

主管单位:中国科学技术协会 主办单位:中国电子学会 中电新一代(北京)信息技术研究院

电子技术与软件工程

ELECTRONIC TECHNOLOGY & SOFTWARE ENGINEERING

ISSN2095-5650

CN10-1108/TP

总第173期

02

上半月刊
2020年

移动信息系统应急通信传输模式

基于Attention-BiLSTM的情感分类模型

基于超像素分割的RGB与高光谱图像融合

一种高精度定时器设计方法

TNPC型并网逆变器嵌位开关管失效分析与改进



ISSN 2095-5650



03>

中国期刊网(CNKI)全文入网期刊
龙源国际期刊网全文收录期刊

中文科技期刊数据库收录期刊

9 772095 565139



电子技术与软件工程

2020年2月上半月刊 总第173期

编委

- 国林 原国家铁道部原副部长
中国铁道学会原理事长
张文范 国家民政部原副部长
中国未来研究会理事长
韩德乾 国家科技部原副部长
中国技术市场协会会长
李厚斌 中国老科学技术工作者协会电子工业分会
理事长
李书训 原电子部军工司原正司级巡视员
中国老科学技术工作者协会电子工业分会
副理事长
徐顺成 原电子部科技司原司长
中国老科学技术工作者协会电子工业分会
副理事长兼秘书长
林元芳 中国电子视听协会秘书长
张宏民 《中国电子报》社原社长
中国计算机协会副秘书长
毛金铸 国家教育部高校司办公室原主任
中国老科学技术工作者协会教育分会副会长
彭大年 北京市科学技术协会联合办公室主任
北京老科学技术工作者总会秘书长
肖泽 中国船舶重工集团公司第七一四研究所
研究员

本刊保留一切版权，如欲转载，请致电编辑部。如有
印刷、装订等质量问题由印刷厂负责调换，电话：(010)
88584136。

本刊已许可中国学术期刊（光盘版）电子杂志社在中
国知网及其系列数据库产品中以数字化方式复制、汇编、
发行、信息网络传播本刊全文。该社著作权使用费与本刊
稿酬一并支付。作者向本刊提交文章发表的行为即视为同
意我社上述声明。

主管单位 中国科学技术协会

主办单位 中国电子学会

中电新一代（北京）信息技术研究院

编辑出版 《电子技术与软件工程》编辑部

社长 刘汝林

副社长 王非

总编辑 熊潘梁

编辑部主任 谭哲

社址 北京市海淀区玉渊潭南路普惠南里13号楼

邮政编码 100036

电 话 (010) 88584136

投 稿 邮 箱 dzyxrjtg@126.com

发 行 代 理 北京功到文化艺术有限公司

地 址 北京市海淀区北洼路9号

电 话 (010) 88583770

网 址 www.eg-soft.com

电 子 邮 箱 dzyxrjgd@163.com

印 刷 北京兰星球彩色印刷有限公司

国 内 刊 号 CN10-1108/TP

国 际 刊 号 ISSN2095-5650

定 价 25.00元

出 版 日 期 每月1日、15日

郑重声明

编辑部从未授权任何单位和个人以《电子技术与软件工程》杂志的名义在网络媒体上刊登《电子技术与软件工程》杂志征稿启事，就有些网络媒体刊登的《电子技术与软件工程》杂志的征稿及显示的联系电话均为假冒。

就有关网站和个人自本刊编辑部发布此声明之日起，请立即停止以我刊名义从事论文征稿的行为，编辑部将保留追究有关网站和个人的法律责任。

特此声明。

《电子技术与软件工程》编辑部

2020年1月

196	基于全局偏移表进行通用动态链接库函数跟踪的方法	张木梁 王国庆 张磊
200	高校思想品德教育工作评价系统的设计与实现	阳晓霞 谭卫 李春来
204	基于 Hadoop 的科技成果转化平台设计与实现	王秀慧
206	基于三支决策的缺失数据处理综述	竺凡超 李红宇
208	高速直升机光电任务系统需求分析	霍苗苗 孙黎静 李雷 谷晓星
210	SQLite 数据库在 Android 平台下的加密设计	汪坤
212	基于 Webservice 的系统信息上报平台的设计与实现	严竞雄

