

# 华中农业大学本科生毕业论文(设计)中期进展情况检查表

\_\_\_\_\_年\_\_\_\_月\_\_\_\_日

课题名称	基于 Python 的主题爬虫设计与实现				
学生姓名	郑泉石	学 号	2013310200408	专业班级	计科 1304
指导教师	王建勇			职 称	讲师
主要研究内容及进展	<p>主要研究内容：</p> <p>（1）本文提到的针对各个主流电商平台网站的网页结构设计的爬虫程序，主要目的是通过调用网站 API 或者模拟浏览器获得网页的 json, xml 格式数据，或者网站 html 文件进行解析，获得商品的种子列表，进而使用多线程去分别抓取每一个商品的信息</p> <p>（2）搭建电商爬虫系统，从整体的需求入手进行整体框架和功能模块的设计与实现，整体使用 SQL 数据库+python 后台+web 前端的模式。</p> <p>（3）针对涉及到的主流电商平台的反爬虫技术的研究，包括但不限于 xhr 动态加载，js 动态生成，xsrf，cookie 认证，gzip 压缩返回内容</p>				
尚须完成的任务	<p>系统详细设计</p> <p>调试</p> <p>论文的撰写</p>				
存在的主要问题及解决措施	<p>爬虫系统的健壮性和支持的平台数量需要加强</p> <p>通过系统结构加强，容错机制的建立完成</p>				
指导教师审查意见					