



南京工业大学  
NANJING TECH  
UNIVERSITY

# 用户数据采集与关联分析

## (结课作业)

吴志祥

18205185639

1030624832@qq.com





南京工业大学  
NANJING TECH  
UNIVERSITY

# 01 文本数据分析

吴志祥

18205185639

[cnwzx2012@njtech.edu.cn](mailto:cnwzx2012@njtech.edu.cn)





# 第三讲 词云与可视化

1. 用任意一款词云工具，制作一个好看的词云（内容合理即可），并对词云图有一段话的解释。

The screenshot displays the Weiciyun (微词云) website interface. On the left, there is a sidebar with navigation options: 内容 (Content), 形状 (Shape), 配置 (Configuration), 插图 (Illustration), 字体 (Font), 快捷键 (Shortcuts), and 帮助 (Help). The main area is divided into two sections. The top section is the 'Import Words' (导入单词) panel, which includes a table for managing word lists. The bottom section is the 'Word Cloud' (词云) panel, which shows a generated word cloud with various words and phrases.

查找单词	字体	词频	颜色	角度	重复
<input type="checkbox"/> 无私奉献	默认	2	auto	auto	yes
<input type="checkbox"/> 1958年	默认	2	auto	auto	yes
<input type="checkbox"/> 按照采集工程的要求	默认	2	auto	auto	yes
<input type="checkbox"/> 手稿	默认	2	auto	auto	yes
<input type="checkbox"/> 报纸	默认	2	auto	auto	yes
<input type="checkbox"/> 这段时间也是黄旭	默认	2	auto	auto	yes
<input type="checkbox"/> 其一生的主要学术	默认	2	auto	auto	yes
<input type="checkbox"/> 在核潜艇领域	默认	1	auto	auto	yes
<input type="checkbox"/> 我国已形成一套完	默认	1	auto	auto	yes
<input type="checkbox"/> 设计	默认	1	auto	auto	yes
<input type="checkbox"/> 试验	默认	1	auto	auto	yes
<input type="checkbox"/> 制造	默认	1	auto	auto	yes
<input type="checkbox"/> 测试的核潜艇产业	默认	1	auto	auto	yes
<input type="checkbox"/> 而且装备了一支具	默认	1	auto	auto	yes
<input type="checkbox"/> 成梯次配备的	默认	1	auto	auto	yes
<input type="checkbox"/> 已近实现战备巡逻	默认	1	auto	auto	yes
<input type="checkbox"/> 回顾我国核潜艇的	默认	1	auto	auto	yes
<input type="checkbox"/> 人们自然会想起以	默认	1	auto	auto	yes
<input type="checkbox"/> 筚路蓝缕	默认	1	auto	auto	yes

The word cloud on the right features large, prominent text such as '报纸' (Newspaper), '1958年' (1958), '试验' (Experiment), '而今' (Now), '按照采集工程的要求' (According to the requirements of the collection engineering), '设计' (Design), '手稿' (Manuscript), '无私奉献' (Unselfish Dedication), and '制造' (Manufacture). The background of the word cloud is a collage of historical documents and newspaper clippings related to the development of nuclear submarines.

