深 圳 大 学 实 验 报 告

课程名称： <u>Java 程序设计</u>

实验项目名称： <u>选实验 6 输入输出应用</u>

学院： <u>计算机与软件学院</u>

专业： <u>计算机科学与技术</u>

指导教师： <u>潘微科</u>

报告人：<u>黎浩然</u> 学号： <u>2018112061</u> 班级： <u>01</u>

实验时间：**2019 年 12 月 13 日（周五）-2019 年 12 月 18 日（周三）**

实验报告提交时间：_____

教务部制

实验目的与要求：

实验目的： 掌握 I/O 程序设计，能够读写本地文件等。

实验要求：

(1). 数据解析和统计

● https://snap.stanford.edu/data/web-Amazon.html 网站上有很多 Amazon 的数据集供研究人员下载使用。

● 本次实验使用 Jewelry.txt.gz 数据集，请下载后解压。

● 格式说明请看网页上的"Data Format"部分。

● 在报告中附上程序截图、运行结果（如每个输出文件前 10 行的截图等）和简要文字说明。

（1.1）使用 Java 语言读取解压后的文件（Jewelry.txt），并得到以下文件（40 分）：

● **product.txt：** 每行三列，以分号作为分隔符，第一列是 productID，第二列是 title，第三列是 price；表示该 product 的 ID、title 和 price。

● **user.txt：** 每行两列，以分号作为分隔符，第一列是 userID，第二列是 profileName；表示该 user 的 ID 和 name。

● **score.txt：** 每行三列，以分号作为分隔符，第一列是 userID，第二列是 productID，第三列是 score；表示(user, product, score)三元组。

● **review.txt：** 每行三列，以分号作为分隔符，第一列是 userID，第二列是 productID，第三列是 summary；表示(user, product, summary)三元组。

（1.2）使用 Java 语言对 score.txt 进行统计，并得到以下文件（20 分）：

● **productAvg.txt：** 每行两列，以分号作为分隔符，第一列是 productID，第二列是所有用户对该 product 的平均 score，表示该 product 的受欢迎程度。

● **userAvg.txt：** 每行两列，以分号作为分隔符，第一列是 userID，第二列是该 user 对所有 product 的平均 score，表示该用户打分的高低。

（1.3）其他统计结果，每个统计结果 10 分，共 20 分，可自由发挥，例如对 summary 或 text 中的单词信息进行统计。

报告写作。要求：主要思路有明确的说明，重点代码有详细的注释，行文逻辑清晰可读性强，报告整体写作较为专业。**（20 分，这一项的评分，采用 20 分、10 分和 0 分三个级别）**

说明：

（1）本次实验课作业满分为 100 分，占总成绩的比例（待定）。

（2）本次实验课作业截至时间 2019 年 12 月 18 日（周三）23:59。

（3）报告正文：请在指定位置填写，本次实验不需要单独提交源程序文件。

（4）个人信息：WORD 文件名中的"姓名"、"学号"，请改为你的姓名和学号；实验报告的首页，请准确填写"学院"、"专业"、"报告人"、"学号"、"班级"、"实验报告提交时间"等信息。

（5）提交方式：截至时间前，请在 Blackboard 平台中提交。

（6）发现抄袭（包括复制&粘贴整句话、整张图），该次作业记零分。

（7）延迟提交，不得分；如有特殊情况，请于截至日期之后 48 小时内发邮件到 panweike@szu.edu.cn，并在邮件中注明课程名称、作业名称、姓名、学号等信息，以及特殊情况说明，我收到后会及时回复。

（8）期末考试阶段补交无效。

（1.1）使用 Java 语言读取解压后的文件（Jewelry.txt），并得到以下文件（40 分）：

- **product.txt**：每行三列，以分号作为分隔符，第一列是 productID，第二列是 title，第三列是 price；表示该 product 的 ID、title 和 price。

- **user.txt**：每行两列，以分号作为分隔符，第一列是 userID，第二列是 profileName；表示该 user 的 ID 和 name。

- **score.txt**：每行三列，以分号作为分隔符，第一列是 userID，第二列是 productID，第三列是 score；表示(user, product, score)三元组。

- **review.txt**：每行三列，以分号作为分隔符，第一列是 userID，第二列是 productID，第三列是 summary；表示(user, product, summary)三元组。

