1、下载如下数据集：

<http://www.sogou.com/labs/resource/r.php>

搜狗数据资源——中文词语搭配库

<http://www.sogou.com/labs/resource/e.php>

搜狗数据资源——网页搜索结果评价

实验目的及要求：

* 了解数据集作用及格式（本实验下载精简版即可）
* 在linux中解压缩并用相关文本命令进行数据查看、统计及分析。（包括但不限于：统计某关键词的词频、对文件进行排序（按照某列或某关键字），对大文件的快速浏览，管道命令、模糊查询（使用通配符）等）

2、按照ppt运行并熟悉相关命令。

3、熟悉vi或者vim的用法，掌握PPT中列表中的常用命令（如定位，复制，搜索等）