

蔡宗文, 郑黎辉, 廖丽霞, 等. 基于融媒体体的防震减灾知识挖掘技术与应用[J]. 华南地震, 2023, 43(3): 1-6. [CAI Zongwen, ZHENG Lihui, LIAO Lixia, et al. Knowledge Mining Technology for Earthquake Prevention and Disaster Reduction Based on Media Convergence and Its Application [J]. South China journal of seismology, 2023, 43(3): 1-6]

基于融媒体体的防震减灾知识挖掘技术与应用

蔡宗文¹, 郑黎辉², 廖丽霞², 危福泉², 黄晓华¹, 杨 婕¹

(1. 福建省海洋地震观测中心, 福建 厦门 361009; 2. 福建省地震局, 福州 350003)

摘要: 通过基于大数据融媒体体的防震减灾知识挖掘技术方法研究, 采用词典技术、基于主题模型的智能分词技术, 对所挖掘或已有文章的标题和内容按主题模型进行分词检索和深度分析, 对相应防震减灾知识条目进行分类识别、入库, 从而构建可自我增长的防震减灾知识资源库。利用地震科普和行为指导服务新媒体系统所建立起来的用户流, 实现科普信息即时获取、精准推送, 提高防震减灾科普传播效率。

关键词: 防震减灾; 知识挖掘; 主题模型

中图分类号: P315.9

文献标识码: A

文章编号: 1001-8662(2023)03-0001-06

DOI: 10.13512/j.hndz.2023.03.01

Knowledge Mining Technology for Earthquake Prevention and Disaster Reduction Based on Media Convergence and Its Application

CAI Zongwen¹, ZHENG Lihui², LIAO Lixia², WEI Fuquan²,
HUANG Xiaohua¹, YANG Jie¹

(1. Fujian Provincial Marine Earthquake Observation Center, Xiamen 361009, China;

2. Fujian Earthquake Agency, Fuzhou 350003, China)

Abstract: Through the research on the technology and method of knowledge mining for earthquake prevention and disaster reduction based on big data convergence media, using the dictionary technology and intelligent word segmentation technology based on the topic model, the title and content of the mined or existing articles are subjected to word segmentation retrieval and in-depth analysis, and the corresponding knowledge entries of earthquake prevention and disaster reduction are classified, identified and stored, so as to build a self-growing knowledge resource database of earthquake prevention and disaster reduction. By using the user flow established by the new media system of earthquake science popularization and behavior guidance service, the instant acquisition and accurate push of science popularization information are realized, and the efficiency of science popularization for earthquake prevention and disaster reduction is improved.

Keywords: Earthquake prevention and disaster reduction; Knowledge mining; Topic model

收稿日期: 2023-03-10

基金项目: 国家重点研发计划课题《大用户量地震科普与行为指导新媒体平台研发》(2019YFC1509404)资助

作者简介: 蔡宗文(1970-), 男, 正研级高工, 主要从事震害防御与信息技术开发。

E-mail: caisinfo@126.com

0 引言

我国是世界上地震灾害最为严重的国家之一,分布地域广,发生频率高,造成损失重^[1];防震减灾新技术新成果的科普转化不够,融媒体传播平台建设滞后,高质量科普作品创作与推广不足,防震减灾知识共建共享的核心资源库尚未形成^[2]。

随着基于 Internet 的全球信息系统的发展,数据丰富、知识贫乏已经成为一个典型问题^[3-4]。知识挖掘的目的就是有效地从海量数据中提取出需要的答案,实现“数据→信息→知识→价值”的转变过程^[3-5]。随着云时代的来临,特别是省、地、市地震部门以及社会各界科普创作品日渐丰富,但各部门的科普作品基本都靠各自自媒体平台发布,难于有效及时共享。如何有效获取这些科普作品并形成共享地震科普作品资源库,需要采用大数据技术,以有效地处理大量防震减灾知识数据。知识挖掘技术可以充分地利用广泛分布于互联网上各类融媒体防震减灾知识信息,通过主题模型方法实现知识获取,构建一个可自我增长的防震减灾知识资源库。

因此,适应科技信息化发展大势,通过开展防震减灾知识挖掘及知识精准推荐的相关研究工作,利用地震科普和行为指导服务新媒体平台所建立起来的用户流^[6],引入融媒体知识挖掘技术,对全网融媒体防震减灾科普知识进行挖掘,形成一个自我增长防震减灾知识资源库,极大地丰富了地震科普知识素材,采用面向用户兴趣感知的智能化地震安全咨询与科普知识推荐服务,实现科普信息即时获取、精准推送,能有效提高防震减灾科普传播效率。