- **product.txt：**

程序如下

```java
import java.io.*;
public class Product {

    public static void main(String[] args) {
        // TODO Auto-generated method stub
        String inpath = "C:\\Users\\Administrator\\Desktop\\jewelry.txt";    //读入文件路径
        String outpath = "C:\\Users\\Administrator\\Desktop\\product.txt";   //写入文件路径

        try {
            FileInputStream fileinputstream = new FileInputStream(inpath);          //文件输入流
            InputStreamReader inputstreamreader = new InputStreamReader(fileinputstream,"gbk");
            BufferedReader reader = new BufferedReader(inputstreamreader);

            File fileout = new File(outpath);      //输出文件
            fileout.createNewFile();               //建立新文件
            FileWriter filewriter = new FileWriter(fileout);     //文件写出
            BufferedWriter out = new BufferedWriter(filewriter);

            while(true) {
                String line = reader.readLine();          //每次读入一行
                if(line==null) {                          //空行跳过
                    continue;
                }
                if(fileinputstream.read()==-1)            //文件尾跳出
                    break;

                if(line.contains("product/productId:")) {
                    String newline = line.substring(18);
                    out.write(newline);
                }
                else if(line.contains("product/title:")) {
                    String newline = line.substring(14);
                    out.write(";"+newline);
                }
                else if(line.contains("product/price:")) {
                    String newline = line.substring(14);
                    out.write(";"+newline+"\n");
                }
            }
;           }

            inputstreamreader.close();
            out.close();

        } catch (FileNotFoundException e) {
            // TODO Auto-generated catch block
            e.printStackTrace();
        } catch (UnsupportedEncodingException e) {
            // TODO Auto-generated catch block
            e.printStackTrace();
        } catch (IOException e) {
            // TODO Auto-generated catch block
            e.printStackTrace();
        }

    }
```

运行结果如下： product.txt

```
B000FTPOMK; 14k Yellow Gold Butterfly Pendant, 16"; unknown
B000FTPOMK; 14k Yellow Gold Butterfly Pendant, 16"; unknown
B000FTPOMK; 14k Yellow Gold Butterfly Pendant, 16"; unknown
B000FTPOMK; 14k Yellow Gold Butterfly Pendant, 16"; unknown
B000FTPOMK; 14k Yellow Gold Butterfly Pendant, 16"; unknown
B000PAMAPI; Sterling Silver Scroll Design Garnet Ring by Sajen, Size 7; unknown
B000PEM42S; Sterling Silver Oval Shaped Locket; unknown
B000KN6ZOM; Amazon.com: Silver Loving Family Necklace 1 Parent, 3 CHILDREN: Clothing; unknown
B000PCZ6UC; 14K Yellow Gold Ruby and Diamond Bracelet; unknown
B000PAMAQM; Sterling Silver Malachite Scroll Design Ring by Sajen, Size 5; unknown
B000L3RWXE; Purple Genuine Russian Charoite and Rainbow Moonstone Sterling Silver Slide Pendant; 69.99
B000L3RWXE; Purple Genuine Russian Charoite and Rainbow Moonstone Sterling Silver Slide Pendant; 69.99
B000L3RWXE; Purple Genuine Russian Charoite and Rainbow Moonstone Sterling Silver Slide Pendant; 69.99
B000L3RWXE; Purple Genuine Russian Charoite and Rainbow Moonstone Sterling Silver Slide Pendant; 69.99
B0006SE2Q4; Sterling Silver Garnet Lady Knight Armor Ring(Sizes 5,6,7,8,9,10); 34.97
B0006SE2Q4; Sterling Silver Garnet Lady Knight Armor Ring(Sizes 5,6,7,8,9,10); 34.97
B0006SE2Q4; Sterling Silver Garnet Lady Knight Armor Ring(Sizes 5,6,7,8,9,10); 34.97
B000FIPQYW; Rembrandt Charms Ballet Shoe Charm; 150.00
B0006SE2QO; Sterling Silver Baby Feet Footprints Spin Ring(Sizes 4,5,5.5,6,6.5,7,7.5,8,8.5); unknown
B000IBT1A6; Oval Cut Tigereye Dangling Heart Earrings Sterling Silver; unknown
B0000VRY1C; Elephant; 16.00
B000EVT0CO; 10K White Gold 1/4 ct. Diamond Tennis Bracelet (8"); 800.00
B000EVS4XK; SILVER-Colored Tie Clip Bar with Dolphin Tie Clasp; unknown
B000EVS4XK; SILVER-Colored Tie Clip Bar with Dolphin Tie Clasp; unknown
B000EVS4XK; SILVER-Colored Tie Clip Bar with Dolphin Tie Clasp; unknown
B0002UPI32; Playboy Bunny Light Blue LOGO Silhouette Belly Ring; unknown
B000FIRI1G; Rembrandt Charms Book Charm; 326.00
B000FIRI1G; Rembrandt Charms Book Charm; 326.00
B0006SE2SM; Sterling Silver AMETHYST Armour Ring Size 7.5; unknown
```

