

项目结题验收单

专家验收表（**主持人所在单位**组织 3-5 名专家对项目进行验收、自评。）

项目名称	农业科学数据关键词挖掘研究与实现				
主持人	宁璐	职务/职称	系统支持部主任/研究馆员		
所在单位	北京农学院图书馆（加盖单位公章）				
专 家 意 见	<p>我国在大数据方面的发展十分迅速，继续深入发掘和探索大数据的资源也是我国乃至全世界都在关注的重点与热点。农业科学数据是指在农业领域，通过基础研究、应用研究、试验开发等产生的数据，以及通过观测监测、考察调查、检验检测等方式取得并用于农业科学研究活动的原始数据及其衍生数据。</p> <p>该项目通过研究农业科学数据的多维度、来源复杂等特点，针对农业科学数据领域进行了关键词提取，或对文本中的占有十分中心位置的词与短语进行自动抽取，对挖掘到的关键词进行分析处理、分类等操作，能够帮助用户迅速了解相关文献研究的中心思想，或者从海量语料中快速获得主题；将处理后的关键词进行可视化展示，建立一个相关的农业科学数据可视化界面，让用户在已有可视图表的基础上进行自定义交互可视分析，可以帮助人们从复杂多样的数据中理顺逻辑与思绪，化繁为简，可以让人们对于农业科学数据关键词有着更清晰的认识与理解。</p> <p>该项目研究对农业科学数据利用具有一定的实践意义和应用前景。专家组同意该项目通过验收。</p>				
专家签字					
职务/职称	研究馆员	副研究馆员	研究馆员		

项目结题验收单

专家验收表（主持人所在单位组织 3-5 名专家对项目进行验收、自评。）

项目名称	农业科学数据关键词挖掘研究与实现				
主持人	宁璐	职务/职称	系统支持部主任/研究馆员		
所在单位	北京农学院图书馆（加盖单位公章）				
专 家 意 见	<p>我国在大数据方面的发展十分迅速，继续深入发掘和探索大数据的资源也是我国乃至全世界都在关注的重点与热点。农业科学数据是指在农业领域，通过基础研究、应用研究、试验开发等产生的数据，以及通过观测监测、考察调查、检验检测等方式取得并用于农业科学研究活动的原始数据及其衍生数据。</p> <p>该项目通过研究农业科学数据的多维度、来源复杂等特点，针对农业科学数据领域进行了关键词提取，或对文本中的占有十分中心位置的词与短语进行自动抽取，对挖掘到的关键词进行分析处理、分类等操作，能够帮助用户迅速了解相关文献研究的中心思想，或者从海量语料中快速获得主题；将处理后的关键词进行可视化展示，建立一个相关的农业科学数据可视化界面，让用户在已有可视图表的基础上进行自定义交互可视分析，可以帮助人们从复杂多样的数据中理顺逻辑与思绪，化繁为简，可以让人们对于农业科学数据关键词有着更清晰的认识与理解。</p> <p>该项目研究对农业科学数据利用具有一定的实践意义和应用前景。专家组同意该项目通过验收。</p>				
专家签字	陈建	王颖	何利		
职务/职称	研究馆员	副研究馆员	研究馆员		



项目编号:

2023002

CALIS 全国农学文献信息中心研究项目 结题报告

项目名称: 农业科学数据关键词挖掘研究与实现

项目关键词: 农业科学数据, 关键词, 分类, 可视化

项目单位(盖章): 北京农学院图书馆

通信地址: (详细地址
址含邮编) 北京市昌平区回龙观镇北农路 7 号, 102206

项目主持人: 宁璐

联系电话: 010-80795674 (13701168920)

电子邮件: ninglu@mx.study.org.cn

提交日期: 2024 年 5 月 20 日



项目编号:

2023002

CALIS 全国农学文献信息中心研究项目 结题报告

项目名称: 农业科学数据关键词挖掘研究与实现

项目关键词: 农业科学数据, 关键词, 分类, 可视化

项目单位(盖章): 北京农学院图书馆

通信地址: (详细地址含邮编) 北京市昌平区回龙观镇北农路 7 号, 102206

项目主持人: 宁璐

联系电话: 010-80795674 (13701168920)