1 知识挖掘技术

知识挖掘是从已有的实体及实体关系出发挖掘新的知识,具体包括知识内容挖掘和知识结构挖掘。实体的链接与消歧是知识挖掘重要技术之一,是将文本中的实体指称链向其在给定知识库中目标实体的过程,包括实体指称识别、候选实体生成和候选实体消歧三个步骤。候选实体消歧方法包括基于主题模型的方法、基于图的方法、基于概率生成模型的方法和基于深度学习的方法等^[7-8]。

实体消歧是海量知识挖掘核心步骤,为解决

海量文本表达的歧义性和多样性问题提供技术手段^[9]。基于主题模型的方法是一种高效的特征抽取方法,可以有效地从海量防震减灾知识数据中获取用户所需要的数据提供特征数据;通过对地震知识文档的标题、内容、进行建模,抽取出文档中的隐含主题信息,从而实现融媒体防震减灾知识挖掘。本文采用 PHPAnalysis 中文分词模块法进行主题模型匹配,以知识文本的关键词或主题词为研究单元,统计关键特征词在同一个文本中出现的频率,然后基于词频确定高频关键词构建共词关系矩阵,再结合聚类分析等方法确定文档的主题结构。

2 融媒体防震减灾知识挖掘方法

知识挖掘是从海量数据中发现有用知识的全过程,融媒体知识挖掘就是从各级应急、地震部门通过自媒体平台发布的信息内容,挖掘出有用的防震减灾知识。结合专家经验构建的防震减灾知识分类字典,对指定地震科普相关融媒体的搜索引擎进行网页分析;通过中文分词技术提取相关地震科普知识关键词,利用加权匹配算法对所挖掘或已有文章的标题和内容进行主题模型分词检索和深度分析,输出能够反映文章关键信息的主题、话题、实体等多维度标签以及对应的置信度;对相应防震减灾知识条目分类识别、入库,从而实现全网融媒体防震减灾知识挖掘与共享;最终按照中文分词结果的词频及词条,进行权重加权匹配纳入对应分类库中。图1为融媒体防震减灾知识挖掘和分类过程流程,包括了防震减灾知识挖掘实体指称识别、候选实体生成和候选实体消歧等关键步骤。

2.1 防震减灾知识分类字典建立

防震减灾知识字典建设是防震减灾知识挖掘的基础,采用专家经验法构建防震减灾分类字典以及对应省、地、市优秀防震减灾知识融媒体知识源,为防震减灾知识挖掘实体提供匹配词典。本文采用正向最大匹配、检索树、树尾标签列表实现了词典匹配的实体识别方式,其支持单实体可以对应多标签的情形。

