**李嘉禾**

地址：辽宁省沈阳市浑南新区东北大学浑南校区，110000

**教育背景**

**东北大学 软件学院 14级软件工程 2014.9至今**

主修课程：操作系统、数据库、数据结构、计算机网络、计算机组成原理、算法原理、编译原理、分布式系统导论、设计模式

**实习经历**

**金山词霸 后台开发组 2017.2.1~2017.2.25**

* 独立负责“金山词霸-名著中英互译功能”后端部分，该功能现已上线，预计适用人数高达千万/月；
* 完成解析给出的epub格式小说中的文本部分，分析XML DOM结构，将其中对应文章的中英文部分做一个对应并保存成json格式返回。最终用户可以再阅读英文名著小说时，点击英文部分自动显示对应的中文。其难点在于格式的不规范。

**项目经历**

**推荐系统（R语言）**

* 利用基于用户的协同过滤算法分析输入集中具有相似特性的用户集，找出他们所关注的商品集作为该用户群体的共同商品，对于该用户集中的每一个用户，向其推荐商品集中他们没有发生过关系的商品。

**PeopleRank排名系统**

* 拓展Google网页排序中所使用的PageRank算法的使用场景。原始数据包含用户与用户间的社交指向关系，通过计算其关系的邻接矩阵及概率矩阵，计算并排名用户影响力。使用了MapReduce计数，使其适用于大批量数据的离线计算。

**网站KPI数据统计**

* 内容包括了：browser（用户使用的浏览器统计），ips（页面用户独立ip数统计），pv（网站pv量统计）source（用户来源网址统计），time（时间段用户访问量统计）等。以用户使用浏览器统计为例，在Mapper中将用户原始数据中的Browser项提取出来及ip提取出来，作为key-value发送给Reducer，然后在Reducer中对数据进行计数统计，最终可以得到不同浏览器浏览的人次、和ip数等信息，该信息最终保存在以Hive为基础的数据仓库中，提供数据的存储、查询、分析。

**电信运营商用户停留数据分析**

* 该功能涉及到对用户设备定位、轨迹计算、是否在线判定等方面的计算，原始数据为包含了用户IMSI、位置信息、时间戳、基站信息等的多张表格，从中取出相应的值作为Key-Value，如在计算用户的基站停留时间时，可以使用用户IMSI作为Key，基站及时间信息作为Value，发送到同一Reducer，排序后计算时间差，统计在不同基站的停留时间等。

**基于Flume-Kafka-Storm-HBase的风力发电厂风机日志获取分析及存储系统**

* 这是一个利用已有可公开数据进行的模拟项目。首先利用Flume从指定数据源获取数据，Sink接入到Kafka中，利用Kafka作为一个数据的缓冲，以Storm作为Consumer消费数据，主要是利用Storm的窗口机制，分析指定时长内风机油温是否偏离正常值，从而对油温进行实时监控、预警。

**语音天气查询小助手**

* 使用Java实现，利用科大讯飞的免费语音识别接口，获取用户语音的文字内容，将得到的内容进行语义分析，调用中国天气网的天气接口返回查询结果。

**贪吃蛇小游戏**

* 用Java实现的一个由圆点组成的小蛇，在经过界面上随机生成的小方块时，长度增长1，并可以在读取键盘的上下左右输入之后做出相应的方向转变。利用状态模式，使代码更加清晰的同时，将位置信息的改变转移到状态中，使得相同计算成本总量被分摊，保证游戏的低延迟。

**职业技能**

1. **完成大数据方向课程学习，包括：Hadoop离线计算（包括Map-Reduce，Hive，HBase，Zookeeper等组件），Storm实时流计算（包括Flume日志框架、Kafka、Storm实时数据处理等），Spark内存计算，以及数据挖掘基础（包括基础理论及算法、R语言以及Python的使用）。**
2. 熟悉Linux系统常规shell处理命令，灵活运用shell做的文本处理和系统操作。
3. 熟练掌握Java技术，了解Struts1、Struts2、SpringMVC、Spring、Hibernate等主流。
4. 熟悉Hadoop2.0 、YARN、MapReduce、HBase、Zookeeper、Hive、Spark、Kafka、Storm、Flume。并能对Mapreduce、HBase应用进行设计、优化和开发。
5. 熟练使用MAVEN项目开发及搭建。
6. 熟练运用MySQL数据库，了解Oracle等主流数据库。
7. 熟悉HTML、Javascript等前段编程语言。
8. 熟练使用PHP、Python、R等脚本语言进行开发。

**自我评价**

1. 对于大数据、数据挖掘、机器学习、人工智能等方面具有相当浓厚的兴趣，长期以来没有机会接触，今年有幸加入了学校的大数据班，获益匪浅。
2. 短暂的实习经历
3. 对行业的动态和未来发展方向具有一定自我的见解与认知。对众多互联网公司、科技产业近年的发展都有着一定程度的了解和自己的判断，希望能在国内兴起的大数据浪潮中，展现自己真正的实力:)。
4. 兴趣使然的软件开发经历使我在长期的开发过程中收获了非常多的乐趣，包括自己闲着没事干做的一些小游戏小网站都是有趣的经历。
5. 大学英语6级 545分，托福 95分，较强的英语能力能帮助我更快的学习和阅读文档。
6. 在校期间曾任班级副班长、社团联合会干事。较强的沟通能力和团队协作能力，有能力在短时间进行较强的学习，完成团队中相应的责任部分。