电子邮件: ninglu@mx.study.org.cn

提交日期: 2024 年 5 月 20 日

题目：农业科学数据关键词挖掘研究与实现

关键词：农业科学数据，关键词，分类，可视化

1 研究背景、目的及意义

农业在我国中的地位十分重要，在经济方面占据了重要的位置，这也决定了我国是一个农业类的大国。在社会的不断发展中，经济方面和科技类的发展也日益提升，于此同时，改变更大的就是农业的生产。为全面了解农业发展新态势，2017年完成的第三次全国农业普查，党的十九大报告依据第三次全国农业普查提出“必须始终把解决好‘三农’问题作为全党工作重中之重”，为新时代的“三农”发展指明方向。习近平总书记在党的二十大报告中强调，加快建设农业强国。当今时代，我国在大数据方面的发展十分迅速，继续深入发掘和探索大数据的资源也是我国乃至全世界都在关注的重点与热点。

农业科学数据是指在农业领域，通过基础研究、应用研究、试验开发等产生的数据，以及通过观测监测、考察调查、检验检测等方式取得并用于农业科学研究活动的原始数据及其衍生数据。我国在大数据方面的发展十分迅速，继续深入发掘和探索大数据的资源也是我国乃至全世界都在关注的重点与热点。

农业部门非常重视农业数据的收集，但是我国在农业领域的科学数据，相关资料与信息分布都比较分散，同时也会根据不同类别和特征存储到各种数据中心里，他们的这些数据结构特点十分多样化。同时，在我国关于农业科学数据的相关研究都比较匮乏，尚没有完备、针对性的清洗农业科学数据的方法，这样一来，在农业科学数据的利用方面，以及后续的系统研究中，效率就会大打折扣。

数据如何变得更有意义？这需要可视化的实现来完成它的任务，这样也可以让人们对于数据方面更有清晰的认识与理解。现如今，数据可视化平台已经占据了主要地位，可以帮助人们从复杂多样的数据中理顺逻辑与思绪，化繁为简，来更方便快捷地为用户服务。农业科学数据没有很好的可视化平台，数据利用和处理也会耗时耗力。通过数据的可视化展示，可以让人们对于农业科学数据关键词有着更清晰的认识与理解。

所以，本项目选择了农业科学数据领域进行关键词提取，进行可视化展示。现阶段，很多领域都有自己的关键词挖掘的相关研究，如网络热搜关键词提取，新闻网页关键词提取，网络舆情热点关键词等；但在农业科学数据这一领域上，还未进行过深入研究。关键词提取是自动抽取文本或在文本中的占有十分中心位置的词与短语，这样可以根据它的内容来描述一个文章中的重点内容。关键词提取能够帮助用户迅速了解研究的中心思想，或者从海量语料中快速获得主题。

虽然大量的信息可以为人们提供一些便利，但是这也让我们在搜寻相关信息的时候有很多困难与负担，如果针对这些信息都是用人工标注，会非常的费时费力，这样就会使关键词提取方面大大地降低人类对于效率方面的要求，故设计这样一个平台来给所有人提供便利与帮助，通过系统中所展示的可视化图表来进行分析，利用这些工具来获取相关信息，来帮助用户在查找农业相关信息时节省时

间与精力，可以准确并且快速地找到所需信息，同时可以对我国对农业科学数据进行整合。

2 研究内容及方法（思路、方法、具体内容）

2.1 研究思路

本研究可分为这几个阶段：关键词获取、关键词处理与分析、可视化展示。

关键词来源：网络文献、相关农业网站等；因为在收集到的文献资源和网络信息中的农业科学数据相关词较为分散，且相关词的定性定义不够准确，故选择这几种方法可以更全面地将相关数据进行整合操作。

农业科学数据相关词的获取：利用网络爬虫技术通过对网络文献、文章、理论报告以及部分农业网站信息进行抓取，通过文本分析等技术确定农业科学数据的相关词。

关键词的确定、分类和存储：针对来源不同的相关词，以及农业科学数据周期长、维度大、来源复杂三大主要特征：，对农业科学数据的相关词进行清洗、中文分词、词性标注、去停用词这四项处理，确定农业科学数据领域的关键词；对提取到的关键词进行分类、存储；运用算法对数据进行分析，找出属性之间的隐含关系，针对特定的问题进行研究。

可视化的实现：将处理后的关键词进行可视化展示，以来实现可视化平台，将大数据技术与农业领域深入耦合，从而解决用户在查找相关农业信息时费时费力、查找的内容不精确、阅读大量文字耗费精力并且难以准确快速地找到所需信息等一系列问题，同时对农业科技发展提供帮助与支持。

2.2 研究方法

本项目拟采文献研究法、功能分析法、数量研究法等研究方法和手段。

文献研究法：查阅国内外相关关键词挖掘与可视化展示的文献研究，以及当今对农业科学数据关键词的相关研究现状，从而全面、完整的掌握课题内容，通过收集大量资料，进行分析、整合，为课题提供理论依据。

