

队伍成员：

曾永家右 黄泰北 赵辉

龙兰心 罗雪蕾

# 基于大数据的人岗匹配 系统“职达”

## 需求分析

职达信息技术有限公司

达芬奇队

指导教师：徐进 尹帮旭

目录

第一章 引言 ..... 1

    1.1 目的 ..... 1

    1.2 项目背景 ..... 1

    1.3 术语 ..... 1

    1.4 期望读者 ..... 1

    1.5 参考文献 ..... 2

第二章 任务分析 ..... 3

    2.1 开发背景..... 3

    2.2 项目目标 ..... 3

    2.3 项目介绍 ..... 3

    2.4 用户类型和特性 ..... 4

    2.5 运行环境 ..... 5

        2.5.1 服务器环境..... 5

        2.5.2 用户环境..... 5

    2.6 设计和实现上的限制 ..... 5

第三章 外部接口需求 ..... 6

    3.1 用户界面 ..... 6

    3.2 软件接口 ..... 6

第四章 系统功能需求 ..... 7

    4.1 系统功能结构..... 7

        4.1.1 数据维护..... 7

        4.1.2 系统管理员功能模块..... 8

        4.1.3 爬虫管理员功能模块 ..... 8

        4.1.4 求职者功能模块 ..... 8

        4.1.5 HR 功能模块..... 9

    4.2 业务建模 ..... 9

        4.2.1 系统管理员用例图 ..... 10

        4.2.2 爬虫管理员用例图 ..... 11

---

4.2.3 求职者用例图 .....	12
4.2.4 HR 用例图 .....	13
4.3 分析模型 .....	14
4.4 数据说明 .....	14
4.4.1 相关的数据流说明 .....	14
4.2.2 数据流图 .....	14
第五章 非功能需求 .....	15
5.1 安全性要求 .....	15
5.2 美观性要求 .....	15
5.3 简捷性要求 .....	15
5.4 性能需求 .....	15
5.5 用户要求 .....	15
5.6 软件质量属性 .....	15

## 第一章 引言

### 1.1 目的

本文档是根据目前求职需求信息化的发展进程,依据用户对匹配精度高、时效性强、区分度好的智能岗位推荐系统的迫切需求,经过详细分析后整理出来的,描述开发招聘定制系统的业务流程和业务需求的文档。本文档的目的的一方面是指导项目成员了解最终的业务需求,另一方面是作为后续开发设计业务需求的指导文档。

### 1.2 项目背景

现今市场上的招聘网站通常是作为连接企业雇主和求职者的信息中介平台,对于求职者的求职决策和企业 HR 的人才招聘都没能起到除了信息中介之外更进一步的辅助性帮助,从而导致了招聘效果往往不尽人意,增加了求职者的试错概率,造成了企业人力资源和人才本身的浪费,整个行业正在渴求高效的求职招聘工具。

此外,此外,近几年,随着大数据技术的发展?如何将大数据分析技术与人岗匹配相结合?

海投简历,广撒网往往是求职者最常用的求职方法,但是传统的求职方式费时费力,而且信息容易泄露,对自身个人信息安全产生威胁。

基于这样的行业痛点,职达信息技术有限公司构建了——基于大数据深度学习技术的人岗匹配系统“职达”,旨在为用户精准匹配推荐岗位,精准人才推荐,辅助求职者理性选择职业岗位,助力 HR 快速筛选岗位最佳候选人,使系统具备智能科学分析能力,助力精准求职,实现有效招聘,具有当代实用性和商业价值。

### 1.3 术语

本文档所涉及的一些专业术语及英文缩写如下:

数据:泛指表示一个指定的值或条件的数字、符号(或字母)等。数据是表示信息的,但这种表示要适合传输、分析和处理。此处,常把数据当作信息的同义词。

Hadoop: Hadoop 是一个开源的框架,可编写和运行分布式应用处理大规模数据

MongoDB: 是一个是专为可扩展性,高性能和高可用性而设计的数据库。

Node: 是一个基于 Chrome V8 引擎的 JavaScript 运行环境。

JSON: JSON(JavaScript Object Notation) 是一种轻量级的数据交换格式。

### 1.4 期望读者

本软件产品需求分析报告所针对的预期读者包括:

