

实验五 爬取网站图书数据(数据库)，数据集成、存储

5.1 实验介绍

本实验首先实现带有图书数据库网站建设,然后通过编程实现对网页背后数据库数据读取。

5.2 实验目标

- 1、了解网站中数据库的作用
- 2、掌握深网数据采集方法
- 3、掌握数据库的读取和存储

5.3 实验原理与方法

数据库使用，安装，导入数据库，mysql or sqlite 结合 python 的使用

5.4 实验步骤

- 1、在本地建立网站数据库和爬虫数据库。
- 2、编写本地服务器端代码，启动本地网络服务。
- 3、编写爬虫脚本，获取网页上的关键信息，采用 GET 和 POST 两种方法。
- 4、将爬取到的信息存入爬虫数据库中。

5.5 实验要求

1. 服务器端代码需要从数据库中读取相关数据，并显示到网页上。
2. 爬虫脚本可以使用之前实验中介绍的方法。
3. 需要在爬虫代码中直接将关键数据存入数据库中。

附录

一、网站包括图书内容说明：

包括序号、书名、ISBN 号、作者、图书简介

二、网站功能说明：

最少两个方法即 GET、POST。

方法 1 通过填写作者，查询后返回作者完成书名；

方法 2 通过填写 ISBN 号，查询后获得书名、作者、图书简介信息。