- **user.txt**

变化的代码如下

```java
        File fileout = new File(outpath);    //输出文件
        fileout.createNewFile();             //建立新文件
        FileWriter filewriter = new FileWriter(fileout);    //文件写出
        BufferedWriter out = new BufferedWriter(filewriter);

        while(true) {
            String line = reader.readLine();        //每次读入一行
            if(line==null) {                        //空行跳过
                continue;
            }
            if(fileinputstream.read()==-1)          //文件尾跳出
                break;

            if(line.contains("review/userId:")) {
                String newline = line.substring(14);
                out.write(newline);
            }
            else if(line.contains("review/profileName:")) {
                String newline = line.substring(19);
                out.write(";"+newline+"\n");
            }

        }

        reader.close();
        out.close();
    } catch (FileNotFoundException e) {
        // TODO Auto-generated catch block
```

运行结果如下:

```
A3T4KKUM1JATUX; Disappointed Sony customer
A33LVMKCG9IG15; A. Thorpe "Amazon lover"
A1PNSOEZQ4E9UT; rabbit "amazon rocks"
A3F12GSAOU0WRT; Jennifer S. Barbier
A2H69HCLAE6GME; Julia
A1U36NQMIRUHST; C. Tremari "cactuspetelady"
A2C4IVOBTG9W81; Not my name
A1KH38J5M2WCTX; greenchic
AABOLDP3KO54N; Mac Fan
A31178QVC2BZC6; Hibiscus
A2X6ADHDF17MQB; A.Mojica "fuego magico"
A1PKH1VARVC8R8; Christopher Kuzmich "Chris K."
A480GWA0CM6ZU; suzanna
A1T4TZB1CJNUUL; E. Brooks "jewel lover"
A2T8TQ8EXVCCFF; Christina Christner "Silver Glutton"
A2XSV06YRA9KFG; Y. Vera "FLY N RN"
```

- **scores.txt 变更代码如下：**

```java
int col = 0;
boolean start = false;      //是否已开始
String first = null, second = null, third = null;      //分别代表三列
while(true) {
    String line = reader.readLine();      //每次读入一行
    if(col%3==0&&start) {
        out.write("("+first+";"+second+";"+third+")\n");
    }
    if(line==null) {                      //空行跳过
        continue;
    }
    if(fileinputstream.read()==-1)        //文件尾跳出
        break;

    if(line.contains("review/userId:") ){
        first = line.substring(14);
        col++;
    }
    else if(line.contains("product/productId:")) {
        second = line.substring(18);
        col++;
    }
    else if(line.contains("review/score:")) {
        third = line.substring(13);
        col++;
        start = true;
    }
    else
        start = false;
}
```

运行结果：

```
( A3T4KKUM1JATUX; B000FTPOMK; 4.0)
( A33LVMKCG9IG15; B000FTPOMK; 4.0)
( A1PNSOEZQ4E9UT; B000FTPOMK; 5.0)
( A3F12GSAOU0WRT; B000FTPOMK; 5.0)
( A2H69HCLAE6GME; B000FTPOMK; 5.0)
( A1U36NQMIRUHST; B000PAMAPI; 5.0)
( A2C4IVOBTG9W81; B000PEM42S; 3.0)
( A1KH38J5M2WCTX; B000KN6ZOM; 1.0)
( AABOLDP3KO54N; B000PCZ6UC; 5.0)
( A31178QVC2BZC6; B000PAMAQM; 5.0)
( A2X6ADHDF17MQB; B000L3RWXE; 4.0)
( A1PKH1VARVC8R8; B000L3RWXE; 4.0)
```

- **review.txt：**

变化代码如下:

```java
        int col = 0;
        boolean start = false;       //是否已开始
        String first = null, second = null, third = null;        //分别代表三列
        while(true) {
            String line = reader.readLine();       //每次读入一行
            if(col%3==0&&start) {
                out.write("("+first+";"+second+";"+third+")\n");
            }
            if(line==null) {                 //空行跳过
                continue;
            }
            if(fileinputstream.read()==-1)          //文件尾跳出
                break;

            if(line.contains("review/userId:") ){
                first = line.substring(14);
                col++;
            }
            else if(line.contains("product/productId:")) {
                second = line.substring(18);
                col++;
            }
            else if(line.contains("review/summary:")) {
                third = line.substring(15);
                col++;
                start = true;
            }
            else
                start = false;
        }
```