1、用户:只需关注项目分析、系统功能描述、项目运行环境,根据项目硬件和软件

要求配置自己的计算机已达到运行项目的条件。

2、开发人员：严格按照此文档实施，需要关注项目的系统功能详细分析，系统流程图，数据流图及数据库设计功能，了解项目中的各子系统之间的关系，各功能之间的处理及实现要求。

3、测试人员：以此文档作为测试的依据，需要关注项目的具体功能实现，了解项目报告中的系统功能要求，根据测试结果判断是否达到系统基本的要求。

4、项目经理：依照此文档安排项目进度及人员安排。

5、文档编写人员：需关注项目报告中的每一项任务，针对报告中的每个模块，需要仔细查看，力争将系统的功能，项目分析，后台描述，项目参与人员及整个项目的架构清晰地抽象出来以展示给各种不同的预读人员。

## 1.5 参考文献

[1]李冰,王悦,刘永祥.大数据环境下基于 K-means 的用户画像与智能推荐的应用[J].现代计算机(专业版),2016(24):11-15.

[2]刘海,卢慧,阮金花,田丙强,胡守忠.基于“用户画像”挖掘的精准营销细分模型研究[J].丝绸,2015,52(12):37-42+47.

[3]席岩,张乃光,王磊,张智军,刘海涛.基于大数据的用户画像方法研究综述[J].广播电视信息,2017(10):37-41.

[4]孟巍,吴雪霞,李静,王婧,杜颖,梁雅洁,王艳玫,徐美玲,林晓兰.基于大数据技术的电力用户画像[J].电信科学,2017,33(S1):15-20.

[5]胡媛,毛宁.基于用户画像的数字图书馆知识社区用户模型构建[J].图书馆理论与实践,2017(04):82-85+97.

[6]付继娟.人与岗位匹配的国内外研究综述[J].武汉职业技术学院学报,2004(02):40-43.

[7]史东风.基于岗位胜任力的石油企业中层管理者人岗匹配模型研究[D].西南石油大学,2011.

[8]张志宇,吕明丽,李从东.基于 BP 神经网络的人岗匹配测评模型的研究[J].天津大学学报(社会科学版),2010,12(05):390-395.

[9]袁珍珍,卢少华.BP 神经网络在人岗匹配度测算中的应用[J].武汉理工大学学报(信息与管理工程版),2010,32(03):515-518.

[10]徐锦阳,张高煜,王曼曦,楼焕钰,薛伟程,毛麒麟.招聘网站职位与简历的双向匹配相似度算法[J].信息技术,2016(08):43-46+51.

[11]成功,李小正,赵全军.一种网络爬虫系统中 URL 去重方法的研究[J].中国新技术新产品,2014(12):23.

[12] 王桦. 基于广度优先的主题爬虫的设计与实现[D]. 复旦大学, 2011.

[13] 吴小惠. 分布式网络爬虫 URL 去重策略的改进[J]. 平顶山学院学报, 2009, 24(05): 116-119.

## 第二章 任务分析

### 2.1 开发背景

随着网络招聘的推广和普遍, 越来越多的人通过招聘网站寻找和选择工作, 同时企业也因为网络招聘的范围广泛、便捷、高效性而开始青睐这一渠道。近几年, 随着大数据技术的发展, 爬虫技术、云计算、文本挖掘、统计分析等也开始初步应用到网络招聘领域。而现今市场上的招聘网站通常是作为连接企业雇主和求职者的信息中介平台, 对于求职者的求职决策和企业 HR 的人才招聘都没能起到除了信息中介之外更进一步的辅助性帮助, 从而导致了招聘效果往往不尽人意, 增加了求职者的试错概率, 造成了企业人力资源和人才本身的浪费, 行业渴求高效求职招聘工具。

