****

A4纸

页面设置

上：2.54 cm

下：2.54 cm

左：3.17cm

右：3.17 cm

页眉：1.5cm

页脚：1.75cm

**本科毕业设计（论文）**

居中、黑体、小初、加粗



中文字体：黑体二号

英文字体：Calibri，大小根据字符数调整

其它文字体：相关专业统一规定实施

左缩进：5字符

右缩进：5字符

悬挂缩进：5字符

两端对齐

段前：0行

段后：0行

行距：单倍行距

居中

环绕方式：嵌入型

大小：4.5cm\*4.5cm

题 目：基于网络爬虫的股票信息预警系统的设计与实现

英文题目：DESIGN AND IMPLEMENTATION OF STOCK INFORMATION EARLY WARNING SYSTEM BASED ON WEB CRAWLER

学 号： 20141234567

姓 名： 王一

按教务系统中完整名称填写

班 级： 14软工A1

专 业： 软件工程

学部(院)： 工学部

工学部

文理学部

经济与管理学院

应用艺术设计学院

入学时间： 2014级

指导教师： 李红

日 期： 2020 年 5 月12日

封面末尾插入分节符（下一页）

毕业设计（论文）独创性声明

本人所呈交的毕业论文是在指导教师指导下进行的工作及取得的成果。除文中已经注明的内容外，本论文不包含其他个人已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中作了明确说明并表示谢意。

作者签名：

日 期：

声明末尾插入分节符（下一页）

基于网络爬虫的股票信息预警系统的设计与实现

中文题目

小二、黑体、居中。

毕业设计（论文）中英文摘要

1.5倍行距

摘要

摘要：三号、黑体、居中

本文结合网络爬虫技术实现对于股票交易信息、股票公告信息、股票财务信息的采集、解析、格式化、挖掘、维护与展示。再通过用户预设的条件对抓取的信息进行推送、预警。本文通过需求分析确定了系统应具有的基本功能包括股票数据获取、页面解析、解析内容格式化、数据整理、信息维护、信息浏览、设置预警、发送预警。采用面向对象的方法进行了总体设计、详细设计并最终实现了股票信息预警系统的主要功能。

摘要正文：小四、宋体、首行缩进2字符，字数300－500字

本文设计的股票信息预警系统共分为股票信息网页采集模块、网页解析模块、数据整理模块、数据浏览模块、预警模块共五个模块。股票信息采集模块采用爬虫技术实现，主要解决了如何准确快速获取增量的股票数据的问题。网页解析模块通过使用原生的XPATH 模块进行，获取需要的信息。数据处理模块采用Newtonsoft.Json库对Json 字符串对象化并存入关系型数据库。数据浏览模块是对数据库中数据的可视化展示。预警模块实现用户自我定制需要的信息条件，通过短信及邮件的方式进行推送。

简要陈述研究课题的内容、创新见解、主要论点

目前，系统处于运营维护阶段，可以稳定、高效的进行股票数据及相关信息的采集、解析、预警。

关键词：网络爬虫；股票预警；WEB挖掘

小四、宋体、分号间隔，最后一个关键词后面无标点符号

个数：3-7个，中英文对应

小四、黑体、顶格。

中文摘要末尾插入分页符

DESIGN AND IMPLEMENTATION OF STOCK INFORMATION EARLY WARNING SYSTEM BASED ON WEB CRAWLER

非英语类由专业统一规范字体、格式

英语题目

小二、Times New Roman、居中、加粗，一律大写字母母

ABSTRACT：三号、Times New Roman、居中加粗、大写

ABSTRACT

摘要正文：小四、Times New Roman、首行缩进2字符

This paper combines the web crawler technology of realizing the acquisition, analysis, formatting, excavation, maintenance and display of stock transaction information, stock announcement information and stock financial information. And then push and early warning the information crawled through the user's default conditions. This paper analyzes the basic functions that the system should have the demand analysis has defined, including stock data acquisition, page analysis, parsing content formatting, data collation, information maintenance, information browsing, setting early warning and sending early warning. The object-oriented method used to design the whole design, and the main function of the stock information early warning system finally designed and realized. The stock information early warning system designed in this paper is divided into five modules: stock information webpage acquisition module, web page analysis module, data collation module, data browsing module and early warning module. The stock information acquisition module is implemented by reptile technology, which solves the problem of how to get the incremental stock data accurately and quickly. The web analytics module makes use of the native XPATH module to get the information you need. The data processing module uses the Newtonsoft.Json library to object to the Json string and store it in a relational database. The data browsing module is a visual display of the data in the database. Early warning module achieves the user needs to customize the information conditions, and pushes it through the SMS and e-mail.At present, the system is in the stage of operation and maintenance, and stock data and related information can be collected, analyzed and warned stably and efficiently.

