



项目计划书

所属学校：青岛科技大学

参赛项目：基于互联网数据的大数据分析

团队名称：GOODLUCK 团队

团队成员：张莹、刘桐源、邵华、乔善宝、寻宇星

2017 年 9 月

目录

目录	2
1 第一部分 前言	3
2 第二部分 项目概述	6
3 第三部分 实施计划	7

1 第一部分 前言

1.1 编写目的

本文档说明程序设计的关键部分,即详细设计说明书, 另其他详细代

码实现请查看代码源。且本文档是描述基于互联网的大数据分析的设计文档。

1.2 应用背景

1.2.1 大数据趋势

趋势一：数据的资源化

何为资源化，是指大数据成为企业和社会关注的重要战略资源，并已成为大家争相抢夺的新焦点。因而，企业必须要提前制定大数据营销战略计划，抢占市场先机。

趋势二：与云计算的深度结合

大数据离不开云处理，云处理为大数据提供了弹性可拓展的基础设备，是产生大数据的平台之一。自 2013 年开始，大数据技术已开始和云计算技术紧密结合，预计未来两者关系将更为密切。除此之外，物联网、移动互联网等新兴计算形态，也将一齐助力大数据革命，让大数据营销发挥出更大的影响力。

趋势三：科学理论的突破

随着大数据的快速发展，就像计算机和互联网一样，大数据很有可能是新一轮的技术革命。随之兴起的数据挖掘、机器学习和人工智能等相关技术，可能会改变数据世界里的很多算法和基础理论，实现科学技术上的突破。

趋势四：数据科学和数据联盟的成立

未来，数据科学将成为一门专门的学科，被越来越多的人所认知。各大高校将设立专门的数据科学类专业，也会催生一批与之相关的新的就业岗位。与此同时，基于数据这个基础平台，也将建立起跨领域的数据共享平台，之后，数据共享将扩展到企业层面，并且成为未来产业的核心一环。

趋势五：数据泄露泛滥

未来几年数据泄露事件的增长率也许会达到 100%，除非数据在其源头就能够得到安全保障。可以说，在未来，每个财富 500 强企业都会面临数据攻击，无论他们是否已经做好安全防范。而所有企业，无论规模大小，都需要重新审视今天的安全定义。在财富 500 强企业中，超过 50%将会设置首席信息安全官这一职位。企业需要从新的角度来确保自身以及客户数据，所有数据在创建之初便需要获得安全保障，而并非在数据保存的最后一个环节，仅仅加强后者的安全措施已被证明于事无补。

趋势六：数据管理成为核心竞争力

数据管理成为核心竞争力，直接影响财务表现。当“数据资产是企业核心资产”的概念深入人心之后，企业对于数据管理便有了更清晰的界定，将数据管理作为企业核心竞争力，持续发展，战略性规划与运用数据资产，成为企业数据管理的核心。数据资产管理效率与主营业务收入增长率、销售收入增长率显著正相关;此外，对于具

有互联网思维的企业而言，数据资产竞争力所占比重为 36.8%，数据资产的管理效果将直接影响企业的财务表现。

趋势七：数据质量是 BI(商业智能)成功的关键

采用自助式商业智能工具进行大数据处理的企业将会脱颖而出。其中要面临的一个挑战是，很多数据源会带来大量低质量数据。想要成功，企业需要理解原始数据与数据分析之间的差距，从而消除低质量数据并通过 BI 获得更佳决策。

趋势八：数据生态系统复合化程度加强

大数据的世界不只是一个单一的、巨大的计算机网络，而是一个由大量活动构件与多元参与者元素所构成的生态系统，终端设备提供商、基础设施提供商、网络服务提供商、网络接入服务提供商、数据服务使能者、数据服务提供商、触点服务、数据服务零售商等等一系列的参与者共同构建的生态系统。而今，这样一套数据生态系统的基本雏形已然形成，接下来的发展将趋向于系统内部角色的细分，也就是市场的细分；系统机制的调整，也就是商业模式创新；系统结构的调整，也就是竞争环境的调整等等，从而使得数据生态系统复合化程度逐渐增强。

2 第二部分 项目概述

在这样的大背景下，我们设计了网络爬虫从各大招聘网站上爬取尽可能多的招聘数据，对原始数据进行清理工作，对清理之后的数据

进行可视化。

2.1 项目主旨：

本产品的主要对象是大学生或者是准备应聘计算机岗位的人。希望借此可以认清当前计算机岗位的就业情况

2.2 项目说明：

该项目所使用的数据均是从各大招聘网站上爬取下来的数据，数据保证真实有效且未违反爬虫爬取协议，并未违反国家相关法律法规规定。

2.3 项目开发人员分派

- 1.任务提出者：齐鲁大学生软件设计大赛
- 2.开发者：青岛科技大学 GOODLUCK、团队
- 3.面向用户：大学生和准备踏入 IT 行业的人

3 第三部分 实施计划

3.1 任务分解及人员分工

任务名称	计划日期	参与人员
查资料并讨论	6月-7月	全体成员
项目计划书	7.15-7.20	寻宇星、乔善宝
可行性分析	7.20-7.22	全体成员
需求分析	7.22-7.23	全体成员

概要设计	7.23-7.24	全体成员
详细设计编写	7.24-7.26	全体成员
网络爬虫编写	7.26-8.5	张莹、邵华
数据存储	8.5-8.10	张莹
数据清洗	8.10-9.10	刘桐源
数据可视化	8.30-9.13	邵华
演示 ppt 制作	9.13-9.15	张莹
视频的拍摄与制作	9.15	张莹
项目总结报告	9.16-9.19	全体成员

3.2 预算及开发时间

本项目经费自理，开发时间为 2017 年 7 月到 2017 年 9 月。

3.3 软件开发关键问题

1. 网络爬虫。
2. 存储容器的选择。
3. 数据清洗及数据分析的方式。
4. 数据可视化的方式。

3.4 人员及分工问题

策划及软件构造：全体成员

编码：张莹、刘桐源、邵华

文档：寻宇星、乔善宝

视频及 ppt 制作：张莹

3.5 交付期限

本项目最迟交付日期为 2017 年 9 月 20 日。