### 2.2 项目目标

基于这样的行业痛点, 职达信息技术有限公司构建了——基于大数据深度学习技术的人岗匹配系统“职达”, 旨在为用户精准匹配推荐岗位, 精准人才推荐。辅助求职者理性选择职业岗位, 助力 HR 快速筛选岗位最佳候选人, 使系统具备智能科学分析能力, 具有当代实用性和商业价值, 助力精准求职, 实现有效招聘,

### 2.3 项目介绍

基于大数据深度学习技术的人岗匹配系统“职达”, 旨在为用户精准匹配推荐岗位, 精准人才推荐。辅助求职者理性选择职业岗位, 助力 HR 快速筛选岗位最佳候选人, 使系统具备智能科学分析能力, 助力精准求职, 实现有效招聘。

信息综合模块主要是爬虫管理员增添爬取网站, 设置爬重频率、代码, 查看历史信息等, 保证系统信息最新最全, 是系统的数据主要来源, 是系统最必不可少的基础功能。

人岗信息匹配模块主要以求职者和 HR 两个角色划分。求职者可上传简历信息, 系统采用基于爬虫数据自主建立的人岗匹配模型, 为求职者推荐匹配度高的职位; HR 可批量上传建立, 设置岗位信息及权重, 利用匹配算法, 将 HR 上传的职位候选人继续匹配度排序。这是求职招聘高效化的关键一步。

信息画像模块主要包括求职者画像、岗位画像、人岗匹配画像、行业画像、公司画像、面对岗位的个人竞争力分析报告六大主要画像, 画像皆基于系统数据分析结果, 科学全面, 与信息匹配部分同步, 是平台功能框架的核心。全方位的信息分析, 帮助提高求职效率成功率, 减少在求职招聘市场上的无用工作量。

辅助求职模块主要包括求职社区交流与面向岗位的职业指导, 求职社区中, 用户可

进行信息交流、经验交流，为用户拓宽求职渠道，增加职场人脉基础；职业指导是面向用户已投递岗位的针对性职业辅导，包括面试技巧、笔试真题、简历指导等多方面信息，助力提高用户求职成功率。

特色工具模块主要包括简历智能信息识别、分析报告在线保存、岗位信息来源查看、公司安全评分、多岗位简历一键投递、职业指导资料在线保存等等系列为用户考虑的功能，系统从用户角度出发，将人性化的服务理念体现在每一个细节处，极力为用户带来最大的便利，使用户的求职路没有后顾之忧，给求职招聘市场带来新气象！

## 2.4 用户类型和特性

本系统主要给求职者使用，本系统的主要面向以下几类客户：

**系统管理员：**进行角色管理，修改和删选增添用户信息，分配用户角色权限、负责数据库管理及运维、数据安全备份还原等操作，保障信息安全。

管理员是系统的后台总体管理人员，负责系统的正常运行和问题处理，能够根据需要对用户进行管理、对数据进行及时的备份还原查看等操作。

**爬虫管理员：**对所爬虫网站的增加删减，设置爬取网站频率，自定义爬虫代码，近期任务及时查看，爬取信息内容查看管理。

爬虫管理员是系统数据信息的管理者，保障系统岗位信息最新最全，为系统画像分析、匹配等提供数据基础。

**求职者用户：**登陆后上传简历文档，匹配合适岗位，查看岗位详细信息和来源，查看面向岗位的人岗匹配、公司、行业画像，以及个人竞争力分析，多岗位一键投递简历，查看职业指导，参与求职论坛交流。

求职者是平台的主要用户，包括各大中小城市的各年龄段、各层次所有求职者群体，均可在职达平台中体验到前所未有的便利高效一站式求职服务。

**HR 用户：**批量上传收到的简历，添加岗位信息，设置匹配权重，匹配岗位候选人。

HR 是招聘市场的人群，这类用户数量大，HR 每天都会收到大量的简历信息，逐一查看的工作量巨大，他们对系统的使用频率较高，因此要求前台功能使用设计简单，易操作，同时要求系统具有较高的并发性和稳定性，能够帮助其减少简历筛选工作量。