Key words: web crawler；stock early；warning；Web mining

摘要末尾插入分节符（下一页）

小四、Times New Roman、加粗、顶格、key words之间有一空格。

小四、Times New Roman、各关键词之间分号间隔，最后一个关键词后面无标点符号。

目录

目录：小二、黑体、居中

1.5倍行距

一级标题：四号、宋体、顶格

二级标题：小四、宋体、缩进2字符

三级标题：小四、楷体、缩进4字符

1 绪论 1

1.1 研究的背景 1

1.2 研究现状 1

1.3 研究的意义 1

1.4 研究的目标与内容 1

1.5 论文的组织安排 1

2 股票信息预警系统的相关理论与技术概述 2

2.1 信息采集系统概述 2

2.2 网络爬虫概述 2

2.2.1 网络爬虫的工作流程 2

3 结论 5

致谢 6

参考文献 7

附录 8

目录末尾插入分节符（下一页）

1. 绪论

外语类正文由专业统一规范字体、格式，报学部（院）备案

二级标题：序数居左顶格，空一格写标题，黑体、四号，1.5倍行距。

绪论通常为第一章，三号、黑体、居中，下空一行

* 1. 研究的背景

随着我国改革开放的脚步，股票日益成为人们生活中不可或缺的投资理财工具之一。股票作为重要经济活动之一，对于国内市场经济的繁荣与国民经济的发展都起到了至关重要的作用。

绪论：说明本课题的意义、目的、研究范围及要达到的技术要求；简述本课题在国内外发展概况及存在的问题；说明本课题的指导思想；阐述本课题应解决的主要问题

正文

首行缩进2字符、1.5倍行距

中文：小四、宋体

英文、数字：小四、Times New Roman

……

* 1. 研究现状

网络爬虫亦称信息采集系统是将网页中的非结构化信息进行抓取、清洗最终存入到关系型数据库中的软件。

针对股票数据具有实时更新的特点，本文采用的网络爬虫为增量采集系统。其大致的工作原理如下：

（1）对所有目标网页进行抓取

对总项包括的分项采用（1）、（2）、（3）……的序号。

（2）在之后的数据抓取过程中比较原网页与新抓取网页，对于没有更新的网页不进行采集。

……

* 1. 研究的意义

……

* 1. 研究的目标与内容

……

* 1. 论文的组织安排

……

各章节之间插入分页符

页脚为页码，小五、黑体、居中。

注：封面、声明、摘要、目录不加页码

1. 股票信息预警系统的相关理论与技术概述

正文一级标题

三号、黑体、居中、下空一行。

* 1. 信息采集系统概述

二级标题

序数居左顶格、空一格写标题、黑体、四号、1.5倍行距。

信息采集系统指从非结构化的信息、或者有大量冗余、噪声的文件中将所需的信息抽取出来保存至关系型数据库中的软件系统。

正文主体：阐述课题的主要研究内容、方案、研究过程、结果分析等内容，理工科的论文要有相应的数学模型、数据分析等，要求论理正确、论据确凿、逻辑性强、层次分明、表态确切及结论合乎逻辑

对于数据源为网页的采集系统往往采用网络爬虫技术

* 1. 网络爬虫概述

网络爬虫（Web Crawler）是指按照一定的规则，自动地抓取互联网信息的程序或者脚本。常见网络爬虫根据实现技术分类有通用(General Purpose)、增量(Incremental)、聚焦(Focused)、深层(Deep)等。在实际应用中往往需要将几类技术相互结合。

第三级、第四级标题均空两格书写序数、再空一格写标题、小四、黑体、居左顶格。

（分级阿拉伯数字的编号一般不超过四级，两级之间用下角圆点隔开，每一级的末尾不加标点）

正文

首行缩进2字符、1.5倍行距

中文：小四、宋体

英文、数字：小四、Times New Roman

* + 1. 网络爬虫的工作流程

对于本程序由于股票的页面相对固定，因此可以采取将股票代码作为一个线性表，对每个股票代码进行遍历获取网页。另外还要对获取的信息与数据库中保存的信息进行比较，避免重复。网络爬虫工作流程图见图2-1。



