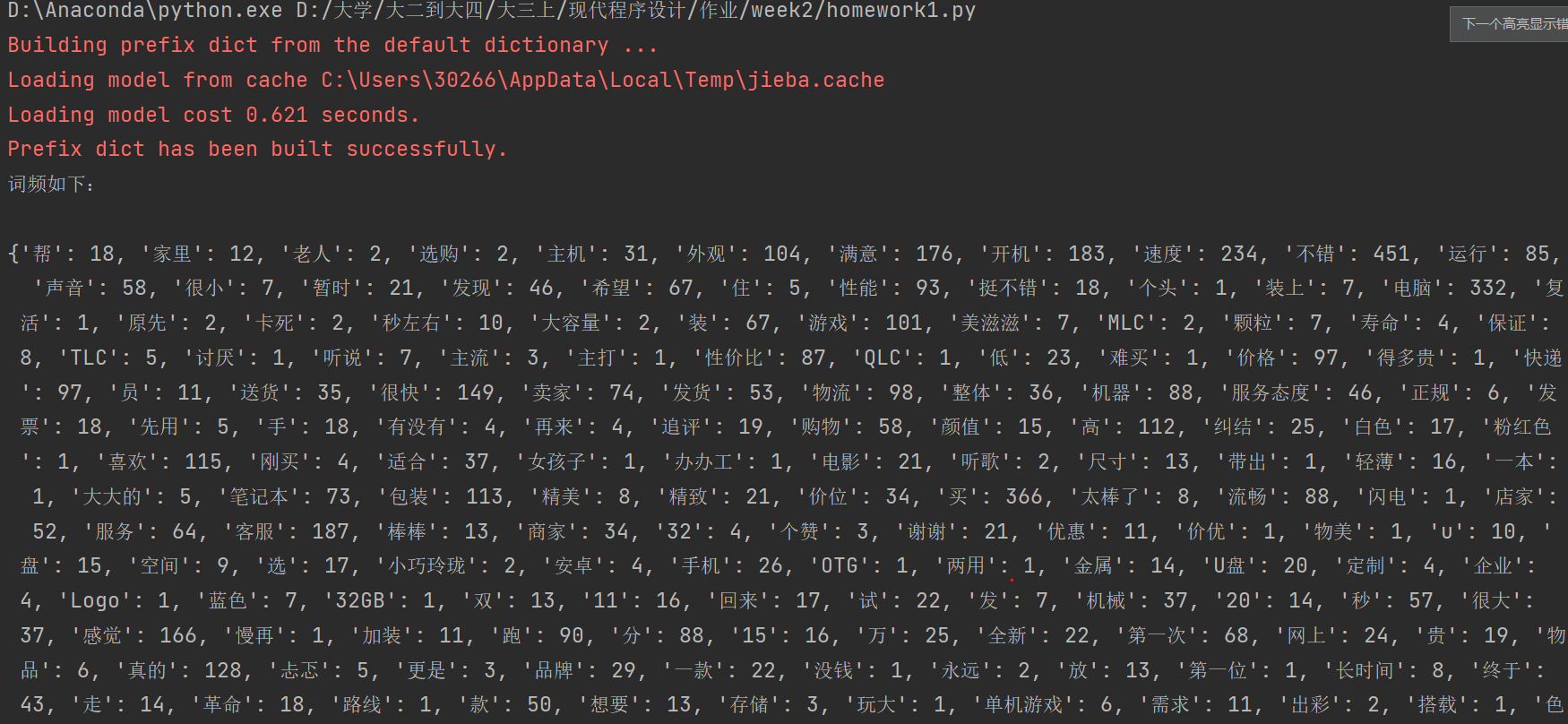
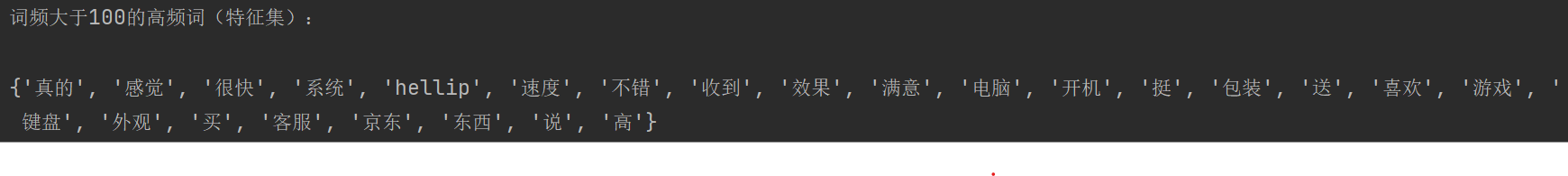
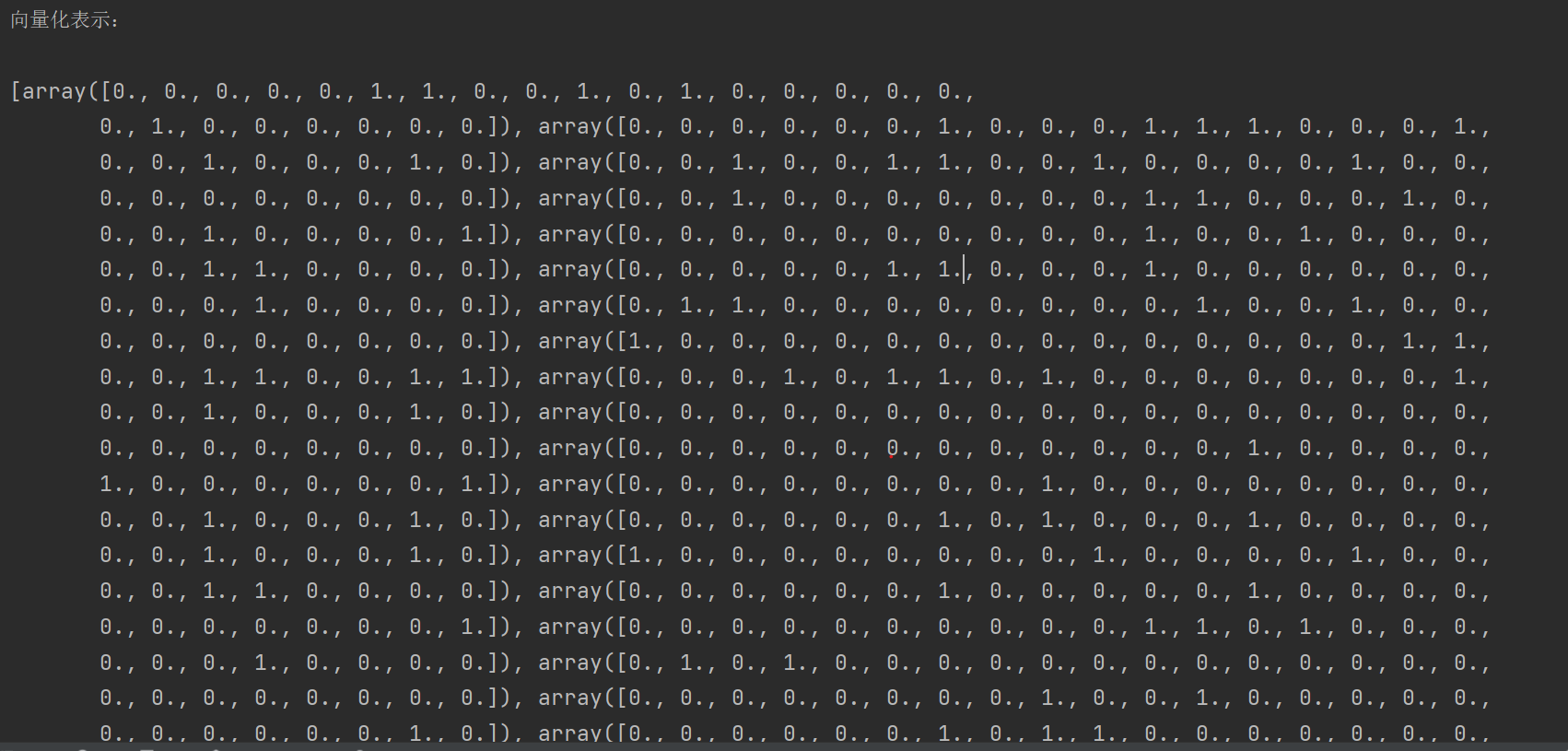
词频展示如下，由于完整词频字典太大，因此只展示部分，完整词频可运行源代码后查看：



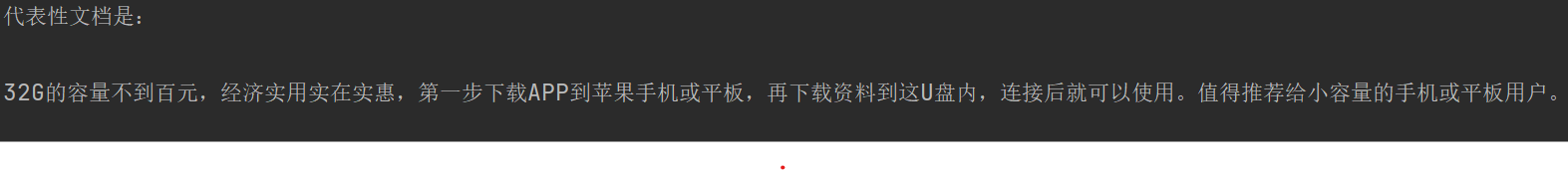
根据词频在100以上筛选的特征集如下：



使用one-hot表示的向量化文档部分展示如下：



根据欧氏距离计算出重心后，找出距离重心最近的文档作为代表性文档，输出如下：



绘制的词云图如下：