功能分析法：通过本课题所要研究的具体内容，所要进行分析的相关数据，对其进行模块分类，对各个模块的功能进行具体分析，从而了解各模块的相关功能，为后期研究提供帮助与技术支持。

数量研究法：通过对研究对象的规模、范围等数量关系进行分析研究，揭示各关键词直接的相互关系、发展趋势，以便进行正确的关键词分类。

2.3 具体研究内容

2.3.1 数据准备

requests 库是一种十分常见的库，它可以处理很多复杂的内容，例如 Cookie、header 等等请求，同时它是和相关的库一起完成的爬虫内容，虽然是一些简单的爬虫内容，但是确实我们研发时最常用的爬虫技术。它可以对一些农业相关资源的网站进行爬取资源，这些网站的数据十分充足，可以充分利用 requests 库来进行爬虫，之后再爬取到的数据进行相关的处理，最后将处理后的数据存储到相

应的数据库中。

由于农业科学数据分布较为广泛，我国对于农业科学数据方面的研究还是比较多的，但是它们的特点都比较复杂，例如季节性、传播比较分散、它们所展示的农业数据所处的地域也都不同，它们的分布规律为较为集中，但是大部分的又很分散，农业科学数据的种类又各不相同，我们在获取农业科学数据的时候，方法也很原始；通过在中国农业科学院的国家农业科学数据中心网站中，其中的数据库可以将农业科学数据的分类进行具体的划分：植物保护科学、作物科学、农业资源与环境科学等，故本系统的数据通过网络爬虫从中国农网、中国农业科技信息网、中国农业科学院等农业专业网站进行爬取相关农业科学数据文献 9178 篇，其中包括文献的标题、摘要、正文、发布日期、作者。该类网站的相关农业数据的权威性较高，可以清晰地了解到当前农业科学数据的走向与发展。得到数据后，首先将数据处理后存入数据库中，然后对文献摘要、关键词和题目这部分数据进行分词、过滤停用词、词干提取、近义词替换等数据预处理，以上所描述的步骤下文将详细阐述。

2.3.2 数据处理

通过对农业科学数据进行分析，本系统针对农业科学数据的特点与第三章对算法的分析，先对爬取的 9178 篇文献进行处理，可以看出文章里面有许多标点符号、特殊字符、换行符“\n”等，这些字符都会给后面的分词带来影响，所以我们首先要对文章进行预处理，把这些无用字符去掉。

```
def read_txt(path):'''读文件'''txt = open(path, 'r', encoding='utf-8').read()for x in '，。‘’“”：！、《》；？」「…’：':txt = txt.replace(x, "")txt = txt.replace("\n","")return txt
```

接下来对采集到的每条数据用 jieba 库进行分词，使用中文停用词库，将不太重要的词或者是无关紧要的词筛选为停用词，例如仍然、各种、消息等，清除掉每条文本中的停用词，截止目前，共计 1872 个停用词，后续的研究中会根据实际需求随时对停用词列表进行添加或者去除。这样，得到了每条文本中的词；对于所有的词，统计其出现的次数；随后，使用 wordcloud 库，绘制词云。

结构化数据与非结构化数据是采集的两种不同资源，接下来就要对他们内部不同的结构特点以及后续工作、分析、可视化的需求，来进行相应的处理，如：清洗与格式转换。因为数据采集处理得不够细致，暂时不能进行存储和可视化展示，所以要进行下一步处理。数据清洗处理流程如图 1 所示。