表 2-1 用户类型和特性表

编号	用户类别	特征描述
1	系统管理员	系统设置，数据库运维
2	爬虫管理员	设置爬虫方案、历史数据信息查看、维护信息又新又全



3	求职者	使用目标明确，期待求职高效，简历上传，匹配职位，查看简历信息，投递简历，求职指导，求职社交
4	HR	批量上传简历，设置岗位信息，候选人匹配

## 2.5 运行环境

### 2.5.1 服务器环境

硬件环境：Intel Pentium4 Xeon 2.5GHz

100GB Raid 5 磁盘阵列

支持环境：Node + nginx + JDK1.8 + Hadoop2.7.5

数据库：MongoDB 3.6.0

### 2.5.2 用户环境

硬件环境：CPU 1.0G 以上

内存 256M 以上

硬盘 20G 以上

操作系统：Windows、Linux、Android、IOS 等主流操作系统

支持环境：建议使用 IE9.0 以上及 Chrome、Firefox、Safari、Opera 等主流浏览器

## 2.6 设计和实现上的限制

本项目系统基于 Node.js 技术，并结合运用 express 框架整合开发，在实现部分的子功能上采取了 Ajax，Vue 等技术，构建了 SPA 单页面应用，以实现更好的用户体验。

开发人员需掌握的技术：

- 1、Node.js 技术熟练应用
- 2、Sublime Text 熟练使用；
- 3、npm 包管理技术
- 4、MongoDB 熟练应用；
- 5、Ajax，Vue 技术基本应用；
- 6、基本的 CSS，Javascript 技术；





## 第三章 外部接口需求

### 3.1 用户界面

根据用户操作需要，设计简洁美观的、友好的用户界面，为用户提供简单、直观的操作界面。

主要遵循以下原则：

- 1、界面元素符合用户习惯；
- 2、操作简易性；
- 3、界面一致化；
- 4、以用户体验至上为原则。

### 3.2 软件接口

服务器程序使用 `mongoose` 提供的对 `MongoDB` 的接口进行对数据库的访问。

Web 前端通过 `ajax` 对后端发起请求，后端通过与数据建立连接，进行数据的传输。

## 第四章 系统功能需求

### 4.1 系统功能结构

#### 4.1.1 数据维护

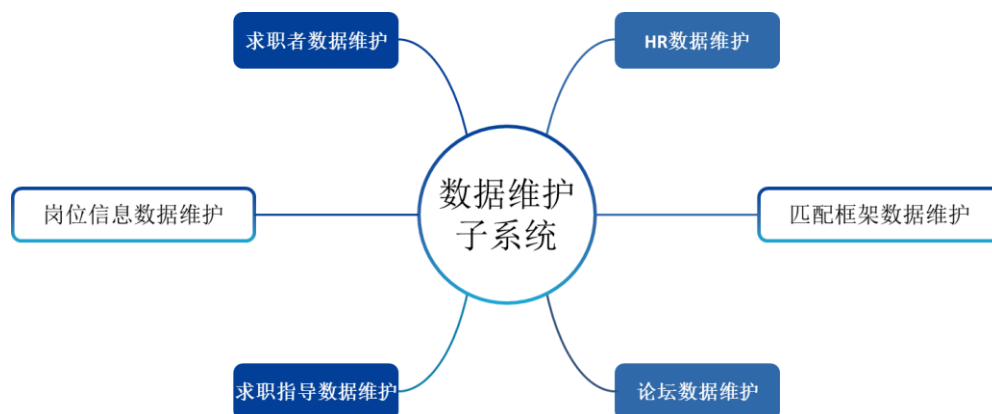


图 4-1 数据维护系统图

##### 1、求职者数据维护

对求职者在平台上的数据的导入、修改与查看。求职者数据主要包括：姓名、学历、性别、应聘岗位、登录密码、技能、工作经验等详细信息。若用户拥有权限，可在该模块对单独求职者数据进行修改和查看，也可以对全部求职者数据进行新增、修改、删除操作。

