



**本科学士毕业论文**

基于网络爬虫的房屋租赁信息筛选系统的设计与实现

姓 名： 云浩

学 号： 20151104745

院 系：计算机科学技术学院

年 级： 2015级

专 业： 网络工程

指导教师： 张志平、郭友全

日期： 2018 年 11月10 日至 2019 年 06 月10 日

目 录

[第一章 绪论 - 1 -](#_Toc390075741)

[1.1网售房屋的背景及意义 - 1 -](#_Toc390075742)

[1.2推荐系统研究现状 - 1 -](#_Toc390075743)

[1.3推荐系统研究内容及论文组织结构 - 3 -](#_Toc390075744)

[第二章 相关理论技术及工具 - 5 -](#_Toc390075745)

[2.1推荐系统算法及比较 - 5 -](#_Toc390075746)

[2.2基于内容的推荐算法 - 7 -](#_Toc390075747)

[2.3实现使用的环境工具及技术 - 10 -](#_Toc390075748)

[2.4本章小结 - 11 -](#_Toc390075749)

[第三章 系统设计 - 12 -](#_Toc390075750)

[3.1系统概述 - 12 -](#_Toc390075751)

[3.1.1系统目标设计 - 12 -](#_Toc390075752)

[3.1.2系统设计思想 - 12 -](#_Toc390075753)

[3.1.3系统总体设计图 - 13 -](#_Toc390075754)

[3.2需求分析 - 13 -](#_Toc390075755)

[3.2.1系统总体功能图 - 13 -](#_Toc390075756)

[3.2.2用例图 - 14 -](#_Toc390075757)

[3.3数据库设计 - 15 -](#_Toc390075758)

[3.3.1数据库需求分析 - 15 -](#_Toc390075759)

[3.3.2数据库逻辑设计 - 15 -](#_Toc390075760)

[3.4本章小结 - 18 -](#_Toc390075762)

[结束语 - 42 -](#_Toc390075778)

[致 谢 - 43 -](#_Toc390075779)

[参考文献 - 44 -](#_Toc390075780)

基于网络爬虫的房屋租赁信息筛选系统的设计与实现

计算机科学技术学院 2015级网络工程班 云浩 20151104745

指导教师 张志平 讲师、郭友全 讲师

摘要 在网络高速发展的现在，计算机已经离不开我们的生活，城市人口的饱和度增加，也导致着房屋价格的提升，刚毕业的大学生和一些上班族只能选择租房落脚，而选购租购房的平台，在互联网上已经多的不能再多了。这时如果能提供一个智能筛选房屋租赁信息的平台，无疑会为用户节省很大一部分时间。

关键词 互联网；房屋；爬虫

房屋租赁筛选系统采用爬虫技术将网上数家租房信息网的房屋出售信息进行爬取统一对比展示，研究用户对所购房屋的需求，使用户方便的利用本系统快速的查找出符合自己意愿的房屋信息。

本系统的设计意义是实现一个房屋信息筛选系统，对于房源信息中出现的垃圾房源信息进行自动删除，缩短了用户查找房源的时间，同时爬取数家网站的房源信息，可以给用户形成一个对比，找到合适的价位以及就近房源。

## 1.1 课题产生的背景及意义

1.1.1 课题产生的背景

1.1.2 课题的目的及意义

## 1.2推荐系统研究现状

二．国内房屋筛选系统研究现状

三．基于内容的推荐系统的研究进展

## 1.3推荐系统研究内容及论文组织结构

一．推荐系统研究内容

1.用户信息获取

2.推荐算法的研究

3.推荐系统的评价问题

二．论文的组织结构

本文的结构安排如下：本章分析了

# 第二章 相关理论技术及工具

完整支持HTML，Struts，JSF，CSS，Javascript,，SQL，Hibernate。同时支持AJAX，Hibernate，JSF，Struts，Spring， EJB3，JSP，JDBC，JavaServlet和数据库链接工具等多项功能。