电力电子

215	变电站设备端子箱内部运行工况智能在线监控平台的应用	潘巍
217	一种便携式单相地面电源设计与实现	陆欢 许晶菁
219	电力系统中智能变电站的继电保护技术	董朝理
221	220kV 发电机变压器的结构设计	肖朋 徐春苗
223	基于物联网技术的端子箱驱潮系统设计	赵灿辉
225	基于故障树的某电源模块可靠性	宋志宇
227	基于 AR 的变电动态运维系统研究与设计	蒋奇
229	紧凑相控阵用大功率脉冲电源设计	林庄 倪扬 杨新国
232	长距离超高压输电线路覆冰期地线温度监测	袁华璐 朱一峰 田霖 周黎明 程凌浩
236	正激式开关电源变压器磁饱和机理分析与实验	张石磊 邹扬 张威 刘佳慧
240	基于 DSP 技术的直流伺服电机调速系统设计	周燕菲 袁文玉 孙晓婷 黄毅

信息技术与安全

242	密码技术在财政信息系统中的应用	肖丽辉 张利明
245	网络信息安全叠加风险的识别与应对	张倩 颜丽渊
247	医疗档案管理信息化优势及实现路径	韩勇
249	大数据背景下行政事业单位信息安全保护	陈超
251	数字化校园在中职学校中的应用	陈伟杰
253	解析智能变电站视频监控系统的应用	谢佳伟 马云
255	基于人工智能时代下计算机网络信息安全防范	秦丞 贺渝镇
257	计算机网络数据库的安全管理技术	侯晓凌
259	利用 X-Forwarded-For 伪造客户端 IP 漏洞成因及防护	李治城

基于 Hadoop 的科技成果转化平台设计与实现

王秀慧

(山西大同大学教育科学与技术学院 山西省大同市 037009)

摘要:本文针对科技成果数据呈几何式增长的需求,提出了构建基于 Hadoop 的科技成果转化平台,并详细阐述了该平台的设计与实现。通过此平台可存储海量的科技成果,同时又为企业提供了快捷、精准的科技成果数据推送方式,对企业的科技成果转化率起到了积极的推动作用。

关键词:Hadoop; HDFS; MapReduce; 余弦相似度

1 引言

近些年随着社会信息化的不断发展,信息数据呈现出爆炸式的增长,科技成果数据也不例外。对于如此庞大的数据,单一的存储服务器模式已经不能满足企业的需求,如何存储海量科技成果数据成为亟待解决的问题;同时企业针对海量科技成果数据不知所措,因为无法准确的查找到本身需要的数据,导致了科技成果转化率低,成为制约企业发展的瓶颈。

针对以上问题,结合科技成果数据的特点与企业对科技成果需求的实际情况,构建了基于 Hadoop 的科技成果转化平台。该平台数据是以分布式的结构进行存储,通过扩展相应存储设备来对数据进行数据扩容,并且分布式结构极大地降低了平台的耦合度,使其可独立部署^[1];同时运用余弦相似度算法计算数据之间的相似程度,给企业推送所需科技成果数据以及相关数据。这样不仅提升了企业科技成果转化率,而且还提高了科技成果转化的质量,进而达到使企业核心竞争力增强的目的。

2 总体设计

2.1 体系结构

基于 Hadoop 的科技成果转化平台分为两部分:

(1) 通过 HDFS(分布式文件系统)来存储海量的科技成果数据,同时应用 MapReduce 并结合余弦相似度算法计算数据之间相似情况,以便为企业推送所需数据。

(2) 采用 B/S 结构、按照 JavaEE 标准框架标准设计与实现用户服务,以便企业通过注册登录后可查询所需数据,平台架构图如图 1 所示。

2 关键技术

2.1 HDFS

HDFS 是 Hadoop 项目的核心子项目,是 Hadoop 中的存储组件,由于它涉及到数据存储,MapReduce 等计算模型都要依赖于存储在 HDFS 中的数据,因此它也是最为基础的一部分。HDFS 是一个分布式文件系统,它采用了主从(Master/Slave)结构模式,由 NameNode 和 DataNode 组成。其中 NameNode 是主服务器,统筹管理着文件系统与文件数据的操作,而 DataNode 作为服务节点,管理着数据的存储,其中数据是以块的形式存储,Hadoop2.0 中数据块默认大小是 128MB。

HDFS 的数据访问使用流式模式:一次写入,多次读取,这样更加关注的是读取整个数据集的整体时间,提升了数据的读写吞吐量,但由于流式读取,无法高效存储大量小文件,也不支持多用户写入,写操作总是以添加形式在文件数据末尾追加,并且还无法在任意位置进行修改。

2.2 MapReduce

MapReduce 是一种编程模型,适应于大规模数据的并行运算。主要思想是通过 Map(映射)把一组键值对映射成一组新的键值对,并自动划分计算数据和计算任务,然后指定并发的 Reduce 函数来收集计算结果并再次规约计算,使得所有映射的键值对中的每一个



