# 项目说明

项目的code文件夹里有五个代码文件，这五个代码是层层递进的关系，展现了词云库由浅入深的应用。

文件1展现了词云库对英文文本的处理，在本示例中，打开了美国宪法的文本文件，并统计文件各词出现的频率，并绘制词云图。

文件2展现了词云库对中文文本的处理。由于英文单词之间以空格隔开，但中文没有这种分隔方式，所以若不进行分词处理的话，程序就不能统计规范词语的词频。

文件3中，我们导入了jieba库对中文文本进行分词处理，其原理是利用一个中文词库，将待分词的内容和分词词库进行比较，通过图结构和动态规划方法找到最大概率的词组。因此在进行分词处理之后，我们能够得到规范词语的词云图。

词云库还提供生成指定形状词云图的功能。因此在文件4中，我们导入了用于图像处理的PIL库。并从网上找了两个图片模板，将其转换为一个numpy库的多维数组，然后用函数就可以生成对应形状的词云图。

有时候我们希望，词频越高的词语在词云图中的大小就越大，因此我们在文件5中通过对词频排名前200的词语返回权重，并根据权重调整其在词云图中的大小。同时我们还可以根据导入图片的颜色，用ImageColorGenerator()函数导出其颜色，就可以让生成的词云图对应的位置和图片是一样的颜色。