基于互联网数据的大数据分析

详细设计说明书

引言

1. 编写目的

信息科技经过 60 余年的发展，已经从高端科技普及到社会运转的每一个角落中。随着信息技术在国家治理、经济运行的方方面面的运转，大量的数据随之产生。而互联网技术的爆发式发展使得近两年来产生的数据总量超过了人类有史以来的数据总和。如何利用这些数据，挖掘数据的深层价值，是未来信息科技发展的趋势，也是大数据技术产生的背景。因此我们抓取了一些信息，通过集合与计算机相关的职业的招聘数据来探索计算机行业的运行发展规律。

* 1. 适用范围

适用于各高校未毕业学生，让其在校学习时了解企业对人才需求的动向，综合开率专业，知识技能，收入，地域等因素。既进行操作实践，又可以对未来的学习有所规划。  
适用于各高校应届毕业生，其通过本网站可以轻松查询工作相关信息，可更好的对比自身条件，扬长避短，找到更适合自己的工作。

* 1. 参考资料

周志华, 杨强. 机器学习及其应用[M]. 清华大学出版社, 2011.

姜杉彪, 黄凯林, 卢昱江,等. 基于Python的专业网络爬虫的设计与实现[J]. 企业科技与发展, 2016(8):17-19.

1. 总体结构说明
   1. 系统结构
   2. 系统内外部关系图

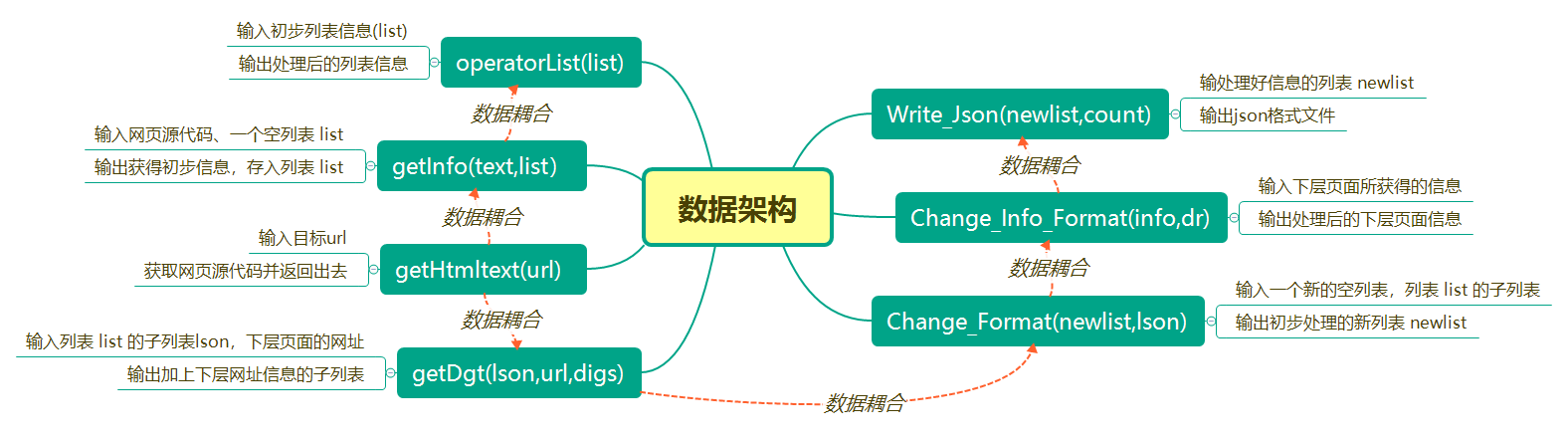


图2.2-1数据架构

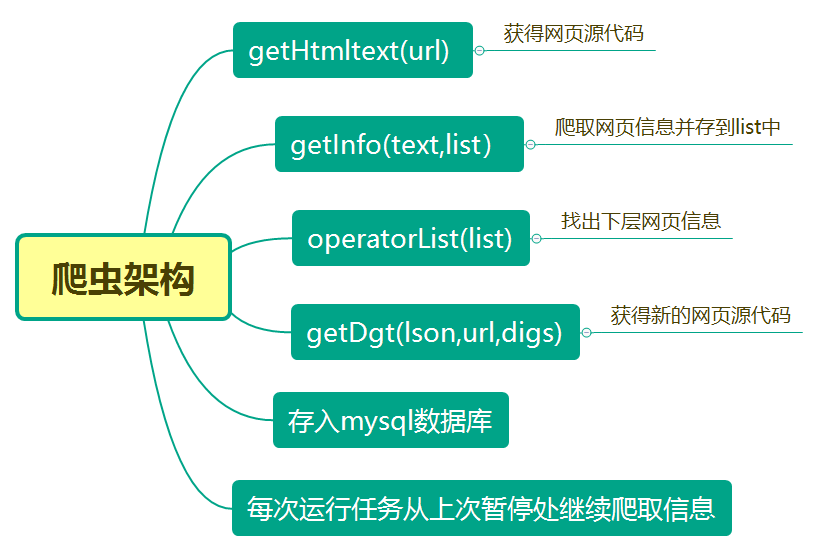


图2.2-2爬虫架构

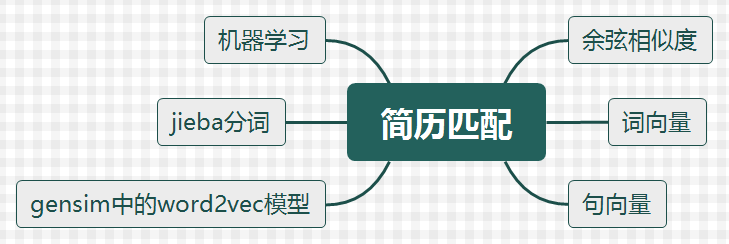


图2.2-3简历匹配

1. 数据模型（Data Model）设计
   1. 逻辑实体模型

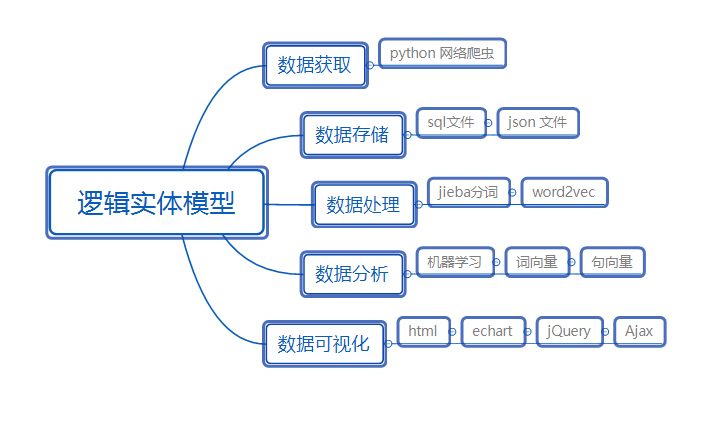


图3.1-1