运行结果：

```
( A3T4KKUM1JATUX; B000FTPOMK; pretty necklace)
( A33LVMKCG9IG15; B000FTPOMK; pretty necklace)
( A1PNSOEZQ4E9UT; B000FTPOMK; Nice and weighty)
( A3F12GSAOU0WRT; B000FTPOMK; Great necklace for kids or adults)
( A2H69HCLAE6GME; B000FTPOMK; Absolutely satisfied with it)
( A1U36NQMIRUHST; B000PAMAPI; Beautiful ring)
( A2C4IVOBTG9W81; B000PEM42S; Just ok...)
( A1KH38J5M2WCTX; B000KN6ZOM; Ordered 2; Both Chains Were Tarnished and Pendants Scratched!)
( AABOLDP3KO54N; B000PCZ6UC; Good Value)
( A31178QVC2BZC6; B000PAMAQM; nice quality)
( A2X6ADHDF17MQB; B000L3RWXE; Dainty)
( A1PKH1VARVC8R8; B000L3RWXE; beautiful)
( A480GWA0CM6ZU; B000L3RWXE; Bought as gift, beautiful...)
( A1T4TZB1CJNUUL; B000L3RWXE; Love it!)
( A2T8TQ8EXVCCFF; B0006SE2Q4; Excellent value!)
( A2XSV06YRA9KFG; B0006SE2Q4; RING)
( A1SGC0Z8SG2YDE; B0006SE2Q4; Thin and bends easily, but very pretty)
( AG9NZH76KYULL; B000FIPQYW; Pointe Shoe)
( A1I0ZC4LL59MHH; B0006SE2QO; Nice Ring)
( A35TOVFPICXJBU; B000IBT1A6; Completely worth it)
( A31QJ8DWYXVWD6; B0000VRY1C; one elephant charm? you've got to be kidding.)
( A2LNHEOFI0I91Z; B000EVT0CO; Poor quality and Customer Service- buyer beware)
( A2601GEMAVMIZH; B000EVS4XK; good pin for casual occasions)
```

（1.2）使用 Java 语言对 score.txt 进行统计，并得到以下文件（20 分）：

- **productAvg.txt**：每行两列，以分号作为分隔符，第一列是 productID，第二列是所有用户对该 product 的平均 score，表示该 product 的受欢迎程度。

- **userAvg.txt**：每行两列，以分号作为分隔符，第一列是 userID，第二列是该 user 对所有 product 的平均 score，表示该用户打分的高低。

productAvg.txt：

程序代码如下：

```java
String [] productid = new String[50000];    //假定产品种类不会超过50000
int [] usernum = new int[50000];             //每种产品的用户数量
float [] scores = new float [50000];         //每种产品的总分
for(int i = 0; i < 50000; ++i) {    //初始化
    productid[i] = null;
    usernum[i] = 0;
    scores[i] = (float) 0;
}

int num = 0;    //产品总数量
```

```java
        while(true) {
            String line = reader.readLine();              //每次读入一行
            if(line==null)  continue;                     //空行跳过
            if(fileinputstream.read()==-1) break;         //文件尾跳出

            if(line.contains("product/productId:")) {     //找到一个产品
                String newline = line.substring(18);      //获取ID
                int i;
                for(i = 0; i < num; ++i)                  //查看是否已记录过该产品
                    if(productid[i].equals(newline))
                        break;
                usernum[i]++;                             //用户数量加一
                if(i==num) {                              //没有该产品
                    productid[i] = newline;
                    num++;
                }

                do {        //查看对应分数
                    line = reader.readLine();
                    if(line==null) {                      //空行跳过
                        continue;
                    }
                }while(!line.contains("review/score:"));

                try {       //考虑到部分格式有问题则不统计
                    newline = line.substring(14);
                    float score = Float.valueOf(newline);  //获取分数
                    scores[i] += score;                    //加到总分
                } catch(Exception e) {                     //有异常不统计该用户
                    usernum[i]--;
                    continue;
```

```java
                    usernum[i]--;
                    continue;
                }

            }
        }

        for(int i = 0; i < num; ++i)                 //输出
            out.write(productid[i]+";"+scores[i]/usernum[i] + "\n");

        reader.close();
        out.close();
    } catch (FileNotFoundException e) {
        // TODO Auto-generated catch block
        e.printStackTrace();
    } catch (UnsupportedEncodingException e) {
        // TODO Auto-generated catch block
```