图 1: 平台架构图

共享相同的键组。这样编程人员在没有掌握分布式并行编程的情况下,就可以将自己的程序运行在分布式系统上,同时数据的处理、通信和运算等操作都交由 MapReduce 负责处理,大大的减少了软件开发人员的工作量。

2.3 余弦相似度

余弦相似度^[2]是一种数学模型,用于计算两个向量的夹角角度,0 度角的余弦值是 1,90 度角的余弦值是 0,而其他任何角度的余弦值都不大于 1,180 度角的余弦值是 -1,且是其最小值。由于余弦相似度的特性,人们发现其计算数据间的相似程度有着天生的优势,而且计算比较简单方便,因此在计算文本数据之间相似度的过程中得到了广泛的应用,同时规定其取值范围是 0 到 1 之间。其原理是:把文本数据转化成向量数据,即文本含有的特征项的权重来表示向量数据。设文本的向量表示: $d_i = (w_1, w_2, \dots, w_p, \dots, w_n)$, w_p 表示某文本特征项的权重值, $d'_i = (w'_1, w'_2, \dots, w'_p, \dots, w'_n)$, 则 d_i 与 d'_i 之间的余弦相似度计算公式如下:

$$Sim(d_i, d'_i) = \frac{\sum_{i=1}^n d_i * d'_i}{\sqrt{\sum_{i=1}^n d_i^2} * \sqrt{\sum_{i=1}^n (d'_i)^2}}$$

3 功能设计

3.1 设计目标

●基金项目: 大同市科技计划基金项目: 大同市科技成果转化信息服务平台建设(2018187)。

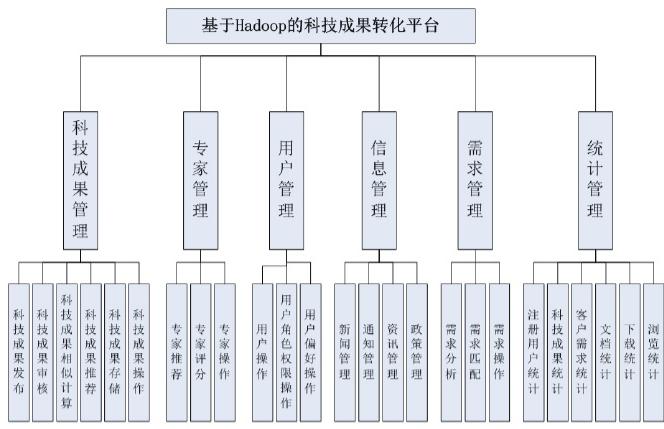


图 2: 平台功能模块图

根据科技成果转化平台的实际需求进行功能设计，主要完成以下几个目标：

- (1) 上传各类科技成果数据，并审核；
- (2) 展示平台各类科技成果数据；
- (3) 浏览查看科技成果相关文档；
- (4) 浏览查看科技成果相关专家；
- (5) 推送科技成果数据；
- (6) 计算科技成果数据间相似度；
- (7) 科技成果数据分布式存储；
- (8) 提交企业用户的需求信息，并审核；
- (9) 注册、登录平台；
- (10) 浏览查看平台新闻、咨询、政策、通知等。

3.2 主要功能

该平台主要功能包含科技成果管理、用户管理、专家管理、信息管理（其中包括新闻管理、通知管理、资讯管理、政策管理）、需求管理、统计管理等。平台主要功能如图 2 所示。

3 平台实现

3.1 平台开发环境

平台选用 Hadoop2.7 作为科技成果数据存储介质，并承载数据间相似度计算；同时应用 Jdk9.0 配置 Java 开发环境，运用 Tomcat8.0 作为 Web 服务器，数据库使用 MySql5.5，页面设计工具使用 Dream weaver2018，程序开发工具运用 MyEclipse2014。

