1. 同学们可以按照文章说明，参考给定语料进行预处理操作；当然也可以自己找语料，参考信息增益或者TF-IDF算法进行文本的特征值抽取工作；
2. 同学们分小组完成，也可独立工作，最终汇总，形成ppt陈述

<https://www.kdnuggets.com/2017/12/general-approach-preprocessing-text-data.html>