

韩语新闻数据库的构建与高校韩语教学中的应用研究

郭新洁

(四川外国语大学东方语言文化学院 重庆 400030)

[摘要]传统的高校韩语教学面临学科转型升级的当下,以融合的姿态与前沿IT技术结合有助于教学和研究质量的提升。在本论文中,利用爬虫技术从韩国新闻网站中快速、高效、准确地提取出所需新闻关键信息,并构建本地韩语新闻数据库,再利用词云可视化技术从本地数据库中提取信息进行分析。通过利用以上技术手段,为高校韩语教学提供一些借鉴。

[关键词]韩语教学;韩语新闻数据库;网络爬虫与词云可视化;python

[DOI] 10.12252/j.issn.2096-6288.2021.03.259

韩语的学习与教学一般围绕听、说、读、写四个方面展开,但传统的课堂教学目的性过强,只注重教授语言,容易使学生们丧失学习兴趣,也不利于学生对韩国进行全面、整体的了解。新闻具有可读性和趣味性的特点,在传统课堂中引入韩语新闻教学课堂,可以增强课堂的趣味性,使学生们学习到关于韩国多方面的知识。但因韩语新闻文本数量庞大、种类繁多,在查找新闻资料时,不仅耗时、费力,还容易出现遗漏。借助python语言中的网络爬虫技术可以在海量的韩语新闻数据中快速、准确地爬取所需新闻,然后运用可视化技术快速提取出高频词汇和关键词并将其制作成词云图,使学生借助词云图中展现的高频词快速掌握多篇新闻的梗概,提高学习韩语新闻的效率。将python语言应用到韩语新闻教学课堂,有助于提高高校韩语新闻教学的质量。

一、基于Python的网络爬虫技术与词云可视化技术

互联网的快速发展使信息呈爆发式增长,关于同一事物的相关新闻报道可能会出现成百上千条,且散布在不同的新闻网站中。利用网络爬虫技术,可以在短时间内迅速爬取各个网站上所需的相关新闻报道数据,建立一个本地文件,将爬取到的数据封装在文件中,存储到本地,从而构建一个属于自己的相关数据库(在本文中的数据库是本地的excel数据库,数据库构成是标题、发布日期、内容)。从数据库中提取出高频词汇或关键词进行可视化分析,将其应用到自己所需领域中。通过构建这样的相关新闻数据库,可以快速、高效地筛选无关信息,获取到自己所需的信息。

爬虫,从本质上来说,就是利用程序在网上拿到对自身有价值、有需要的数据。爬虫的工作原理分为以下四步:获取数据、解析数据、提取数据、存储数据。网络爬虫技术可以快速、高效地提取出海量所需信息,正好适用于本研究在韩国新闻网站中对所需韩语新闻进行提取的用例。技术的改革使一些简单重复性的劳动或思考活动逐步被机器取代,因此将简单的爬虫技术应用到语言学习中是十分有益、且有必要的。

词云又称可视云或标签云,是文本数据可视化的一种工具。在文本数据所占比重日益居高的如今,为了更好地利用这些文本数据,使其发挥功效,人们对于文本数据处理技术的需要也随之产生。词云可视化技术可以在繁琐、冗杂的信息中提取出高频率的关键词以及渲染词,采用可视化的方式将关键信息通过词云图展示出来。使人们迅速掌握文本的关键信息,还可以辅助分析大篇幅的文本信息。由其制作出来的词云图可以制作成各式各样的图片样式,在简洁、美观的同时,极大地提升了读者的阅读兴趣。

二、新闻数据库的构建在高校韩语新闻教学中的应用

(一) 韩语新闻爬取用例

2020年伊始,不速之客新冠疫情的到来,给世界带来巨大灾难。因为疫情,整个世界都被按下暂停键。从疫情开始的第一天直到现在,每天都会有关于疫情的新闻被报道。因此本研究运用爬虫技术提取了 연합뉴스(联合新闻:韩国人常用新闻网站)上于2020年4月15日至2021年4月15日一年以

来韩国报道的有关新冠疫情经济方面的新闻。

具体编程操作如下:

```
import requests, csv, json ①
from bs4 import BeautifulSoup
file = open('news.csv', 'w') ②
writer = csv.writer(file)
list = ['标题', '发布时间', '内容']
writer.writerow(list)
for i in range(1, 51): ③
    url = "韩语新闻网站地址".format(i)
    r = requests.get(url)
    results = json.loads(r.text[25:-2])["KR_ARTICLE"]["result"]
    for result in results:
        try:
            article_url =
                "https://www.yna.co.kr/view/{}?section=search".format(result["CONTENTS_ID"])
            r = requests.get(article_url)
            soup = BeautifulSoup(r.text, 'lxml')
            content = soup.find("article", class_="story-news article").get_text()
        except Exception as e:
            print(e, article_url)
            list1 = [result["TITLE"], result["DIST_DATE"], content]
            writer.writerow(list1) ④
file.close()
```

①导入python爬虫模块, requests为获取数据模块; json为解析数据模块; BeautifulSoup为提取数据模块; csv为存储数据模块。②创建名为“news”的csv文件,用于存储爬到的韩语新闻数据。③进入韩国新闻网站爬取所需内容,并通过列表对字符串进行切片,去掉不需要的内容。④将爬取到的韩语新闻数据封装在之前创建的csv文件中,存入本地。

