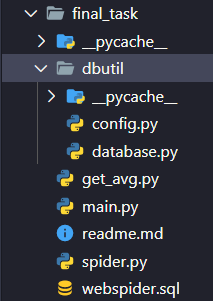


图1

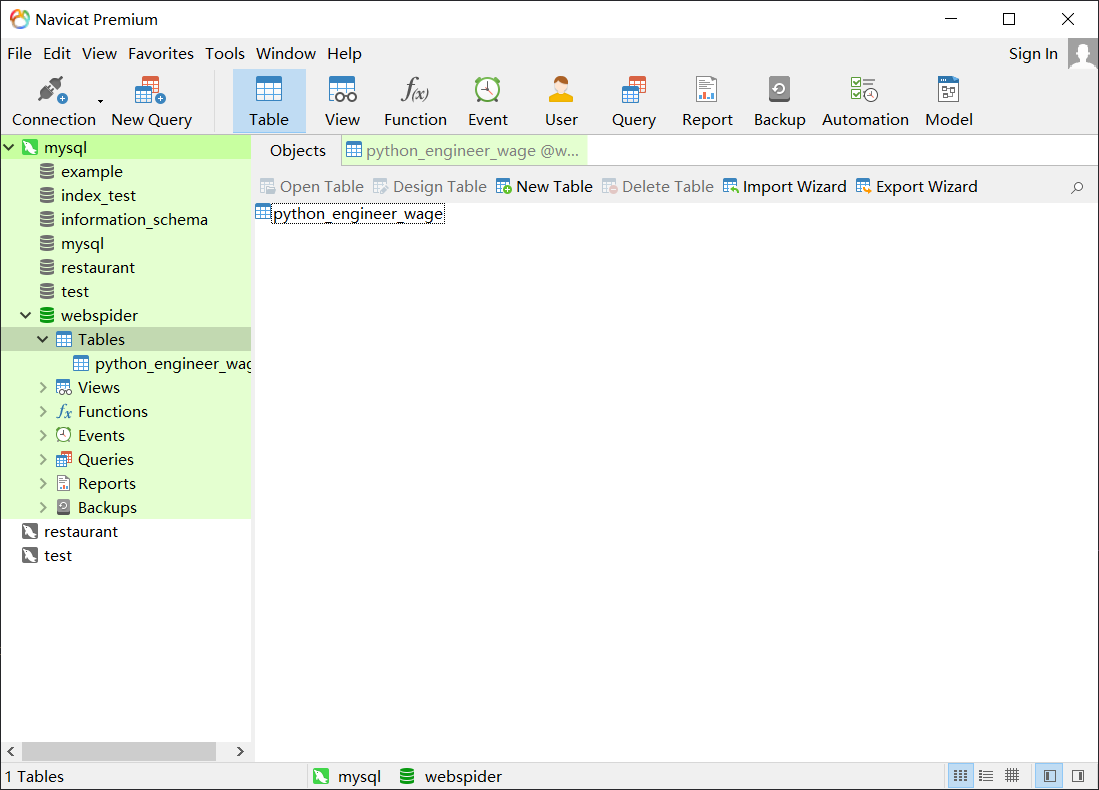
1. **功能设计**

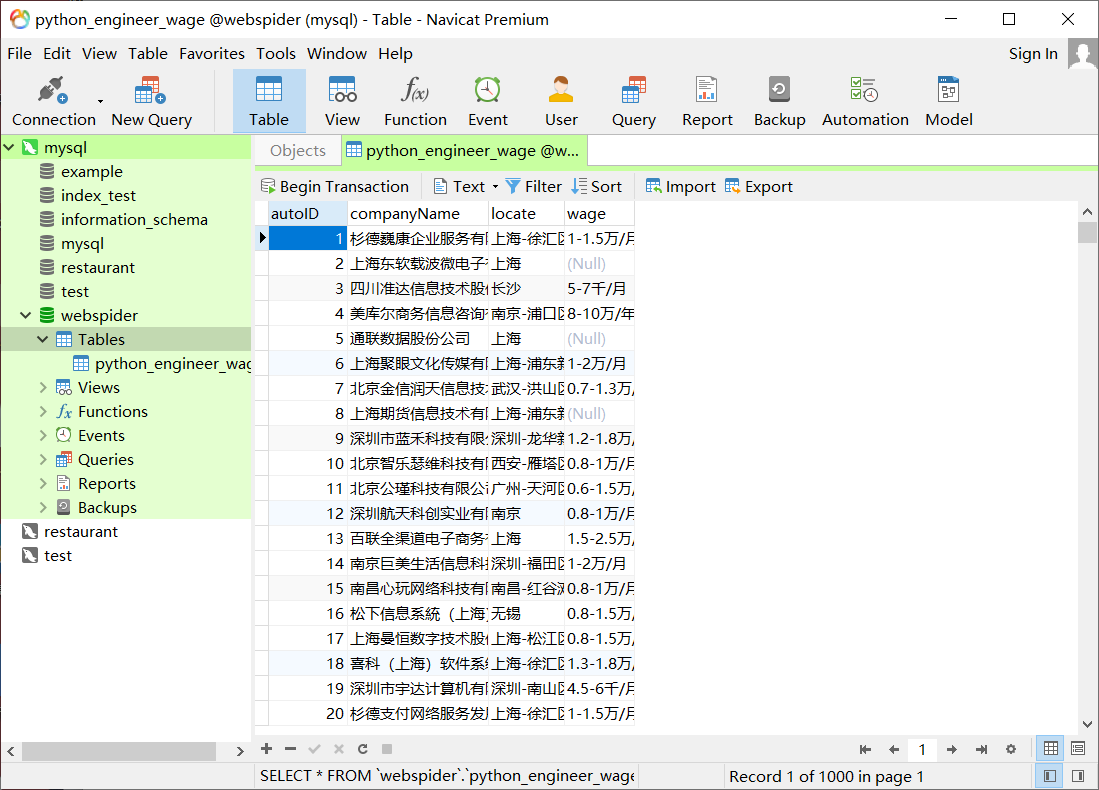
文件说明见readme.md文件。

程序流程：main文件启动爬虫，spider文件运行爬虫，在爬取网页信息的同时调用dbutil包中文件，将信息存入数据库中。dbutil包中封装了数据库连接的初始化与关闭，数据的存入与查询，以及数据库配置文件。爬虫运行结束后，get\_avg文件从数据库中取出北京地区数据并求得平均薪资。

1. **数据库设计**

数据库中仅有python\_engineer\_wage一个表单，用于存储python开发岗位薪资信息（共约26000条记录）。





1. **经验和教训**

**经验：获得了更多爬虫开发的技能，学会了最基本的软件架构模式，了解了MVC的基本开发形式。**

**教训：不要再写完大部分程序之后再进行调试，组织程序结构的时候要注意减小耦合度，否则修改起来非常麻烦。这个程序很早之前就完工了，但是修改消耗了非常多的时间。以后一定要提前架构，合理安排时间，避免仓促赶工的情况出现。**