2.2 防震减灾知识搜索与挖掘

在进行融媒体防震减灾知识挖掘时,直接采用融媒体搜索引擎,结合防震减灾知识字典库所

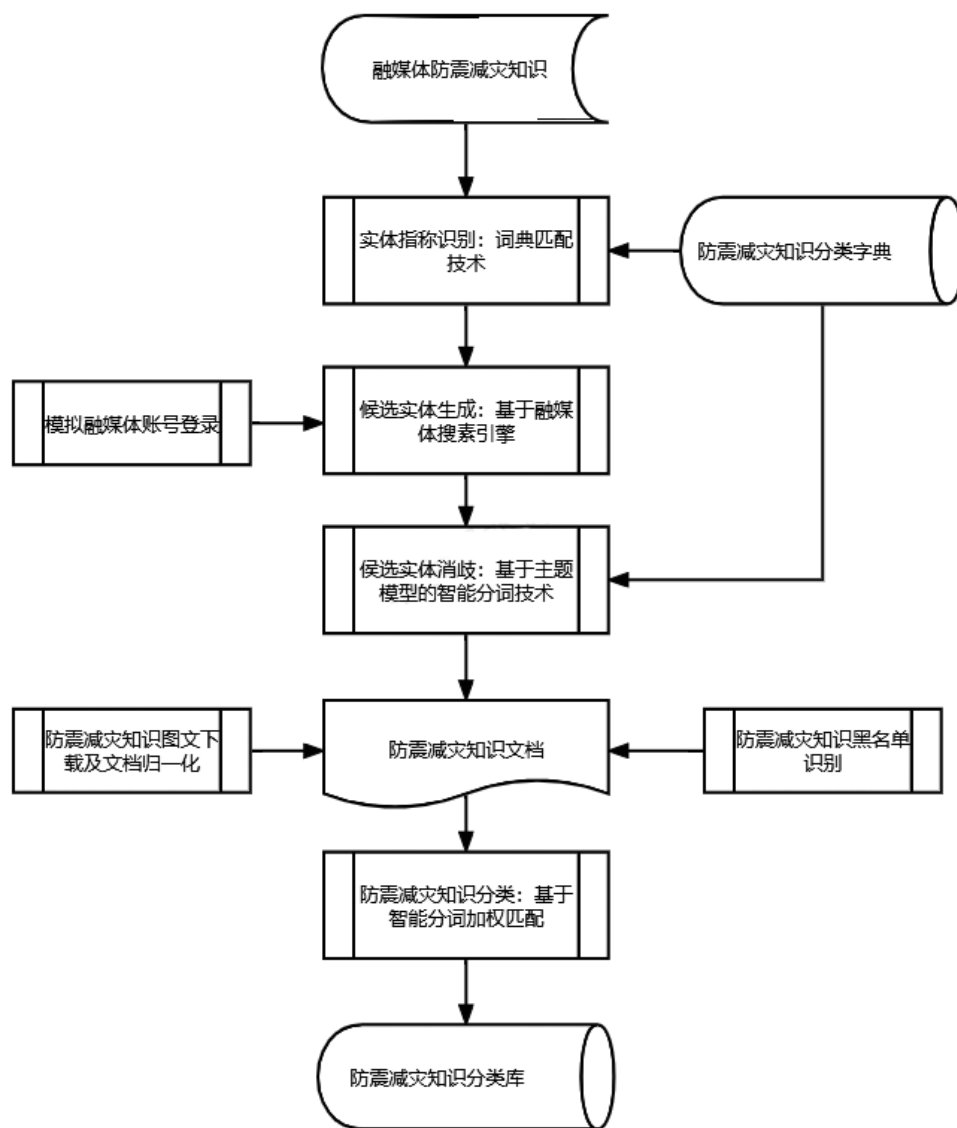


图1 防震减灾知识挖掘与分类

Fig.1 Mining and classification of earthquake prevention and disaster reduction knowledge

对应标签,进行防震减灾知识检索,对所检索知识的标题以及内容进行转码处理,形成统一格式的候选实体,采用中文分词技术确定特征关键词及词频,通过聚类分析方法实现防震减灾知识挖掘。对有登录要求的融媒体平台,通过系统已有的账号及密码进行模拟二维码扫描账号登录。

2.3 防震减灾知识识别与分类

主题检索即将预先定义好的查询语句作为主题。通过防震减灾知识字典中的关键词及其对应权重系数,利用主题检索技术进行查询运算,命中含有权重高的关键词的文档得分高低确定防震减灾知识文档,从而实现防震减灾知识数据的自

动识别。为了更好地识别一些非必要信息,同时构建了防震减灾知识黑名单,对一些可能造成误判知识点进行识别。最终通过图文下载、转码技术、智能分词、加权匹配实现防震减灾知识分类入库。

3 知识挖掘在防震减灾新媒体平台中的应用

3.1 知识挖掘清单

在防震减灾知识挖掘过程中,为了更好地实现融媒体知识获取,使所获取的防震减灾知识更具有科学性,需要预先指定需要挖掘的融媒体名

称及对应账号。通过把知识挖掘技术引入到地震科普与行为指导新媒体平台知识数据库中，预先指定需要挖掘的融媒体名称、要挖掘的信息内容、文章序号等栏目，可以有效提升知识挖掘效率。

