

**课 程 实 验 报 告**

**课程名称： 大数据分析**

**专业班级：**

**学 号：**

**姓 名：**

**指导教师：**

**报告日期：**

**计算机科学与技术学院**

**目录**

[实验三 关系挖掘实验 1](#_Toc58793746)

[**3.1 实验内容** 1](#_Toc58793747)

[**3.2 实验过程** 1](#_Toc58793748)

[3.2.1 编程思路 1](#_Toc58793749)

[3.2.2 遇到的问题及解决方式 1](#_Toc58793750)

[3.2.3 实验测试与结果分析 1](#_Toc58793751)

[**3.3 实验总结** 2](#_Toc58793752)

# 实验三 关系挖掘实验

## **3.1 实验内容**

**必做：**

1. 实验内容

编程实现Apriori算法，要求使用实验一得到的前1000个title及其引用关系作为实验数据。

从实验一中得到的引用关系数据为<<title,<title1,…,titlek>>,…>，将其处理为<<title,title1,…,titlek>,…>作为算法输入。

2. 实验要求

输出1~4阶频繁项集与关联规则，各个频繁项的支持度，各个规则的置信度，各阶频繁项集的数量以及关联规则的总数。

固定参数以方便检查，频繁项集的最小归一化支持度为0.15（与basket总数的比值），关联规则的最小置信度为0.3。

**加分项：**

1. 实验内容

在Apriori算法的基础上，要求使用pcy或pcy的几种变式multiHash、multiStage等算法对二阶频繁项集的计算阶段进行优化。

2. 实验要求

输出1~4阶频繁项集与关联规则，各个频繁项的支持度，各个规则的置信度，各阶频繁项集的数量以及关联规则的总数。

输出pcy或pcy变式算法中的vector的值，以bit位的形式输出。

固定参数以方便检查，频繁项集的最小支持度为0.15，关联规则的最小置信度为0.3。

## **3.2 实验过程**

### 3.2.1 编程思路

### 3.2.2 遇到的问题及解决方式

### 3.2.3 实验测试与结果分析

## **3.3 实验总结**