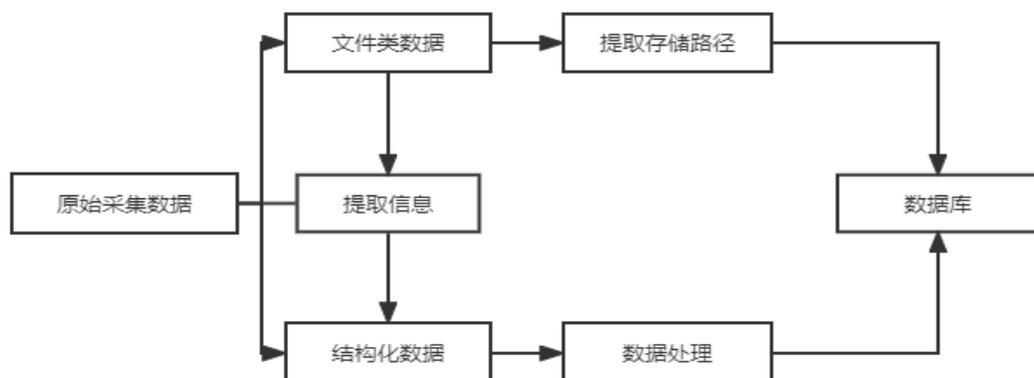


图 1 数据处理流程

2.3.3 数据库分析与设计

CSV 可以通过使用相关文本内容,进行数据的存储,将内容存储到表格数据中,CSV 可以记录任意的内容,但是每条记录都不是连在一起的,它们都会被相关的换行符进行分隔,这样也有便于查找与发现。CSV 中的每条记录都有很多字段,它们之间使用逗号这个分隔符来进行分隔,通常也会使用制表符来进行分隔。CSV 是以特定字符作为分隔符的纯文本,结构简单清晰。本次系统使用的是 CSV 数据存储,CSV 模块实现了 CSV 格式表单数据的读写。这可以以一个兼容 Excel 的方式读写其数据文件, csv 模块中的 reader 和 writer 类被用来读写序列化的数据,也可以使用 DictReader 类和 DictWriter 类以字典的方式读取数据。如图 2 所示。

芹菜菌核	芹菜菌核病可危害叶片、叶柄和茎。一	2022-01-12	中国农业科技信息网
中药材如	为有效做好防御应对、防寒减灾工作,	2022-01-12	中国农业科技信息网
蔬菜保鲜	低温贮藏。恒温库低温贮藏是一种简单	2022-01-12	中国农业科技信息网
冬季大棚	1、注意长势 看植株长势,通常植株	2022-01-12	中国农业科技信息网
葱蒜类蔬	锈病是大葱、大蒜、韭菜、洋葱等葱蒜	2022-01-12	中国农业科技信息网
预防棚室	适时浇水。浇水要控两头促中间,即结	2022-01-12	中国农业科技信息网
如何有效	菠菜适宜种植,很容易栽种的,菠菜有	2022-01-12	中国农业科技信息网
看叶片区	眼下,正是大棚黄瓜生产季节。冬季低	2022-01-12	中国农业科技信息网
猕猴桃树	降低果园湿度,做好清园工作,修剪工	2022-01-12	中国农业科技信息网
冬季梨园	中耕除草。结合施冬肥,铲除梨园内的	2022-01-12	中国农业科技信息网
灰霉病反	进入冬季以后,棚内高湿的环境使灰霉	2022-01-12	中国农业科技信息网
果树防冻	当果树受到寒冷天气侵袭的时候,树木	2022-01-12	中国农业科技信息网
菌种活力	1、保证菌种的纯培养,注意不要用被杂	2022-01-12	中国农业科技信息网
果树冬刮	果树随着年龄增加,树皮会增厚,缺乏	2022-01-12	中国农业科技信息网
芹菜空心	科学浇水。为避免芹菜空心现象的发生,	2022-01-12	中国农业科技信息网
壮苗防冻	油菜进入越冬期后,种植户应围绕壮苗	2022-01-07	中国农业科技信息网
莲藕 莴笋	随着气温的不断降低,如何使用莲藕和莴	2022-01-07	中国农业科技信息网
蔬菜结果	伴随着大雪节气的来临,总算有冬天的	2022-01-07	中国农业科技信息网
蔬菜冲施	不少菜农由于施用冲施肥过勤、过量,	2022-01-07	中国农业科技信息网
梨树腐烂	梨树腐烂病发病的最适温度为25℃左右,	2022-01-07	中国农业科技信息网
苹果树腐	苹果树腐烂病菌为弱寄生性真菌,病害	2022-01-07	中国农业科技信息网
防治苹果	又到一年果树管理时节,冬季苹果树免	2022-01-07	中国农业科技信息网
冬季防治	把好清园关,减少菌源量。将猕猴桃园	2022-01-07	中国农业科技信息网
如何自制	为减少化学农药在稻田的使用量,降低	2021-12-31	中国农业科技信息网
植物抗病	(1) 田间鉴定和温室鉴定 田间	2021-12-31	中国农业科技信息网
怎样合理	保管肥料应做到“六防”: (1) 防止混放	2021-12-31	中国农业科技信息网
及早查治	温室白粉虱又名小白蛾子,是北京地区温	2021-12-31	中国农业科技信息网

