热词发现

**实验内容：**

热词发现，通过统计给定数据中的单词出现的频数，找到最常用的单词。

**实验要求：**

本次实验要求分别使用 MapRedcue和Spark 计算框架，处理给定文本集，挖掘其中使用最频繁的 200个词汇，要求在大数据计算的过程中：1）将所有大小字母都变成小写字母；2）过滤掉所有标点符号和常见的停用词；3）过滤掉所有数字。

**作业提交材料：**

1. 实验报告；
2. 实验源代码（请只提交自己写的部分，而不是将整个工程打包）