3.2 主要实现代码

```
class RecordSimilar{
    public double caculateSimilar(Map hashMapOne, Map
hashMapTwo){
        Map hashMap=consistVetor(hashMapOne, hashMapTwo);
        Set set[hashMap.keySet()];
        List list3=consistEigenvalue(set, hashMapOne);
        List list4=consistEigenvalue(set, hashMapTwo);
        double totleHashMapOne =0;
        double totleHashMapTwo=0;
        double totleSum=0;
        for(int i=0;i<list3.size();i++){
            double e1=((Double)list3.get(i)).doubleValue();
            double e2=((Double)list4.get(i)).doubleValue();
            totleHashMapOne+=e1*e1;
            totleHashMapTwo+=e2*e2;
            totleSum+=e1*e2;
        }
    }
}
```

```
double similar=totleSum/((Math.sqrt(totleHashMapOne))*(
Math.sqrt(totleHashMapTwo)));
//double similar=totleSum/(Math.sqrt(totleHashMapOne)+Math.sqrt(totleHashMapTwo)-totleSum);

return similar;
}

public Map consistVetor(Map hashMapOne,Map hashMapTwo){
    Map hashMap= hashMapOne;
    Set set= hashMapTwo.keySet();
    Iterator iterator=set.iterator();
    while(iterator.hasNext()){
        String str=(String)iterator.next();
        if(!hashMap.containsKey(str)){
            hashMap.put(str,new Double(0));
        }
    }
    return hashMap;
}

public List consistEig(Set set,Map hashMap){
    List listEig=new ArrayList();
    Iterator iteratorEig=set.iterator();
    while(iteratorEig.hasNext()){
        String strEig=(String) iteratorEig.next();
        if(hashMap.containsKey(strEig)){
            Double douEig=(Double)hashMap.get(strEig);
            listEig.add(douEig);
        }else
            listEig.add(new Double(0));
    }
    return listEig;
}
}
```

4 结语

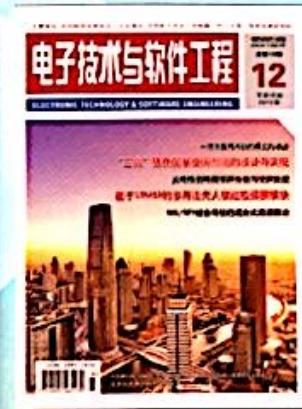
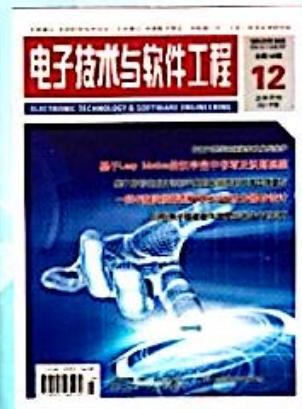
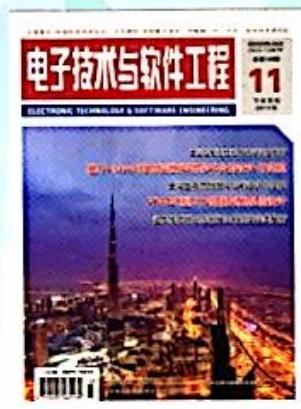
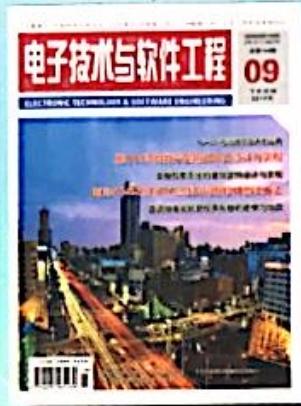
该平台采用 Hadoop、按照 JavaEE 框架标准、使用余弦相似度等信息技术，根据科技成果转化实际需要进行开发，实现了全部设计需求，做到了界面友好、平台运行安全、可操作性强等特点。此平台的实现将科技成果数据以分布式的结构进行存储，解决了海量数据存储难的问题；同时使用余弦相似度计算数据间相似程度，解决了为企业推送所需数据难的问题。我们相信随着此平台在科技成果转化工作中的逐步深入，将会大大提升企业科技成果转化率、提高科技成果转化质量。

参考文献

- [1] 温志芳. 山西科技成果转化平台建设与运营的思考 [J]. 山西电子技术, 2019 (05): 57-60.
- [2] 黄运稳, 陈光, 叶建芳. 基于余弦相似度的加权 K 近邻室内定位算法 [J]. 计算机应用与软件, 2019, 36 (02): 159-162.
- [3] 高凤丽, 孙连山. 个性化推荐系统概述 [J]. 技术市场, 2015, 22 (2): 78-79.
- [4] 赵康, 杨余旺. 基于 Hadoop 的物联网监控系统的设计与实现 [J]. 计算机与数字工程, 2019, 47 (07): 1738-1742.
- [5] 杨凯. 基于 Hadoop 平台的个性化新闻推荐系统的设计与实现 [D]. 北京交通大学, 2019.
- [6] 李庆君. Hadoop 架构下海量空间数据存储与管理 [D]. 武汉: 武汉大学, 2017.

作者简介

王秀慧 (1981-) , 女, 河北省石家庄市人。山西大同大学副教授, 硕士研究生, 计算机应用技术专业。研究方向为数据挖掘、教育技术。



欢迎订阅

2020 全年征订中 咨询电话
010—88584136



电话：010-8858 4136

传真：010-6846 2122

投稿邮箱：dzyxrttg@126.com

网址：www.eg-soft.com