图 2 csv 数据存储

2.3.4 分词处理与去停用词

构建农业科学数据模型的过程中,在对关键词进行分析之前,需要先把相关文献进行归属,将在农业权威网站上爬虫获取的文献进行分词处理,使用 jieba 分词将词语切分,例如果树病害防治,经分词处理后变为果树-病害-防治,分词后进行去停用词处理,为了使存储数据的空间降低,同时加快解析效率,在对文本数据或者自然语言数据进行处理时,提前或者完成后去除掉部分词;截止目前,共计 1872 个停用词,后续的研究中会根据实际需求随时对停用词列表进行添加或者去除。相关代码如图 3 所示。

```

# 获取有效单词
def cal_useful_words(texts):
    if type(texts) == float:
        return []
    # 创建空列表, 存储单词
    useful_words = []
    # 打开停用词文件
    with open("../Resouces/stopwords.txt", "r", encoding="utf-8") as f:
        # 读取内容
        stopwords = f.read().split("\n")
    # 用jieba切词
    words = jieba.cut(texts)
    # 遍历每个词
    for word in words:
        # 如果这个词是中文, 且不再停用词中
        if is_all_chinese(word) and word not in stopwords:
            # 将其加入有用词列表
            useful_words.append(word)
    # 返回有用词列表
    return useful_words

```

图 3 分词处理代码图

2.3.5 词频统计

在对农业科学数据关键词进行分词处理以及去停用词后，再对其进行词频统计，通过词频统计抽取关键词，字面意思即需求，引出假设，文本中出现次数最多的词就是关键词。运用词频统计方法计算出文本中所有词在文本中出现的全部次数，根据实际需求，确定一定的数量进行关键词抽取。对去停用词处理后的文献数据进行词频统计，为了避免定义数量的数据起不到补充实验过程和实验数据的效果，没有对数量进行定义，统计记录的是全部数据的词频。在后续的研究中，可以根据实际需求，筛选所需的数量的数据词频。对于所有的农业科学数据的词，统计其出现的次数，在农业科学数据领域，对于在相关权威农业网站爬取的文献中，词频高的前十个词分别为“防治”、“植保”、“技术”、“蔬菜”、“病虫害”、“农业”、“植物保护”、“病害”、“农药”、“防控”。高频词分布分为标题高频词和内容高频词，标题高频词为该关键词在文章标题中出现的次数，内容高频词为该关键词在全文出现的次数，柱状图越高，说明该关键词出现的次数越多，其频率也越高。如图 4 所示。加强对植物病虫害防治存在的问题及应对措施进行深入的探讨，意义重大。相关代码如图 5 所示。

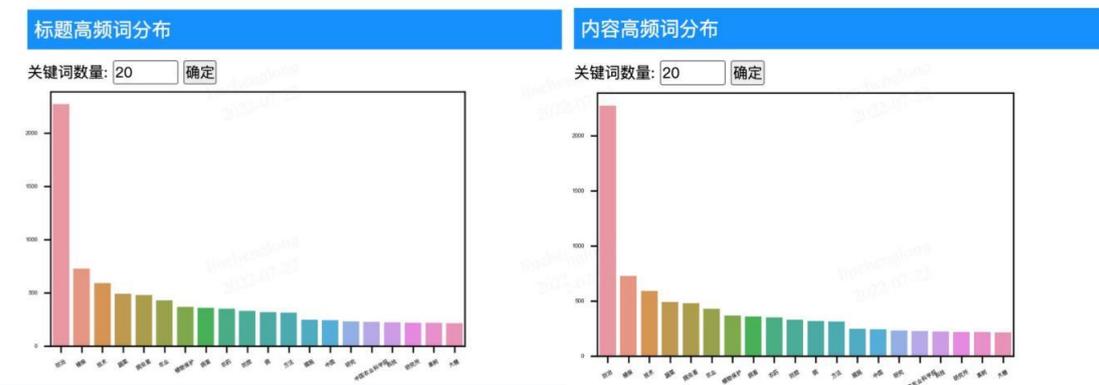


图 4 高频词图

```

# 统计词频
def get_word_freq(texts, ascending = False):
    word_freq = {}
    for text in texts:
        for word in cal_useful_words(text):
            # 如果该词语已经出现过, 出现次数+1
            if word in word_freq.keys():
                word_freq[word] += 1
            # 否则将该词语加入字典, 出现次数记为1
            else:
                word_freq[word] = 1

word_freq = pd.DataFrame({"词语": list(word_freq.keys()), "出现次数": list(word_freq.values())})
word_freq.sort_values("出现次数", ascending=ascending, inplace=True)
return word_freq