##### 2、HR 数据维护

对招聘人员的数据导入、修改与查看。HR 数据主要包括：HR 姓名、HR 个人登录密码、HR 批量导入的简历信息、HR 设置的岗位信息等详细信息。若用户拥有相应的权限即可在该模块对单独 HR 数据进行修改和查看，也可以对所有 HR 数据进行新增、修改和删除操作。

##### 3、岗位信息数据维护

对经爬虫获取的岗位相关数据的导入、修改与查看。岗位相关数据主要包括：职位名称、职位所需求求职者学历、岗位信息要求、相应公司信息、相应行业信息等详细信息。若用户拥有相应的权限即可在此模块对单个岗位数据进行修改和查看，也可以对全部职位数据进行新增、修改和删除操作。

##### 4、匹配框架数据维护

对人岗匹配框架等数据进行导入和修改与查看。匹配框架主要包括：匹配项、指标权重、用户反馈数据等详细信息。若用户拥有相应的权限即可在此模块可对信息进行新增修改和删除。

##### 5、求职指导数据维护



对各岗位对应的求职指导信息数据进行导入和修改与查看。求职指导数据主要包括：岗位名称，面试咨询，笔试题目，简历指导等详细信息。若用户拥有相应的权限即可在此模块可对单个岗位信息进行修改和删除，也可以对平台所有求职指导数据进行新增和修改，删除操作。

#### 6、论坛数据维护

对论坛信息数据进行查看、管理和修改。论坛数据主要包括：用户提问，回帖信息，点赞信息，收藏数据等等详细信息。若用户拥有相应的权限即可在此模块对单个论坛数据进行删除和管理，也可以对平台论坛所有数据进行新增和修改，删除等操作。

### 4.1.2 系统管理员功能模块

管理员可以对系统用户进行用户信息增减、修改，针对不同的用户进行权限设置，方便 HR、求职者对系统的使用高效便捷，也便于管理员对系统的管理维护。

#### 1、用户管理

用户信息管理、权限分配管理

#### 2、数据库管理及运维

管理员可对数据库里的信息进行备份还原，快速响应安全突发事件。

### 4.1.3 爬虫管理员功能模块

爬虫管理员根据 ID 和密码进行登录，组织爬虫程序抓取网站数据，登录后初始化界面，含爬虫部分全部功能。

#### 1、爬虫列表管理

爬虫管理员能够自行设置网站名称、爬取地址、爬取频率，对网站列表增添删减，同时可自定义爬虫代码，实时的全网数据更新，海量数据触手可及。

#### 2、近期任务及结果查看

爬虫管理员可查看近期任务，确定后期爬虫工作，对网站信息库进行全面管理，保障岗位信息最新最全。

### 4.1.4 求职者功能模块

#### 1、登陆及信息填写

求职者登录后登陆后可进入初始化界面，之后选择用户图标进入信息填写页，用户可以通过添加简历上传，系统利用文本分析智能信息识别添加，也可由用户自主手动填写。

#### 2、查看匹配结果

求职者完善自身信息后点击确认即进入匹配完成界面。各岗位信息按匹配度高低进

行排列。

### 3、画像查看

用户可进入岗位匹配情况详细页，依次查看面向该岗位的人岗匹配画像，行业画像以及公司画像，并根据系统职达邮局中关于该岗位的用户信息，形成个人竞争力分析，且支持导出为 PDF 格式。

### 4、简历一键投递

系统为求职者设计了简历一键投递功能，用户可进入职达邮局，查看之前浏览过程中添加至职达邮局的岗位信息，可对岗位进行勾选，选择“一键投递”，将用户简历信息通过来源网站投递至 HR 处。