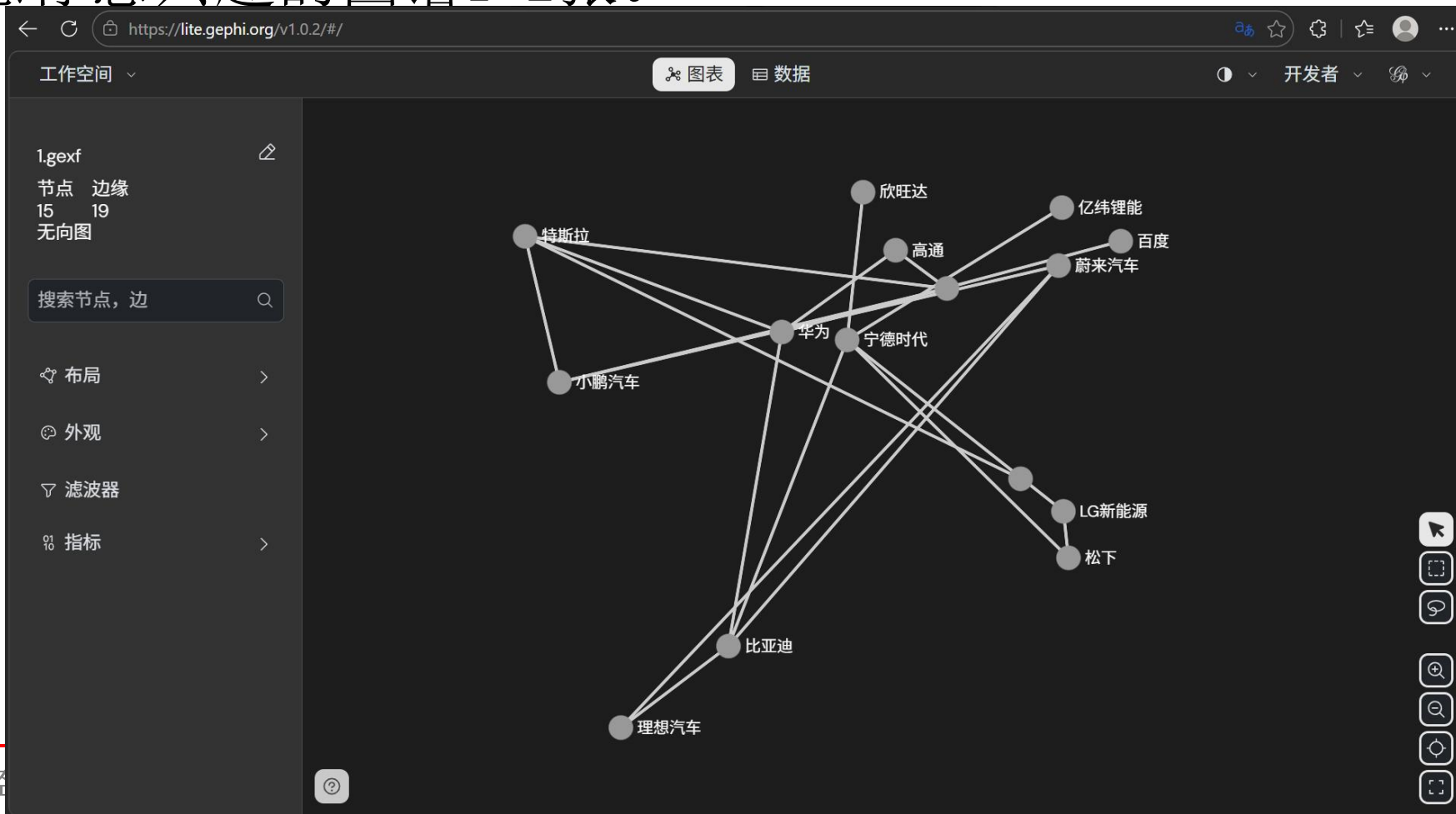
# 第三讲 词云与可视化

2. 使用Echarts，制作3个以上图，其中一个必须是“关系”，图的概念越明确（可解释，而不是自带的模板）越好。



# 第三讲 词云与可视化

3. 使用Gehpi、VOSViewer、CiteSpace…其中任意一款工具，绘制任意你感兴趣的图谱1-2张。



# 第三讲 词云

## 4. 采用给定的文本生成词云图 (生成的词云图)

```
1.py x
4 from PIL import Image
5
6 # --- 1. 准备文本数据 ---
7 # 这是一段关于“人工智能”的中文文本
8 text = """
9 人工智能（Artificial Intelligence，简称AI）是一门让计算机能够执行通常需要人类智能才能完成的任务的科学与工程学科。
10 它涵盖了机器学习、深度学习、自然语言处理、计算机视觉、专家系统等多个领域。
11 机器学习是人工智能的一个核心分支，它研究如何使计算机能够从数据中学习并改进，而无需显式编程。
12 深度学习作为机器学习的一个重要子集，通过构建深层神经网络模型，在图像识别、语音识别等领域取得了突破性进展。
13 自然语言处理致力于使计算机能够理解、解释和生成人类语言，为智能客服、机器翻译等应用提供了技术支持。
14 随着大数据和计算能力的飞速发展，人工智能技术正深刻影响着医疗、交通、金融、教育等各行各业，展现出巨大的发展潜力和前景。
15 """
16
17 # --- 2. 准备自定义形状的蒙版（Mask） ---
18 # 读取蒙版图片
19 try:
20     mask_image = np.array(Image.open("mask.png")) # 替换成你的蒙版图片文件名
21 except FileNotFoundError:
22     print("未找到蒙版图片 'mask.png'，将生成默认矩形词云。")
23     mask_image = None # 如果找不到图片，则不使用蒙版
24
25 # --- 3. 配置并生成词云 ---
26 # 设置中文字体路径（请根据你的操作系统和字体位置进行修改）
27 # Windows 示例: 'C:/Windows/Fonts/simhei.ttf'（黑体）
28 # macOS 示例: '/System/Library/Fonts/PingFang.ttc'（苹方）
29 # Linux 示例: '/usr/share/fonts/truetype/wqy/wqy-zenhei.ttc'（文泉驿正黑）
30 font_path = 'C:/Windows/Fonts/simhei.ttf'
31
32 # 创建 WordCloud 对象
33 wc = WordCloud(
34     font_path=font_path, # 设置字体，解决中文显示问题
35     background_color="white", # 设置背景颜色为白色
36     max_words=200, # 设置最大显示的词数
37     mask=mask_image, # 设置蒙版，使词云形状与蒙版一致
38     contour_width=3, # 如果使用蒙版，可以设置轮廓宽度
39     contour_color='steelblue', # 设置轮廓颜色
40     max_font_size=100, # 设置最大字体大小
41     random_state=42, # 设置随机种子，使每次生成的布局一致
42     colormap='viridis' # 设置颜色方案
43 )
44
45 # 从文本生成词云
46 wc.generate(text)
47
48 # --- 4. 显示和保存词云图 ---
49 # 使用 matplotlib 显示词云图
50 plt.figure(figsize=(10, 10)) # 设置图像大小
51 plt.imshow(wc, interpolation='bilinear')
52 plt.axis("off") # 关闭坐标轴
53 plt.show()
54
55 # 保存词云图到文件
56 wc.to_file("ai_wordcloud.png")
57 print("词云图已保存为 'ai_wordcloud.png'")
```

文本的词云图绘制，越清晰越好