操作结果如图<1-1>

如图<1-1>(由于空间限制,仅展示了前20条)所示,借助网络爬虫技术提取了共同涵盖新冠疫情和经济两个关键词的1813条韩语新闻数据,将其存储到本地创建的名为“news1”的csv文件中。从爬取上千条信息到存储到本地电脑为自己所有,这个过程仅仅花费了几分钟的时间。由此可见熟练操作运用爬虫技术,可以从海量信息中轻松提取出自己想要的信息,从而构建一个属于自己的韩语新闻数据库。如将此技术应用到教育行业,可以极大地节省时间和人力,提高工作或学习效率。

(二) 词云可视化用例

在知识可视化的教学理念下,“词云”在韩语学习中有着良好的应用前景。词云分析不仅可以应用于高校韩语专业的《报刊选读》《文学作品选读》等课程中,还可以应用于韩语新闻课程中。以上文中作者在韩国新闻网站中爬取到的

创客文化是创客教育发展的土壤，学校可以举办创客讲座，邀请各行各业的创客佼佼者宣扬创客精神、讲解典型的创客案例，利用设计学院学生专业的优势，在课程中融入实际项目，创办创客节目，宣传创客文化，实时展现创客风采，营造良好的创客生态环境。

3.6将专创融合的理念深度融合到专业教学的全流程

3.6.1将专创融合理念融入人才培养方案中

专创融合改革需要把创新精神、创业意识和创新创业能力作为评价人才培养质量的重要指标，使创新创业教育理念和内容融入专业教育教学中，贯穿人才培养全过程，使专创融合理念融入人才培养方案中。

3.6.2将专创融合理念融入课程标准中

课程标准要体现与课程内容相关的创新创业知识、技能和素质等；通过聘请行业协会优秀人才作为客座讲师，给学生面对面授课，实现专业与创业，专业与企业无缝对接，为学生创新创业提供条件与发展空间。改革考试考核内容和方式，注重考查学生分析、解决问题的能力，探索多样化、多形式、多标准的考试模式。

3.6.3将专创融合理念融入实践教学中

在构建专创融合实践教学体系上，采取“训赛融合”的理念，构建由仿真到实践，由浅到深，由仿真到实践的专创融合实践教学体系，促进了专创融合理念。

3.7构建教师研修创客技能的循环学习平台

创客教师是创客教育的重要主体，以“按需培训，训用结合”为理念，构建一个教师培训学习平台。对教师进行创客导师培训，让教师通过线上线下的学习模式及时掌握信息，真正实现随时随地学习，从而建立一个涵盖各学科领域的创客教师联盟。

3.8建立多元激励的创客实训平台管理模式

设立明确的创客教育目标，根据专业的特点，对于学生和教师的考核采取过程评估和结果评估、自我评估与同伴

评估、上级评估与下级评估相结合。在激励措施上，设立专门的创客奖学金、开展创客作品展览会、举办创客大赛等，努力激发学习者的创新思维，实现创意，推动创客教育的发展。

四、创客实训平台的特点

运用专创融合的理念，在充分学生的特征和需求的基础上，剖析创新创业教育融合专业教育后的教学内容设计、教学方法技巧的创新、教学场景的打造、教学工作的有序安排，分析融合的有效路径，根据学生创客能模型和知识技能类型，采用认知式学习、虚拟体验式学习、项目制学习和研讨式学习四种学习方式构建一套适合艺术设计类专业的创客教育体系。

五、结语

以服务为导向，以参与为前提，以成果转化为目标，形成学、研、产一体的学生+的创客空间实验平台。通过这个系统平台的搭建，基于新的设计教育范式，通过教育的理念、观念、模型、场景、系统的设计，实现创客实训平台的创新路径，打造创客学习生态。提升学生的创新思维、创业能力、引导学生成为能发现问题和解决问题，具有将专业知识和跨界知识进行整合的“T型人才”。构建服务地方经济“四位一体”的文创孵化平台，与产业发展紧密结合，促进创新创业项目转化落地。

参考文献

- [1]王维嘉. “双创”时代文化创意人才培养的探索与思考[J]. 创新与创业教育, 2017.8(4)
- [2]谢天. 创意产业类专业教学现状、问题及对策研究[J]. 湖南包装, 2018(6)

作者简介:

姜娜, 女, 1981, 湖北黄冈, 汉, 副教授, 硕士, 研究方向: 产品设计。

(上接第282页)

出关键词以及高频词汇，在实际教学中，提高实时性，使学生掌握最基本的新闻内容，第一时间获取相关新闻要点、热点话题；还可以使学生们在学习知识的同时，看到不同的国家在报道与国内相似新闻时有何不同的观点、角度，从而拓展学生对于事物思考的一个新的角度和国际视野，有助于拓宽学生的思考空间、培养学生的国际意识。

其次，教师在进行韩语新闻教学时，借助网络爬虫技术和云图可视化分析技术可以快速、高校地提取不同种类的新闻、热点，做成不同样式的云图，在课堂上展示给学生，让学生进行对比学习，不仅可以减轻教师在备课时的压力和负担，同时还可以增强课堂趣味、活跃课堂气氛，提高学生学习的积极性，有助于教学工作的开展和实施。

参考文献

- [1]张文宣. 浅谈韩语新闻对韩国语学习的促进作用[J]. 新闻研究导刊, 2016, 7(21): 28.

- [2]潘亚星. 基于Python的词云生成研究——以柴静的《看见》为例[J]. 电脑知识与技术, 2019, 15(24): 8-10.

- [3]赵永升, 郭新洁. 自然语言处理技术在高校韩语教学中的应用初探. 《高等教育教学改革研究(第七辑)》, 2020.

基金项目: 课题: 此论文为重庆市外国语言文学市级一流学科研究生科研创新项目研究成果

项目类别: 一般项目.

项目名称: 韩语新闻数据库的构建与高校韩语教学中的应用研究.

项目编号: SISUWYXY202049.