### 5、查看求职指导

用户可在职达邮局中查看历史投递的信息，并查看面向历史投递岗位的各画像分析报告和求职指导，求职指导资料也支持在线保存。

### 6、求职经验交流池

用户可点击论坛交流，进入互动交流平台，学习求职经验。

## 4.1.5 HR 功能模块

### 1、批量导入简历信息

HR 登陆后，可批量导入收到的简历信息至职达平台，系统利用文本分析智能信息识别添加候选人信息。

### 2、设置岗位信息

HR 可设置岗位要求、权重，以便后续与候选人信息进行匹配。

### 3、匹配候选人排序

HR 添加岗位要求后，可选择执行匹配，系统将根据候选人信息与岗位信息，将候选人按匹配度高低依次排序，辅助 HR 快速筛选候选人。

### 4、论坛交流

用户可进入互动交流平台，学习分享行业资讯，拓宽职场人脉。

## 4.2 业务建模

针对系统的几类用户，分别建立如下用例图：

#### 4.2.1 系统管理员用例图



图 4-2 系统管理员用例图

表 4-1 系统管理员用例表

用例编号：01
用例名称：系统管理员用户
主要参与者：系统管理员
项目相关人员：
前置条件：用户已登录进该系统
后置条件
基本事件流：角色管理，系统参数设置，数据库备份与恢复，用户权限管理
扩展事件流：
特殊需求：
异常处理：当输入信息出错时给出提示信息，并返回重新输入

4.2.2 爬虫管理员用例图



图 4-3 爬虫管理员用例图

表 4-2 爬虫管理员用例表

用例编号：02
用例名称：爬虫管理员用户
主要参与者：爬虫管理员
项目相关人员：
前置条件：用户已登录进该系统
后置条件
基本事件流：爬虫列表管理、近期任务与结果查看
扩展事件流：
特殊需求：
异常处理：当输入信息出错时给出提示信息，并返回重新输入

4.2.3 求职者用例图



图 4-4 求职者用例图

表 4-3 求职者用例表

用例编号：03
用例名称：求职者用户
主要参与者：求职者
项目相关人员：
前置条件：用户已登录进该系统
后置条件：
基本事件流：登陆及信息填写、查看匹配结果、画像查看、简历一键投递、求职经验交流池
扩展事件流：
特殊需求：





