用python 完成词云图

词云图是一种用来展现**高频关键词**的可视化表达，通过文字、色彩、图形的搭配，产生有冲击力地视觉效果，而且能够传达有价值的信息。词云图过滤掉大量的文本信息，使浏览网页者只要一眼扫过文本就可以领略文本的主旨。本次实验旨对已获取的一份文档进行预处理、文本分词、词频统计、词云展示，让同学们熟悉制作词云图的基本方法。



词云图样例

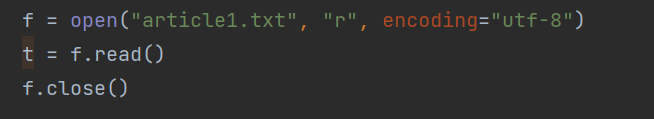
# 实验介绍

本次作业给出了初步的实验代码。

本节给出初步的实验代码WordCloud.py介绍：

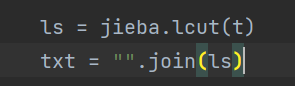
* 读取数据集

首先是读取待展示的数据



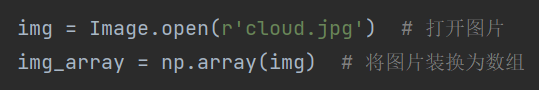
* 处理数据集

使用jieba提供的默认的分词机制进行数据处理



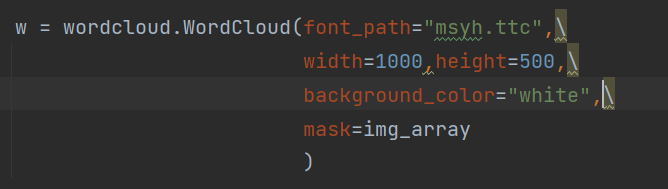
* 读取背景图片

背景图片必须是白色的背景

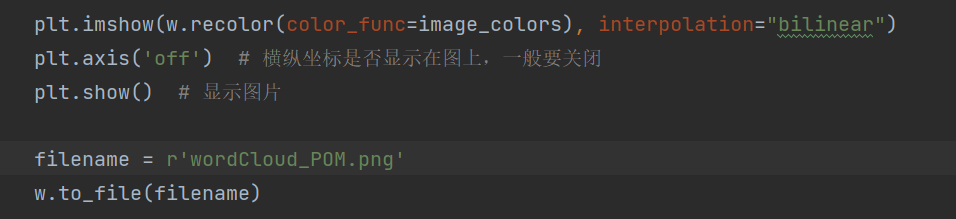


* 设置词云图格式

比如，字体格式、背景图片、大小等等



* 展示与保存



# 实验要求

本实验有以下三个方面的要求，根据自己的实际情况完成。

* 自己去查找不同的数据集，可以是歌词、小说、新年祝福语等等，基于自己找到的数据集进行绘制词云图。
* 能够改变词云图的基本格式，比如更换背景图片、更换字体格式、词云图大小等等。
* 本实验提供的代码是基于generate\_from\_text函数进行词频统计，该函数能够**自动统计词频**，并展示生成后的图。第三个要求是希望学生能够**自己实现**词频统计功能，并借助generate\_from\_frequencies绘制词云图（附了一份参考答案）。

最终需要完成实验报告，实验报告包含实验过程以及截图，我们会根据同学的实际完成情况来打分。

作业截止日期2022年1月5日23:59:59