统计结果如下：

```
B000FTPOMK;4.6
B000PAMAPI;5.0
B000PEM42S;3.0
B000KN6ZOM;1.0
B000PCZ6UC;5.0
B000PAMAQM;5.0
B000L3RWXE;4.5
B0006SE2Q4;4.0
B000FIPQYW;5.0
B0006SE2QO;4.0
B000IBT1A6;5.0
B0000VRY1C;1.0
B000EVT0CO;1.0
B000EVS4XK;5.0
B0002UPI32;3.0
B000FIRI1G;5.0
B0006SE2SM;3.0
B000FIRI2O;5.0
B000HWS1AW;5.0
B000O3KHWE;5.0
B000EDH1KA;3.0
B000IBT1EC;5.0
```

userAvg：程序代码：

```java
String [] userid = new String[50000];        //假定用户会不会超过50000
int [] productnum = new int[50000];           //每个用户的产品数量
float [] scores = new float [50000];          //每个用户的总分
for(int i = 0; i < 50000; ++i) {     //初始化
    userid[i] = null;
    productnum[i] = 0;
    scores[i] = (float) 0;
}

int num = 0;    //用户总数量
```

```
        while(true) {
            String line = reader.readLine();          //每次读入一行
            if(line==null)  continue;                 //空行跳过
            if(fileinputstream.read()==-1) break;      //文件尾跳出

            if(line.contains("review/userId:")) {      //找到一个用户
                String newline = line.substring(15);    //获取用户ID
                int i;
                for(i = 0; i < num; ++i)                //查看是否已记录过该用户
                    if(userid[i].equals(newline))
                        break;
                productnum[i]++;                        //产品数量加一
                if(i==num) {                            //没有该用户
                    userid[i] = newline;
                    num++;
                }
                do {     //查看对应分数
                    line = reader.readLine();
                    if(line==null) {                            //空行跳过
                        continue;
                    }
                }while(!line.contains("review/score:"));
                try {       //考虑到部分格式有问题则不统计
                    newline = line.substring(14);
                    float score = Float.valueOf(newline);       //获取分数
                    scores[i] += score;          //加到总分
                } catch(Exception e) {             //有异常不统计该产品
                    productnum[i]--;
                    continue;
                }
```

统计结果如下：

```
A3T4KKUM1JATUX;4.0
A33LVMKCG9IG15;4.0
A1PNSOEZQ4E9UT;5.0
A3F12GSAOU0WRT;5.0
A2H69HCLAE6GME;4.5
A1U36NQMIRUHST;5.0
A2C4IVOBTG9W81;3.0
A1KH38J5M2WCTX;1.0
AABOLDP3KO54N;5.0
A31178QVC2BZC6;4.8
A2X6ADHDF17MQB;4.6666665
A1PKH1VARVC8R8;4.625
A480GWA0CM6ZU;5.0
A1T4TZB1CJNUUL;5.0
A2T8TQ8EXVCCFF;5.0
A2XSV06YRA9KFG;4.0
A1SGC0Z8SG2YDE;4.0
AG9NZH76KYULL;5.0
A1I0ZC4LL59MHH;4.9545455
A35TOVFPICXJBU;5.0
A31QJ8DWYXVWD6;1.0
A2LNHEOFI0I91Z;1.0
A2601GEMAVMIZH;5.0
```