图2为地震科普与行为指导新媒体平台中对知识挖掘清单设置实例，系统在知识挖掘过程中会根据所设定清单自动挖掘最新的防震减灾知识。

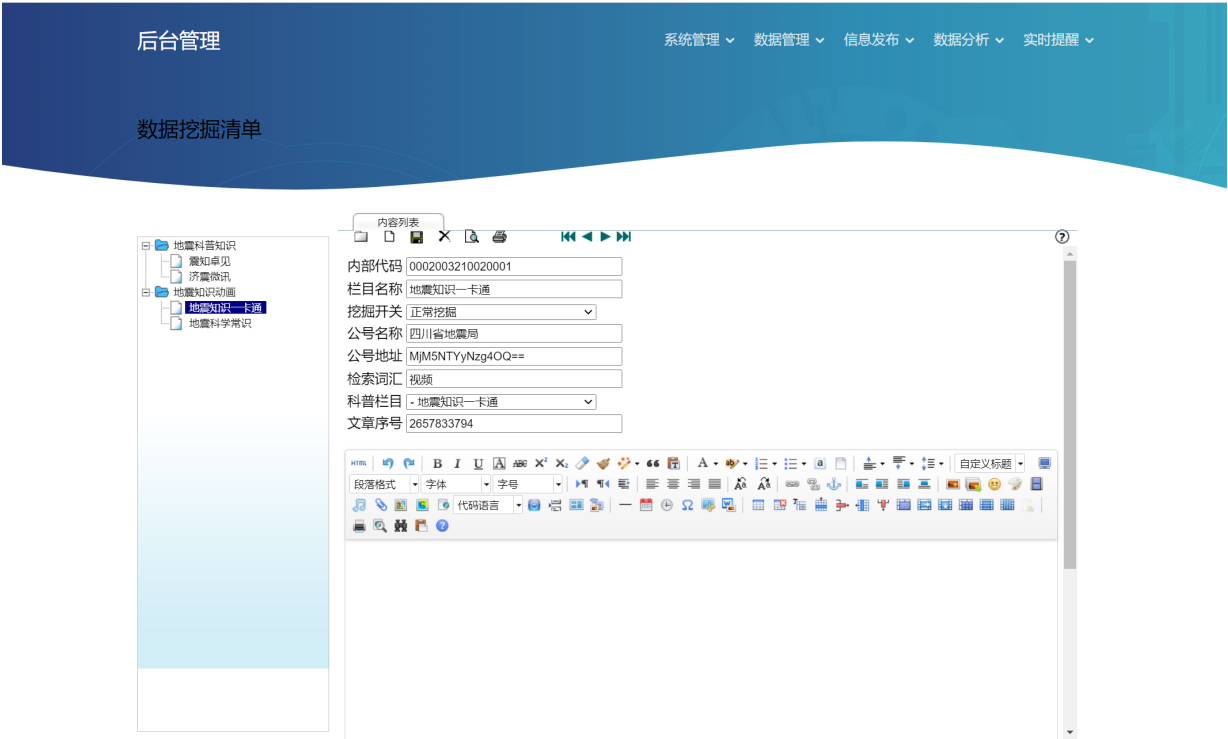


图2 知识挖据清单设置实例
Fig.2 Setting example of knowledge mining list

3.2 防震减灾知识挖据

防震减灾知识搜索与挖据是知识挖据的关键，图3为在新媒体平台中进行防震减灾知识挖据过程。系统会自动按照挖据清单自动采用模拟登录方式连接对应部署微信公众号服务号的账号，通过目标清单的融媒体发布相应关键知识内容检索结果分析，结合中文分词模块对知识内容进行分词处理，提取防震减灾知识按频度排序的关键词特征；基于主题模型分析模型确定知识符合度，利用网页页面解析技术对符合度较大的知识信息进行内容获取、链接分离、图文识别、图片下载、视频保存、文档转码等，自动分解出文字、视频、图件等信息进行本地化入库处理。对已获取相关地震知识，通过黑名单识别排除不当的文档信息，形成有效的防震减灾知识条目。对于每个新形成

知识条目，通过与已有知识数据库比较，自动排除重复条目；同时也提供知识条目人工审核入库功能，确保所建设的防震减灾知识库的科学性。

3.3 防震减灾知识推荐

充分利用融媒体形成的防震减灾知识大数据，利用知识挖据技术构建集图文于一体的防震减灾知识库，形成一个可自我增长的防震减灾知识资源库；基于主题模型智能分词技术、关键词匹配推荐引擎感知用户相关兴趣内容，采用基于实例知识推荐算法，实现文章聚合服务，形成推荐知识列表，并按匹配度尽量不重复精准地对用户进行个性化推荐，使每个知识点呈现到感兴趣的用戶面前。最终通过微信公众号、小程序等融媒体平台实现用户日常定期科普精准推荐以及防震减灾咨询知识精准应答服务(图4)。