异常处理：当输入信息出错时给出提示信息，并返回重新输入

#### 4.2.4 HR 用例图

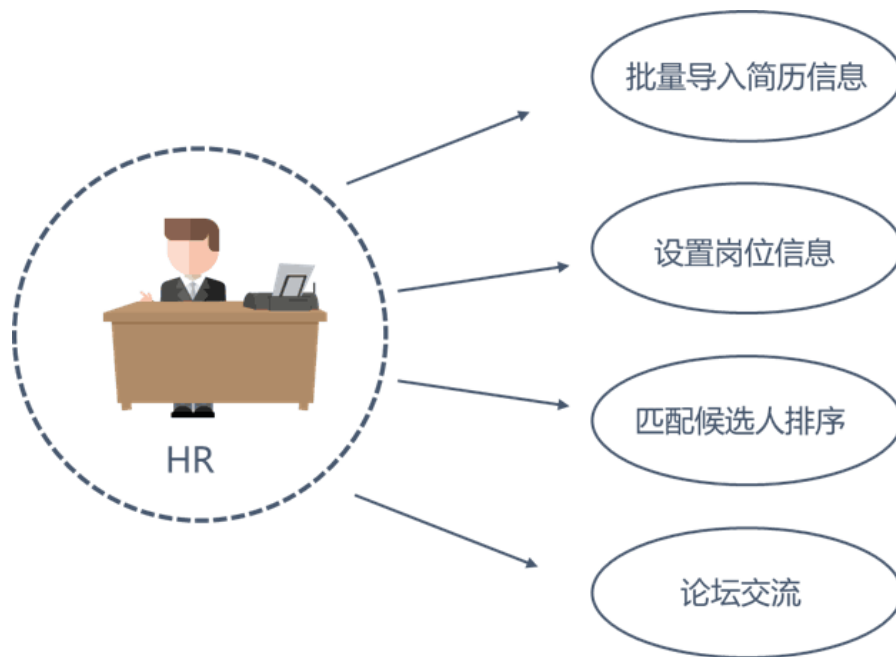


图 4-5 HR 用例图

表 4-4 HR 用例表

用例编号：03
用例名称：HR 用户
主要参与者：HR
项目相关人员：
前置条件：用户已登录进该系统
后置条件：

基本事件流：批量导入简历信息、设置岗位信息、匹配候选人排序、论坛交流

扩展事件流：

特殊需求：

异常处理：当输入信息出错时给出提示信息，并返回重新输入

### 4.3 分析模型

以数据流图对本系统进行分析。

### 4.4 数据说明

#### 4.4.1 相关的数据流说明

1、求职者基本数据=求职者姓名+求职者毕业院校+求职者学历+求职者专业+求职者性别+求职者工作经验+求职者技能+求职者邮箱+求职者电话+密码+上次登录时间。

2、HR 基本数据= HR 姓名++HR 岗位+HR 邮箱+HR 电话+密码+上次登录时间。

3、系统管理员基本数据=管理员工号+管理员姓名++管理员邮箱+密码+上次登录时间。

4、爬虫管理员信息=管理员工号+管理员姓名++管理员邮箱+密码+上次登录时间。

5、岗位数据=岗位名称+岗位类别+公司名称+公司类别+所在城市+待遇薪资+招聘人数+所需技能+岗位福利+招聘人数+发布日期。

6、公司信息=公司名称+公司评分+公司简介。

#### 4.2.2 数据流图

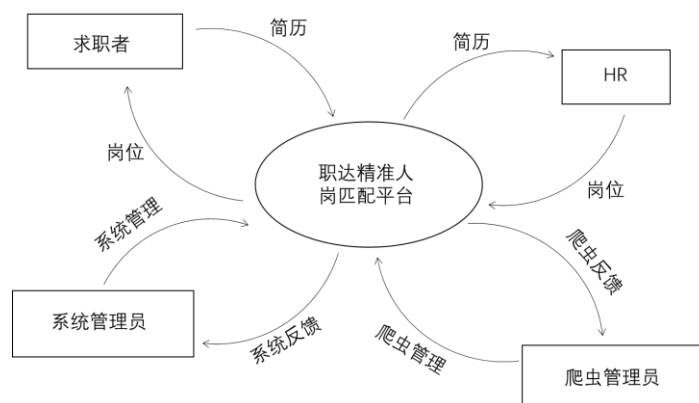


图 4-6 HR 数据流图

## 第五章 非功能需求

### 5.1 安全性要求

“职达”——基于大数据深度学习技术的人岗匹配系统，本平台涉及用户的隐私，存放着极为重要且敏感的数据，对安全性能有极为严格的要求。在本平台中，主要设计到数据的安全保护、用户权限检查、数据传输安全、数据备份恢复机制以及容灾方案的设计等。

### 5.2 美观性要求

界面要简洁美观，避免过于繁乱复杂。

### 5.3 简捷性要求

根据用户操作需要，设计简洁美观的、友好的用户界面，为用户提供简单、直观的操作界面。

### 5.4 性能需求

1、在最大并发量访问下，系统仍能提供服务，不出现致命性错误。

2、一般操作的响应时间应在 3~5 秒内，响应时间也应在可接受的时间内完成，具体情况以网络速度为限制；

### 5.5 用户要求

具备一定的计算机操作能力；

具备上网操作能力；

具备微信的基本操作能力。

### 5.6 软件质量属性

1、健壮性：系统应能够长时间无故障运行；

2、易用性：界面应采用图形化操作方式，便于业务人员操作，保证中专以上文化程度的人员可以无障碍操作；

3、容错性：用户输入错误数据系统均有相应错误提示信息，确保非法数据不录入系统，具有较好的容错性能；

4、可扩展性：充分考虑到将来需求的修改或增加，采用模块化设计，避免需求变更时大规模修改程序。