（1.3）其他统计结果，每个统计结果 10 分，共 20 分，可自由发挥，例如对 summary 或 text 中的单词信息进行统计。

统计一：

对每组数据统计 summary 单词数，每行格式为 product id + title + summary 词数

代码如下：

```java
            File fileout = new File(outpath);        //输出文件
            fileout.createNewFile();                  //建立新文件
            FileWriter filewriter = new FileWriter(fileout);  //文件写出
            BufferedWriter out = new BufferedWriter(filewriter);

            while(true) {
                String line = reader.readLine();        //每次读入一行
                if(line==null)  continue;               //空行跳过
                if(fileinputstream.read()==-1) break;    //文件尾跳出

                if(line.contains("product/productId:")) {
                    String newline = line.substring(18);
                    out.write(newline);
                }
                else if(line.contains("product/title:")) {
                    String newline = line.substring(14);
                    out.write(";"+newline);
                }
                else if(line.contains("review/summary:")) {
                    String newline = line.substring(16);
                    String regex = "[\\s]+";       //分割
                    String [] words = newline.split(regex);
                    int num = words.length;         //获取长度
                    out.write(";     "+num+"\n");
                }

            }
```
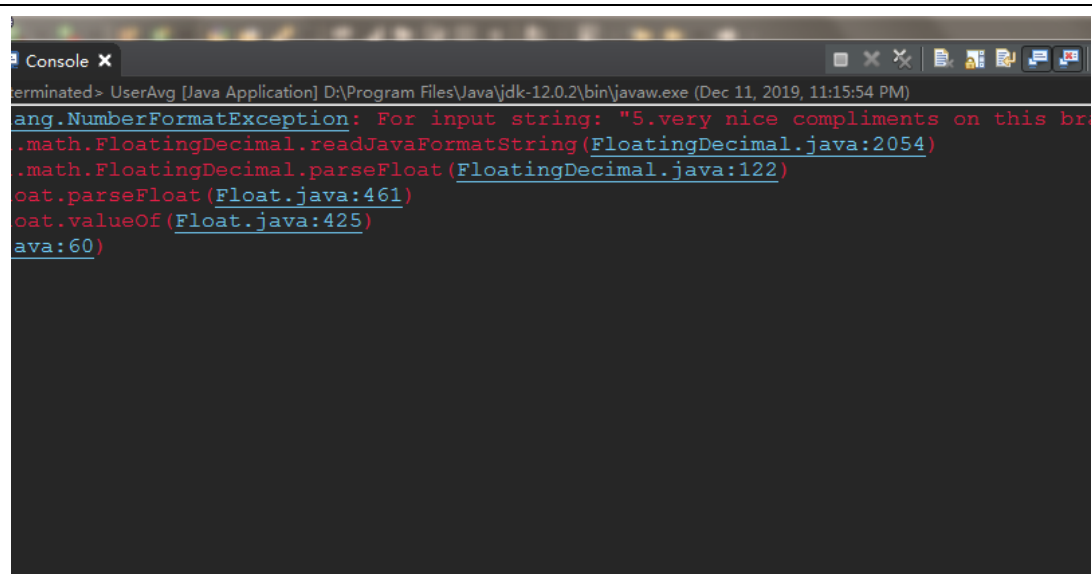
运行结果如下：

```
B000FTPOMK; 14k Yellow Gold Butterfly Pendant, 16";    2
B000FTPOMK; 14k Yellow Gold Butterfly Pendant, 16";    2
B000FTPOMK; 14k Yellow Gold Butterfly Pendant, 16";    3
B000FTPOMK; 14k Yellow Gold Butterfly Pendant, 16";    6
B000FTPOMK; 14k Yellow Gold Butterfly Pendant, 16";    4
B000PAMAPI; Sterling Silver Scroll Design Garnet Ring by Sajen, Size 7;    2
B000PEM42S; Sterling Silver Oval Shaped Locket;    2
B000KN6ZOM; Amazon.com: Silver Loving Family Necklace 1 Parent, 3 CHILDREN: Clothing;    9
B000PCZ6UC; 14K Yellow Gold Ruby and Diamond Bracelet;    2
B000PAMAQM; Sterling Silver Malachite Scroll Design Ring by Sajen, Size 5;    2
B000L3RWXE; Purple Genuine Russian Charoite and Rainbow Moonstone Sterling Silver Slide Pendant;    1
B000L3RWXE; Purple Genuine Russian Charoite and Rainbow Moonstone Sterling Silver Slide Pendant;    1
B000L3RWXE; Purple Genuine Russian Charoite and Rainbow Moonstone Sterling Silver Slide Pendant;    4
B000L3RWXE; Purple Genuine Russian Charoite and Rainbow Moonstone Sterling Silver Slide Pendant;    2
B0006SE2Q4; Sterling Silver Garnet Lady Knight Armor Ring(Sizes 5,6,7,8,9,10);    2
B0006SE2Q4; Sterling Silver Garnet Lady Knight Armor Ring(Sizes 5,6,7,8,9,10);    1
B0006SE2Q4; Sterling Silver Garnet Lady Knight Armor Ring(Sizes 5,6,7,8,9,10);    7
B000FIPQYW; Rembrandt Charms Ballet Shoe Charm;    2
B0006SE2QO; Sterling Silver Baby Feet Footprints Spin Ring(Sizes 4,5,5.5,6,6,6.5,7,7.5,8,8.5);    2
B000IBT1A6; Oval Cut Tigereye Dangling Heart Earrings Sterling Silver;    3
B0000VRY1C; Elephant;    8
B000EVT0CO; 10K White Gold 1/4 ct. Diamond Tennis Bracelet (8");    7
B000EVS4XK; SILVER-Colored Tie Clip Bar with Dolphin Tie Clasp;    5
B000EVS4XK; SILVER-Colored Tie Clip Bar with Dolphin Tie Clasp;    4
B000EVS4XK; SILVER-Colored Tie Clip Bar with Dolphin Tie Clasp;    6
B0002UPI32; Playboy Bunny Light Blue LOGO Silhouette Belly Ring;    2
B000FIRI1G; Rembrandt Charms Book Charm;    6
```