```

图 5 词频统计代码图

2.3.6 绘制词云

在对农业科学数据的词频统计完后使用 wordcloud 库绘制词云，wordcloud.WordCloud() 代表一个文本对应的词云，在绘制相关词语的词云时，就需要利用标题和文本中关键词出现的频率进行绘制，以 WordCloud 对象为基础，配置参数、加载文本、输出文件。词云绘制后就可以看出相关关键词出现的频率，关键词出现的频率越高，说明在这篇文章中所涉及到的重点就为该关键词，绘制词云时，获取前 100 最高频的词，标题词云与内容词云较为显著的几个关键词为“植保”、“防治”、“农药”、“病虫害”等，这里也可以看出我国在植物保护中，关注度高的仍旧是对于植物病虫害的防治、预防和如何使植物更好的生长。

词云图分为标题词云图和内容词云图，均表示该关键词出现的频率高低，字体越大，说明该关键词的频率越高，反之，字体越小，说明频率越低。如图 6 所示。

标题词云图



内容词云图



图 6 词云图界面

2.3.7 农业科学数据的可视化系统

由需求分析得知，农业科学数据的可视化系统主要是为了使用户能够更有效直观地观察农业科学数据关键词的相关内容，从而使该系统为农业科学数据乃至整个农业提供相应的价值，在系统分析阶段，得到了一个比较完备的设计流程与方案；由于我国现如今在很多数据挖掘方面都有投入，但在农业科学数据处理方面还未有一个完整的体系，可视化平台也没有看到有关农业科学数据的提出，这两方面在我国都鲜少有发现，故而设计本系统，也可以方便用户查找农业科学数据领域的相关内容，从而对农业数据进行利用与发现，来进一步推进我国的农业发展。用户可根据自己的需要在搜索框中输入相关文字，点击“搜一下”可查看文字类的说明，点击“看统计”可以直观的看到该关键词的词频分布、共现词网络分布以及词云图。主页搜索界面，如图 7、图 8 所示。



图 5-5 主页搜索界面



图 5-6 主页搜索界面

3 结论与建议

3.1 总结、结论

农业生产在经济发展中占据着重要的位置，这也决定了我国是一个农业类的大国。在社会的不断发展中，经济方面和科技类的发展也日益提升，于此同时，改变更大的就是农业的生产。本项目通过对农业科学数据关键词挖掘进行了研究，并设计实现了农业科学数据关键词挖掘与可视化展示系统平台，用户管理、数据管理、数据挖掘与可视化模块均已完成。具体所做的工作如下：

(1) 对于在农业相关网站获取农业科学数据方面。选定了部分农业科学数据中的相关权威性的农业网站，通过对 requests 库来对相关农业文献进行爬取，后在对爬取到的数据进行采集的工作，解决了大规模农业数据集的快速爬取问题，可以完成了千万级数据的采集需求。

(2) 在对农业科学数据采集完成后,进行进一步的清洗。在数据获取后对具有数据挖掘价值的信息进行了字段信息提取,实现了数据整合。

(3) 对于农业科学数据可视化系统设计。使用 Flask 框架以及 requests、pandas、jieba 等第三方库实现了可视化平台,使用 Matplotlib 和 networkx 可视化工具对收集的农业科学数据进行了可视化展示。

本项目实现了对农业科学数据关键词的挖掘与可视化展示,后续还希望该技术可以更加精进,希望在后续的工作中能够提升对于关键词挖掘和提取的精度,更符合农业科学研究的需求。

3.2 展望

本项目主要对农业科学数据关键词挖掘和提取的技术进行了研究,并设计实现了农业科学数据关键词挖掘与可视化展示系统平台,系统平台实现了课题研究设定的基本功能,但是在某些方面还有着很多提升和改进的地方,从而更好地符合农业科学研究的需要,具体的改进内容如下:

(1) 扩大数据采集的来源。我对于本系统的设计中,查询了很多农业权威性网站,但是这在我国农业发展中,只是很小的一部分,在农业科学数据这一领域,还有很多的农业相关网站的数据用来展示该领域的相关研究;在数据采集中,还是有一定的局限性,数据采集的规模还可以扩大,虽然采集的数据种类比较全面,但这只能是代表我国农业科学数据的一部分,还应该根据不同的类别和领域分类来拓宽数据采集的范围,深究采集的内容和需求,这也是该系统需要改进的一点和有待进一步研究的方向。

(2) 使用不同的类别可视化图表来展示不同的内容。因为农业科学数据采集范围的局限,展示的内容也是以文字类为主,整个系统中的图形、图表数量也有限,故需要加强农业科学数据方面的数值型采集需求,通过对数值型的数据进行分析、分类,来进行可视化展示,再根据细节进行分类,不同的数据类型分门别类的进行展现,这样也可以方便用户可以更直观的看到系统所要表达的内容。