图序和图题：标于图的下方居中，图序和图题中间空一格，五号、黑体

图内：五号、中文宋体，英文Times New Roman

图片格式：嵌入文本行中，居中

插图位于引用该插图的字段后面

每幅插图应有图序和图题，插图可以统一编序，也可以逐章单独编序，图序必须连续，不得重复或跳缺。

图2-1 网络爬虫工作流程

……正文内容……

……正文内容……

通过上文的CDM与PDM模型构建数据库结构创建如下表：

表2-1 ANNOUNCEMENT表结构

表格位于引用该表格字段的后面

每张表格应有表序、表题，表格可以统一编序，也可以逐章单独编序，图序必须连续，不得重复或跳缺。

表序和表题：表格上方居中，表序和表题中间空一格、五号、黑体、表题末尾不加标点

表内：五号、中文宋体，英文Times New Roman

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 名称 | 说明 | 数据类型 | 长度 | 主键 | 外来键 |
| CODE | 股票代码 | VARCHAR2(20) | 20 | TRUE | TRUE |
| URL | 公告URL | VARCHAR2(500) | 500 | TRUE | FALSE |
| TITLE | 标题 | NVARCHAR2(200) | 200 | FALSE | FALSE |
| DAYS | 日期 | DATE |  | FALSE | FALSE |
| ALARMED | 是否已预警 | VARCHAR2(20) | 20 | FALSE | FALSE |
| …… |  |  |  |  |  |
| …… |  |  |  |  |  |
| …… |  |  |  |  |  |
| …… |  |  |  |  |  |
| …… |  |  |  |  |  |
| …… |  |  |  |  |  |
| …… |  |  |  |  |  |
| …… |  |  |  |  |  |
| …… |  |  |  |  |  |
| …… |  |  |  |  |  |
| …… |  |  |  |  |  |
| …… |  |  |  |  |  |
| …… |  |  |  |  |  |
| …… |  |  |  |  |  |
| …… |  |  |  |  |  |
| …… |  |  |  |  |  |
| …… |  |  |  |  |  |
| …… |  |  |  |  |  |
| …… |  |  |  |  |  |

续表2-1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 名称 | 说明 | 数据类型 | 长度 | 主键 | 外来键 |
| …… |  |  |  |  |  |
| …… |  |  |  |  |  |

表格跨页

表题省略，表头应重复写，并在右上方写“续表xx”。

公式应另起一行

正文中的公式、算式或方程式等应编排序号，公式的编号用圆括号括起，序号标注于该式所在行（当有续行时，应标注于最后一行）的行末。

公式可按章节顺序编号或按全文统一编号。公式序号必须连续，不得重复或跳缺。重复引用的公式不得另编新序号。

公式和编号之间不加虚线。

……正文内容……

 （2-1）

较长的公式，如必须转行时，最好在等号处转行,如做不到这一点,要在+，-，×，÷等数学符号处转行。数学符号应写在转行处的行首。上下式尽可能在等号“＝”处对齐。

 （2-2）

1. 结论

……正文内容……

对整个研究工作进行归纳和综合，阐述本课题研究中尚存在的问题及进一步开展研究的见解和建议。结论要写得概括、简短。

致谢

在论文即将完成的时候，我要由衷感谢 ：

…

在此，我要再次向他们表示深深的谢意和衷心祝福。

谢辞应以简短的文字对在毕业设计过程中曾给予帮助的人（例如指导教师）表示自己的谢意。

参考文献

[1]李旭乐，宗光华.生物工程微操作机器人视觉系统的研究[J].北京航空航天大学学报，2002（2）：22-25

[2]孙家正，杨长青.计算机图形学[M].北京：清华大学出版社，1995:26-28

不少于10篇（含3篇以上外文类），一半以上需为近4年出版或发表

书写格式要按国家标准GB7714-87规定（例如：[1] 呂炎主编.锻造工艺学.第1版.北京：机械工业出版社,1995）。

参考文献按在正文中出现先后，统一用阿拉伯数字进行编号，序码用方括号括起，且在正文引用处最后一个字的右上角，用方括号标明此序号（如[1]，[3] ），以便于查找。

附录

一些不宜放在正文中，但又直接反映完成工作的成果内容。如作品影像资料、计算机程序等材料附于毕业设计（论文）之后。