**农业大数据应用课程报告**

（20 年春（）学期）

课程名称： 农业大数据应用 任课教师：

学 号： 姓 名：

一、目标及意义

通过学习农业大数据应用课程，从python中学习到词云制作和代码编辑。以此实现数据分析和清洗，将专业文档的高频词汇展示到画布上。一来，检验自己的学习成果，加深对词云代码编写的理解。二来，用可视化技术来研究大数据在农业中的各种应用。比如，提取关键字，分析当下政策；清洗出专业文章的主题内核，使其简单易懂，一目了然。

# 二、过程分析

首先从网上下载相关农业大数据与粮食安全相关的主题专业文章，对文章进行简单处理，删除表格和参考文档等会影响词云展示的内容。根据文章内容提取停用词表，比如‘一些’、‘通常’、‘以及’等这些无意义连接词。使文章更简练，让词云更正确。

然后开始编写python代码，加入关键词汇，同义词等。设置词云格式，背景，颜色，形状等。使其运行后生成png格式的图片文件，并且展示在当前屏幕上。

三、项目代码及分析

# 1. 开始预处理文本

# 读取专业论文内容

with open(r'Z:/study/new/title.txt', 'r', encoding='utf-8') as f:

text = f.read()      #打开该路径下的文本文件。

# 添加专业词汇

terms = [

    '病虫害管理', '气象预测', '精细化种植', '智慧农机', '市场价格监测',

    '国家粮食安全'

]

for term in terms:

    jieba.add\_word(term)   #添加这些词，在后续分词过程中专业词汇不被切分

# 读取停用词表

stopwords = set()

with open(r'Z:/study/new/stopwords.txt', 'r', encoding='utf-8') as f:

    for line in f:

        stopwords.add(line.strip())  #读取该路径下的停用此表，一些数字，连接词都被收录停用切分统计。

# 分词 & 过滤停用词

words = jieba.lcut(text)

filtered\_words = [word for word in words if len(word) > 1 and word not in stopwords]

# 同义词替换

synonyms = {

    '不足': '制约', '国家粮食安全':'粮食安全'

}         #将部分意思相近的词汇做一次同义词替换，以免词云中出现多频同义词。

final\_words = [synonyms.get(word, word) for word in filtered\_words]

# 统计词频

word\_counts = Counter(final\_words)

top\_100 = word\_counts.most\_common(100)

#2. 生成词云

# 设置词云对象

wc = WordCloud(

    font\_path='C:/Windows/Fonts/msyh.ttc',  #设置文字

    background\_color='white',  #设置背景

    max\_words=100,   #设置频率

    collocations=False   #拒绝自动组合

)

# 根据频率生成词云

wc.generate\_from\_frequencies(dict(top\_100))

# 3. 输出词云

plt.figure(figsize=(12,8))#创建一块画布，设置大小

plt.imshow(wc, interpolation='bilinear')

plt.axis('off') #隐藏坐标

plt.savefig('new.png', dpi=300, bbox\_inches='tight')  #保存图片

plt.show()

# 四、运行结果



# 五、遇到的问题及解决方案

在编写程序的过程中，主要遇到了两个问题。

一是，没有添加专业词汇，导致在后续过程中被jeba自动切分，词云展示不完整。在代码里加入专业词汇解决了这一问题。

二是，停用词表并不完善，导致词云看上去有很多无意义的连接词。一边对照词云展示，一边修改stopwords文件，最后解决该问题。

六、结论及总结

通过该项目的实践和学习，我深入理解了数据清洗和词云相关技术。对农业大数据和现在的数据处理之间的联系更熟悉了，也察觉到该项目对于农业的发展有着长远深刻的意义。词云能更好更直白的显示专业文章的核心观点和政策输出，弥补了专业文章生涩冗长的不足。老百姓也能更容易得去理解专业文章输出的内容。

农业发展离不开科技，离不开大数据。如今大数据在各行各业开花结果，我意识到大数据相关的技术在推动农业发展中肯定会扮演越来越重要的角色。