(3) 虽然本次系统的各个功能都很齐全,但是功能中的内容不够丰富,需要进一步的完善,还需要对整个页面进行加工处理,使得使用本系统的用户能够对系统有很好的体验感和深刻的印象。

4 项目成果 (发表的文章、开发的软件、取得的实践效果等)

(1) 指导研究生完成“农业科学数据关键词挖掘与可视化展示系统平台(V2.0)”的平台升级。

(2) 作为通讯作者发表论文一篇:北京房山区蒲洼乡中华蜂产业发展策略(安徽农学通报,2023,29(17):73-76.)。

(3) 指导三名研究生完成相关硕士毕业论文。

5 参考文献

[1]谢金星,邢文训,王振波.网络优化[M].清华大学出版社,2000.

- [2]AricHagberg,DanSchult,PieterSwart.NetworkXReference[DB/OL].https://networkx.github.io/documentation/stable/_downloads/networkx_reference.pdf,2018-1-22.
- [3]时雷.基于物联网的小麦生长环境数据采集与数据挖掘技术研究[D].河南农业大学,2013.
- [4]AhujaRK,MagnantiTL,Orlin JB.Network flows: theory, algorithms,and applications[M].Prentice Hall,1993.
- [5]王航臣.基于级联失效的航路网络中关键航路识别算法[J].航空计算技术,2017,47(06):32-35.
- [6]胡为群,祝利莉,郑可锋,叶少挺,朱旭斌.农业资源属性数据挖掘研究(1).科技情报开发与经济,2006(15):207-208.
- [7]Deichmann,Aparajita Goyal,Deepak Mishra.Will digital technologies transform agriculture in developing countries?[J].Agricultural Economics,2016,47(S1).
- [8]GaiLi,ZhiqiangZhang,LiyangWang,QiangChen,JincaiPan.One-Class Collaborative Filtering Based on Rating Prediction and Ranking Prediction[J].Knowledge-Based Systems,2017.
- [9]吴建军,高自友,孙会君等.城市交通系统复杂性[M].科学出版社,2010.
- [10]马治涛.文本分类停用词处理和特征选择技术研究[D].西安电子科技大学,2014.
- [11]牛萍.TF-IDF 与规则结合的中文关键词自动抽取研究[D].大连理工大学,2015.
- [12]王文生,郭雷风.农业大数据及其应用展望[J].江苏农业科学,2015,43(09):1-5.
- [13]睦海霞,孙清.大数据时代下的网络文化传播分析[J].成都理工大学学报(社会科学版),2017,25(01):102-10
- [14]莫鹏.基于超图的文本摘要与关键词协同抽取研究[D].华中师范大学,2016.
- [15]王立华,肖慧,徐硕,刘树,杜卫利,黄其泉,王宇.基于关联规则的渔业信息推荐系统设计与实现[J].农业工程学报,2013,29(07):124-130.
- [16]柯杨蒙,时政新闻信息可视化设计研究[D].武汉:华中科技大学,2015.
- [17]PAGANO P,CANDELA L,CASTELLI D. Data Interoperability[J].Data Science Journal,2013,12:GRDI19-GRDI25.
- [18]杨京,王效岳,白如江.大数据背景下科学数据互操作实践进展研究[J].图书与情报,2015(3): 97-102.
- [19]Health N I.NIH Data Sharing Policy and Implementation Guidance. Gants&Funding.[EB/OL].(2016-01-03)[2017-01-02].https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm.
- [20]吴彬.生态科学数据元数据及其标准研究[J].中南林业科技大学学报,2010,30(12):75-79.
- [21]战学刚,吴强.基于 TF 统计和语法分析的关键词提取算法[J].计算机应用与软件,2014,31(1):47-49.(ZHAN Xuegang, WU Qiang. Keyword extraction algorithm based on TF statistics and syntactic parsing[J].Computer Application and Software,2014,31(1):47-49.)
- [22]崔运鹏,钱平,苏晓鹭,等.农业科技信息核心元数据标准及其计算机著录[J].中国科技资源导刊,2010,42(3):41-44
- [23]FREITAS A,CARVALHO D S,SILVA J C P D,et al.A Semantic Best-Effort Approach for Extracting Structured Discourse Graphsfrom Wikipedia [C]// Web of Linked Entities Workshop in

Conjunction with the International Semantic Web Conference.2012.

[24]花君林.基于 Selenium 的 Python 网络爬虫的实现[J].电脑编程技巧与维护.2017(15):30—31+36.

[25]侯洁茹,吕继续.基于 Python 的天猫商品爬虫技术[J].科技资讯,2019,17(32):10+12.