图3 防震减灾知识挖掘过程

Fig.3 Knowledge mining process of earthquake prevention and disaster reduction

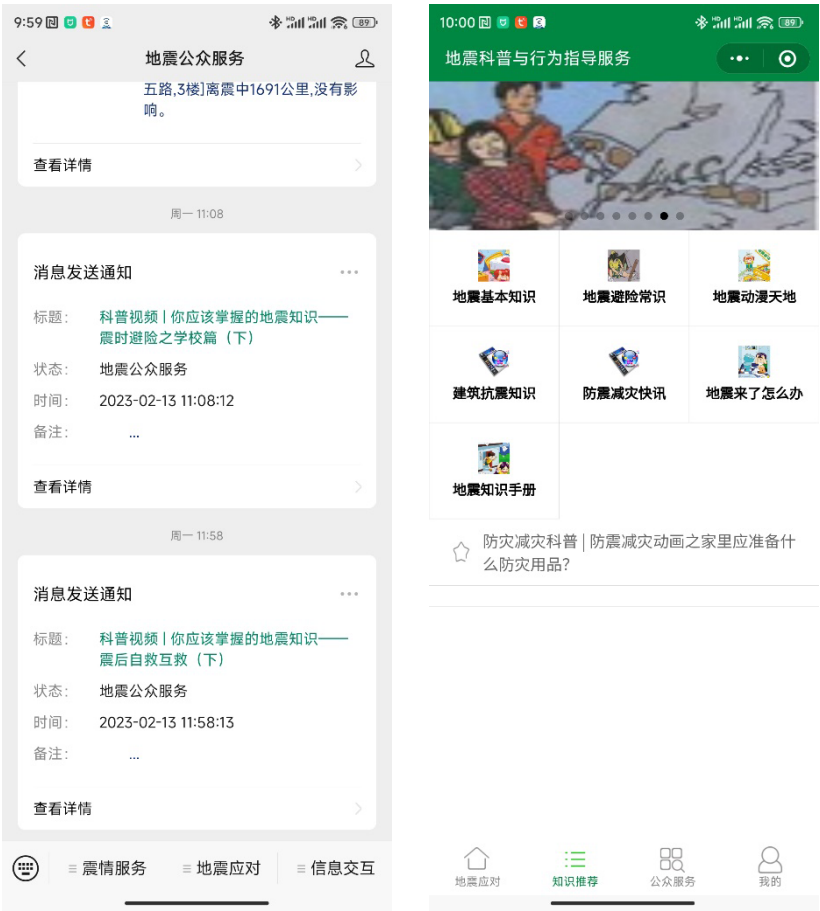


图4 防震减灾知识推荐

Fig.4 Recommendation of earthquake prevention and disaster reduction knowledge

4 应用成效

本文的知识挖掘技术方法已成功地应用到所开发的地震科普与行为指导新媒体平台, 并已成功部署在宁夏自治区、福建省、吉林省、重庆市、贵州省等省级地区, 厦门市、福州市、漳州市、包头市、许昌市、徐州市、长沙市、唐山市、廊坊市、西宁市、六盘水市、铜仁市、安顺市、海西州等市级地区, 寻乌县、福清市等县级地区微信公众服务号及小程序进行试运行。试运行用户达 20 万人, 3 个月内累计新增用户超 10 万人, 挖掘到防震减灾新媒体知识超 1000 条, 地震视频约 180 部, 根据用户兴趣主动推荐地震知识 136 万条。基于融媒体的防震减灾知识挖掘技术, 能不断地自动更新防震减灾知识资源库, 丰富防震减灾知识内容。可有效拓展全国社会大众获取地震科普知识渠道, 为社区、学校、公众的地震科普工作提供了新的传播媒介, 形成面向用户兴趣感知的智能化地震安全咨询与科普知识推荐服务, 实现科普信息即时获取、精准推送, 提高防震减灾科普传播效率。

参考文献

- [1] 游志斌. 全方位做好灾害应对准备切实保障人民群众生命财产安全[J]. 中国减灾, 2022(09): 18-21.
- [2] 刘允林, 徐仙鹏, 姚兰, 等. 地震应急期间现场防震减灾科普宣传的路径研究[J]. 高原地震, 2021, 33(03): 69-73.
- [3] 伊宏. 数据挖掘技术概述[J]. 中国标准导报, 2008(03): 19-22.
- [4] 胡英, 罗大庸. 数据挖掘技术的应用及发展[J]. 湖南第一师范学报, 2007(04): 165-166.
- [5] 马建光, 姜巍. 大数据的概念、特征及其应用[J]. 国防科技, 2013, 34(02): 10-17.
- [6] 刘仲达, 蔡宗文. 创建厦门市地震信息微信公众平台——提升地震部门公众服务能力[J]. 防灾减灾学报, 2017, 33(04): 100-104.
- [7] 李涛, 王次臣, 李华康. 知识图谱的发展与构建[J]. 南京理工大学学报, 2017, 41(01): 22-34.
- [8] 刘志荣. 电子政务的数据挖掘研究[J]. 广东技术师范学院学报, 2008(03): 8-10.
- [9] 王瑞, 李弼程, 杜文倩. 基于上下文词向量和主题模型的实体消歧方法[J]. 中文信息学报, 2019, 33(11): 46-56.