统计一：

对每组数据统计 text 单词数，每行格式为 product id + title + text 词数

代码如下：

```java
        InputStreamReader inputstreamreader = new InputStreamReader(fileinputstream,"gbk");
        BufferedReader reader = new BufferedReader(inputstreamreader);

        File fileout = new File(outpath);    //输出文件
        fileout.createNewFile();             //建立新文件
        FileWriter filewriter = new FileWriter(fileout);     //文件写出
        BufferedWriter out = new BufferedWriter(filewriter);

        while(true) {
            String line = reader.readLine();        //每次读入一行
            if(line==null)  continue;               //空行跳过
            if(fileinputstream.read()==-1) break;         //文件尾跳出

            if(line.contains("product/productId:")) {
                String newline = line.substring(18);
                out.write(newline);
            }
            else if(line.contains("product/title:")) {
                String newline = line.substring(14);
                out.write(";"+newline);
            }
            else if(line.contains("review/text:")) {
                String newline = line.substring(13);
                String regex = "[\\s]+";      //分割
                String [] words = newline.split(regex);
                int num = words.length;      //获取长度
                out.write(";    text words:"+num+"\n");
            }
        }

    }
```

统计结果：

```
B000FTPOMK; 14k Yellow Gold Butterfly Pendant, 16";    text words:15
B000FTPOMK; 14k Yellow Gold Butterfly Pendant, 16";    text words:42
B000FTPOMK; 14k Yellow Gold Butterfly Pendant, 16";    text words:15
B000FTPOMK; 14k Yellow Gold Butterfly Pendant, 16";    text words:24
B000FTPOMK; 14k Yellow Gold Butterfly Pendant, 16";    text words:26
B000PAMAPI; Sterling Silver Scroll Design Garnet Ring by Sajen, Size 7;    text words:145
B000PEM42S; Sterling Silver Oval Shaped Locket;    text words:89