2.MySql数据库

MySql是一个快速、多线程、多用户的SQL数据库服务器。其优点有：

(1)MySql的核心程序采用完全的多线程编程。线程是轻量级的进程，它可以灵活地为用户提供服务，而不过多的系统资源。用多线程和C语言实现的MySql能很容易充分利用CPU。

(2)MySql可运行在不同的操作系统下。

(3)MySql有一个非常灵活而且安全的权限和口令系统。当客户与MySql服务器连接时，他们之间所有的口令传送被加密，而且MySql支持主机认证。

(4)MySql拥有一个非常快速而且稳定的基于线程的内存分配系统，可以持续使用面不必担心其稳定性。事实上，MySql的稳定性足以应付一个超大规模的数据库。

3.Tomcat

Tomcat 服务器是一个免费的开放源代码的Web 应用服务器，属于轻量级应用服务器，在中小型系统和并发访问用户不是很多的场合下被普遍使用，是开发和调试JSP 程序的首选。

二． Servlet技术

Servlet是用Java语言编写的程序，运行于支持Java的Web服务器或应用服务器中。Servlet先于JSP出现，提供和客户端动态交互的功能。Servlet可以处理来自客户端的HTTP请求，并生成响应返回给客户端。Servlet具有如下优点：可移植性，安全性和高效性。

三．JSP技术

J S P是Java Server Page的缩写，是由Sun公司倡导、许多公司参与一起建立的一种应用范围广泛的动态网页技术标准，是Java语言编写的服务器端运行的页面。JSP页面由HTML代码和嵌入其中的Java代码组成，JSP可以被看作一种脚本语言，然而，作为一种脚本语言，在JSP中集合可以使用全部的Java类。支持JSP的服务器在JSP页面被客户端请求以后对这些Java代码进行处理，然后将生成的HTML页面返回给客户端的浏览器。Servlet是JSP技术的基础，而且许多大型的Web应用程序的开发需要Servlet和JSP配合才能完成。

四．JavaBean技术

JavaBean是为Java语言设计的软件组建模型，具有可重复使用和跨平台的特点。可以通过JavaBean来封装业务逻辑，进行数据库操作，从而很好的实现业务逻辑和前台后台的分离。

JavaBean可以分为两类，一类有用户接口，一类没有用户接口。一般在JSP中采用的是没有用户接口的，本系统也是采用此种类型的JavaBean。这类JavaBean只是简单的进行业务封装，如数据运算和处理、数据库操作等。

一个标准的JavaBean应该具有如下几个特点：

1. JavaBean必须是一个公开的类，也就是类访问权限必须是public的。
2. JavaBean必须具有一个无参数的构造方法。如果在JavaBean中定义了自定义的有参构造方法，就必须添加一个无参数构造方法，否则将无法设置属性；如果没有定义的有参构造方法，则可以利用编译器自动添加的无参构造方法。
3. JavaBean一般将属性设置成私有的，通过使用getXXX()方法和setXXX()方法来进行属性的取得和设置。

## 2.4本章小结

# 第三章 系统设计

## 3.1系统概述

### 3.1.1系统目标设计

### 3.1.2系统设计思想

。

### 3.1.3系统总体设计图

## 4.8本章小结

本章是系统的详细设计，详细介绍了数据库和JAVA Web的连接，数据持久层设计，数据访问层设计，和每个功能块的实现，其中每个功能块都详细介绍了实现的流程图。

# 结束语

# 致 谢

# 参考文献

**Design and Implementation of House Rental Information Screening System Based on Web Crawler**

Yunhao 2015 1104745, Network Engineering Class, Grade 2015, College of Computer Science and Technology

Instructor Zhang Zhiping and Guo Youquan

**Abstract** Nowadays, with the rapid development of the network, computers have been inseparable from our lives. The increasing saturation of the urban population has also led to the rise of housing prices. Graduates and some office workers can only choose to rent houses, while there are no more platforms for buying and renting houses on the Internet. At this time, if we can provide a platform for intelligent screening of housing rental information, it will undoubtedly save a lot of time for users.

**Keywords** Internet；Housing；Reptiles