[26]吴陈伟.基于图嵌入模型的复杂文本关键词挖掘方法[D].福建工程学院,2021.DOI:10.27865/d.cnki.gfgxy.2021.000097.

[27]吕佳高,梁奎阳,蔡伟.基于文献计量和众包技术的前沿科技关键词挖掘[J].计算机科学,2019,46(03):275-282.

[28]王琳鑫.基于 CiteSpace 的国内农业现代化研究进展及热点可视化分析[J].现代营销(学苑版),2021(06):3-5.DOI:10.19932/j.cnki.22-1256/F.2021.06.003.

[29]黄家章,卢士军,姚远,吴鸣,孙君茂.基于文献计量的国际营养导向型农业研究进展可视化分析[J].中国农业科技导报,2020,22(09):11-21.DOI:10.13304/j.nykjdb.2019.0642.

[30]顾颖,祁小敏.基于 GIS 的南京市农业大数据可视化平台建设[J].山东农业大学学报(自然科学版),2020,51(04):702-704.

[31]刘银波,边静,蔡红梅,王艳.基于 Python+Pandas+Matplotlib 的学生成绩数据统计与图形输出实现[J].信息与电脑(理论版),2019,31(19):43-45.

[32]陈华.基于 Python 的数据分析与可视化平台研究[J].网络安全技术与应用,2022,(02):57-58.

[33]胡永乐.基于 Flask 的智能家庭种植系统设计[D].北方民族大学,2021.

[34]宋永生,黄蓉美,王军.基于 Python 的数据分析与可视化平台研究[J].现代信息技术,2019,3(21):7-9.

[35]裴丽丽.基于 Python 语言对电影影评数据爬虫与词云制作[J].信息记录材料,2020,21(05):116-118.

[36]刘慧婷,刘志中,王利利.一般间隙序列模式挖掘的关键词抽取[J].电子学报,2018.

[37]董宇,杨景升,秦彤昱,郭海博.基于文本关键词提取的自驾车营地行业标准讨论——以《自驾车旅居车营地质量等级划分》为例,[J].中国标准化,2021(08):125-132.

[38]朱玉梅 Apriori 算法在棉花病虫害分析中的应用[J].农学学报,2016,6(11):23-26+31.

[39]彭秀媛,王枫,周国民.辽宁省农业科学数据共享情况调查与分析[J].农业经济,2017(1):59-61.

[40]王一鹤,杨飞,王卷乐,刘丽峰,马娅.农业大数据研究与应用进展综述[J/OL].中国农业信息,2018,30(04):48-56

[41]眭海霞,孙清.大数据时代下的网络文化传播分析[J].成都理工大学学报(社会科学版),2017,25(01):102-106.

[42]胡亮,雷波,牛凯,等.四川育种攻关数据共享平台系统设计[J].农业网络信息,2011(07):27-30.

[43]李晓雨.利用 python 实现对存储在 Mysql 数据库的日志文件的分析和保存[J].科学技术创新,2021,(29):94-96.

[44]雷鑫.网络舆情引导平台系统及其关键技术的研究与实现[D].成都:电子科技大学,2016.

[45]孙九林,李灯华,许世卫,吴文斌,杨雅萍.农业大数据与信息化基础设施发展战略研究[J].中国工程科学,2021,23(04):10-18.

[46]刘海燕,杨榕,侯彤瑜,赵维,姚兆群,王海江,张泽,高攀,吕新. 新疆棉田土壤微生物资源大数据平台建设与可视化分析[J]. 农业大数据学报,2021,3(01):45-55.

[47]许哲平,邵曾婷,朱学军,王昉,王媛媛,肖曼,马克平.农业生物多样性大数据平台建设研究和展望[J].农业大数据学报,2019,1(02):76-87.

[48]吕帅.基于 Python 开源框架 Flask 的地震信息网络运维管理系统实现[J].高原地震,2018,30(03):59-63+55.

[49]The effects of m-learning on motivation,achievement and well-being: A Self-Determination Theory approach[J].Lucas M.Jeno,Paul J.C.Adachi,John-Arvid Grytnes,Vigdis Vandvik,Edward L.Deci.British Journal of Educational Technology.2019 (2)

[50]牛作东,李捍东.基于 Python 与 flask 工具搭建可高效开发的实用型 MVC 框架[J].计算机应用与软件,2019,36(07):21-25.

[51]A python framework for multi-agent simulation of networked resource systems[J].Stephen Knox,Philipp Meier,Jim Yoon,Julien J.Harou.Environmental Modelling and Software.2018.