

## 命题赛题目三

# 求职者智能分析系统

### 一、选题背景：

大数据和“云计算”像是一枚硬币的正反面一样慢慢勾勒出当今世界的财富价值风向。大数据的出现得益于互联网行业的快速发展、计算机硬件和软件能力的不断提升。大数据技术现已被应用到各行各业，而在招聘求职领域，我们希望通过爬虫技术、机器学习、文本挖掘、统计分析等手段帮助求职者更好的了解市场需求，从而有一个清晰、明确的求职方向。

数据挖掘工程师

合肥李山聚财网络科技有限公司

五险一金

年底双薪

绩效奖金

加班补助

全勤奖

交通补助

餐补

定期体检

立即申请

职位描述

公司介绍

举报★收藏

岗位职责：

1、参与建设大数据分析系统；

2、对海量业务数据、用户数据进行分析挖掘，为公司在行业精准营销、运营及决策提供业务分析及数据支持；

3、根据研究成果及业务需求，形成解决方案，实施应用项目；

4、开发出独立知识产权算法及知识库，固化成数据产品；

任职要求：

1、大专及以上学历，计算机、应用数学、统计学、交通等相关专业优先；

2、具有2-3年以上数据挖掘/数学建模相关工作经验，具有深厚的统计学、数学、人工智能和数据挖掘知识基础，具备数理统计理论基础；

3、熟悉数据挖掘方法并常用的统计工具软件，如R、SAS、EG/EM、SPSS等数据挖掘工具；

4、熟悉Python/Java编程语言、MySQL/MySQL等关系型数据库；

5、拥有海量数据处理和分析能力，熟悉分布式计算架构，有Spark、Hadoop平台的开发经验者优先；

6、精通数据挖掘方法论，熟悉逻辑回归、决策树、随机森林等分类和聚类、预测、关联机器学习算法，并具备相关项目经验；

7、具有高度的责任感、良好的敬业精神和团队合作精神；能够主动获取新知识，具有较强领悟力，有系统性思考解决问题能力；

合肥李山聚财网络科技有限公司

公司规模：100-499人

公司性质：民营

公司行业：互联网/电子商务

公司地址：合肥李山聚财网络科技有限公司

突破女性  
边界束缚

实用指南

### 二、题目概况：

本系统首先需要通过网络爬虫爬取 51job 等招聘网站上大数据相关职位的招聘信息（如下图）。提取出其中的关键数据，通过对这些

信息的挖掘分析,可以更加精准、清晰的指导求职者所在行业的待遇水平、自身可能的待遇、以及对公司、行业的选择。请参赛队伍根据需求描述所提供的思路,完成“求职智能分析系统”的设计、开发、部署工作。

本题要求选手使用 H3C 大数据平台产品 H3C DataEngine 系列产品完成云计算和大数据处理平台的构建,并基于该平台在核心算法上进行创新,挖掘真实招聘数据中有价值的信息。

### 三、需求描述:

#### 需求一(必选): 网络爬虫

网络爬虫根据选定的网站列表(如智联招聘, 51job 等)进行爬取,爬取网站公布的招聘信息。并将爬取到的信息进行整理后存储在数据库中。

#### 需求二(必选): 网络数据分析

解析爬取的数据,进行行业招聘态势的分析。设计文本分析、数据挖掘等相关算法,得到有趣和有价值的结论,同时为求职者的职位智能推荐做准备。鼓励选手大胆发挥创意和想象,通过优秀的算法设计,挖掘更多有价值和有意义的信息。

#### 需求三(必选): 智能检索和推荐

分析选手的求职信息,设计数据分析算法为求职者推荐多元化的信息和匹配的职位。示例场景:当求职者输入学历、专业、学校、求职地、工作年限、技能、岗位名称等基本信息后,系统将智能分析出

该职位的待遇水平、求职者的待遇区间、可能去的公司、公司性质和规模、行业、匹配概率等信息。

### 三、参考资料：

本题配有详细的命题指南，更多详细的命题解析请参考本题配套的命题指南。

*H3C DataEngine 介绍：*

[http://www.h3c.com/cn/Products\\_Technology/Products/Big\\_Data/Catalog/DataEngine/DataEngine/](http://www.h3c.com/cn/Products_Technology/Products/Big_Data/Catalog/DataEngine/DataEngine/)

本赛题提供软件安装包、技术文档和技术支持，请在报名后联系命题联系人索取。