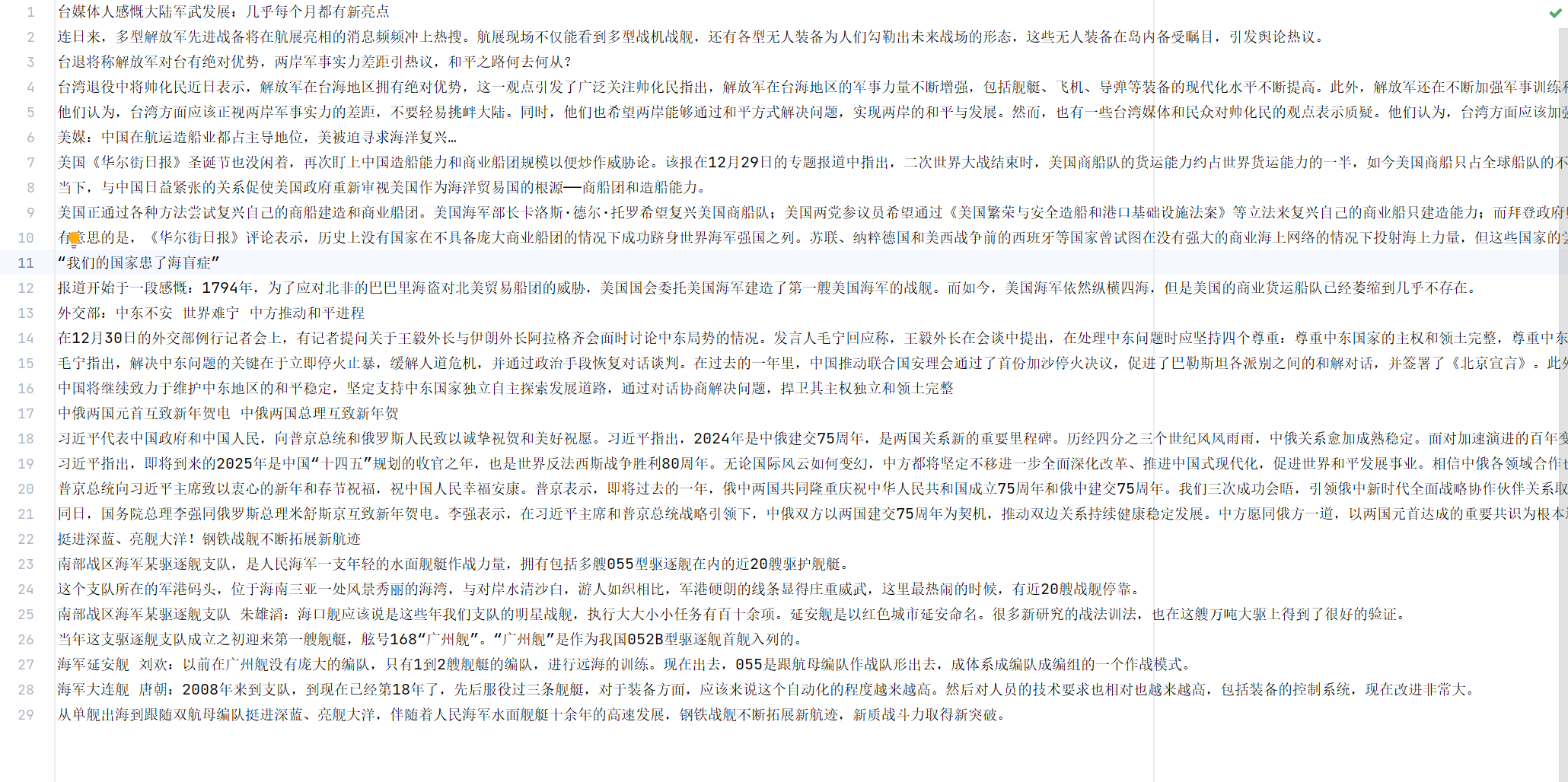
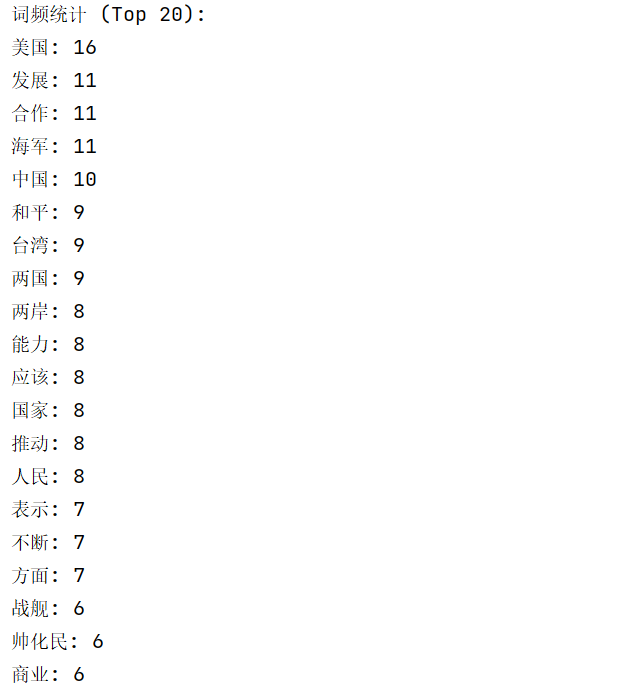
Sy8-军事新闻词云图

1. 数据采集与清洗

首先，在军事新闻网站中挑选5篇内容相关性较强的新闻文章，并将它们保存为一个文本文件（TXT格式）。同时，下载并整理一份中文停用词表，此表包含了常见的不携带实际意义的词语，如“和”、“是”等，用以过滤分词结果。



使用jieba库对保存的新闻文本进行分词处理。分词后，利用之前准备的停用词表过滤掉无关词汇。过滤完毕后，利用Python的collections.Counter类统计并存储每个词汇的出现频率。



1. 词云生成

将处理好的词频数据导入WordCloud库。设置词云图的参数，包括字体路径、背景颜色、最大词数等，确保生成的词云图美观且信息量丰富。使用PIL库将指定的背景图（JPG/PNG格式）加载为词云的形状。生成的词云图随后被转化为HTML格式的交互式图表，使用pyecharts库进行配置，确保用户在浏览词云时能够看到每个词汇的具体出现次数。

生成的HTML格式词云图存储于本地文件系统，用户可以直接用网页浏览器打开查看。词云图清晰地显示了如“美国”、“中国”、“军事演习”等关键词的出现频率，通过不同的字体大小表示词频的高低，为用户提供了直观的新闻关键词展示。



1. 实验结果

生成的词云图中，“美国”、“中国”、“合作”等词汇出现频率较高，这反映了当前军事新闻中常见的主题。此外，词如“海军”、“和平”等也频繁出现，显示了新闻关注的具体领域。

本实验成功应用了文本处理和数据可视化技术，展示了如何利用Python及其相关库来分析军事新闻内容。通过这种方式，不仅可以快速识别信息的重点，还能有效地将复杂的文本数据转化为直观、易理解的视觉格式。