# 基于Aprori算法的医疗诊断规则的挖掘和分析

### 成员

* 严西敏 2120151051
* 欧阳童 2120151025

### 处理步骤

1. 用数据爬虫从网站上面抓取网页。
2. 文本预处理，对文本数据进行去重和过滤掉那些没有实际意义的词。
3. 分词，使用专用的医疗学词库，对文本进行分词
4. 抽取 症状 和 疾病 信息
5. 使用Aprori算法，计算症状和疾病之间的相关性
6. 根据第5步的规则，输出可信度高的规则

### 阶段性成果

1. 完成了爬虫抓取程序。（使用Python语言）



**图 使用爬虫程序获得文本资料**

1. 完成了对文本的分词处理。（使用Java语言）



**图 使用分词程序和词典过滤处理之后的结果**

1. 完成了使用Aprior算法对文本进行处理。（使用Java语言）

### 下阶段任务

串联起各个部分，使各部分都能完美衔接运行。