B000KN6ZOM; Amazon.com: Silver Loving Family Necklace 1 Parent, 3 CHILDREN: Clothing;    text words:53
B000PCZ6UC; 14K Yellow Gold Ruby and Diamond Bracelet;    text words:21
B000PAMAQM; Sterling Silver Malachite Scroll Design Ring by Sajen, Size 5;    text words:36
B000L3RWXE; Purple Genuine Russian Charoite and Rainbow Moonstone Sterling Silver Slide Pendant;    text words:72
B000L3RWXE; Purple Genuine Russian Charoite and Rainbow Moonstone Sterling Silver Slide Pendant;    text words:34
B000L3RWXE; Purple Genuine Russian Charoite and Rainbow Moonstone Sterling Silver Slide Pendant;    text words:35
B000L3RWXE; Purple Genuine Russian Charoite and Rainbow Moonstone Sterling Silver Slide Pendant;    text words:20
B0006SE2Q4; Sterling Silver Garnet Lady Knight Armor Ring(Sizes 5,6,7,8,9,10);    text words:327
B0006SE2Q4; Sterling Silver Garnet Lady Knight Armor Ring(Sizes 5,6,7,8,9,10);    text words:62
B0006SE2Q4; Sterling Silver Garnet Lady Knight Armor Ring(Sizes 5,6,7,8,9,10);    text words:66
B000FIPQYW; Rembrandt Charms Ballet Shoe Charm;    text words:32
B0006SE2QO; Sterling Silver Baby Feet Footprints Spin Ring(Sizes 4,5,5.5,6,6.5,7,7.5,8,8.5);    text words:54
B000IBT1A6; Oval Cut Tigereye Dangling Heart Earrings Sterling Silver;    text words:51
B0000VRY1C; Elephant;    text words:22
B000EVT0CO; 10K White Gold 1/4 ct. Diamond Tennis Bracelet (8");    text words:139
B000EVS4XK; SILVER-Colored Tie Clip Bar with Dolphin Tie Clasp;    text words:11
B000EVS4XK; SILVER-Colored Tie Clip Bar with Dolphin Tie Clasp;    text words:25
B000EVS4XK; SILVER-Colored Tie Clip Bar with Dolphin Tie Clasp;    text words:28
B0002UPI32; Playboy Bunny Light Blue LOGO Silhouette Belly Ring;    text words:19
B000FIRI1G; Rembrandt Charms Book Charm;    text words:16
```
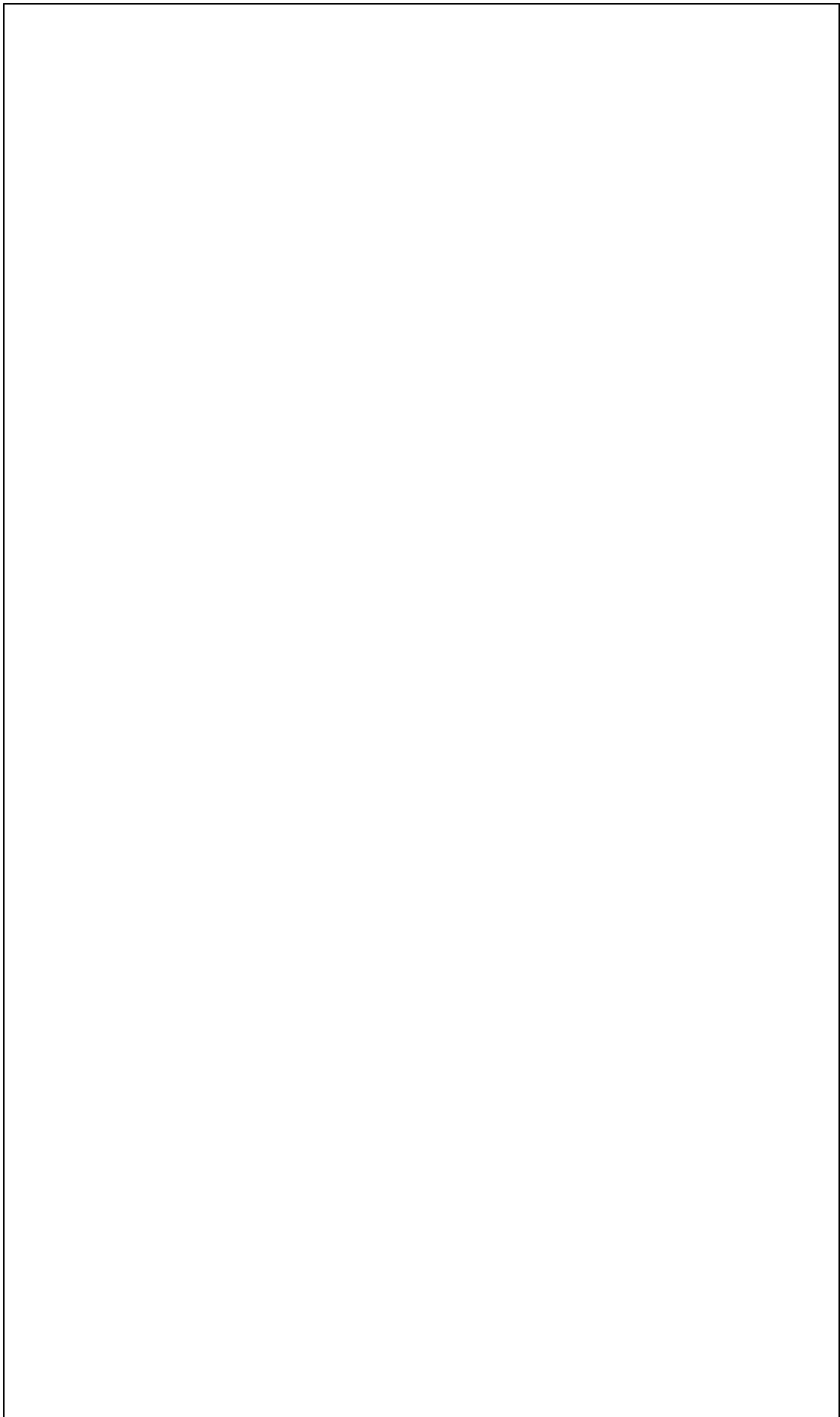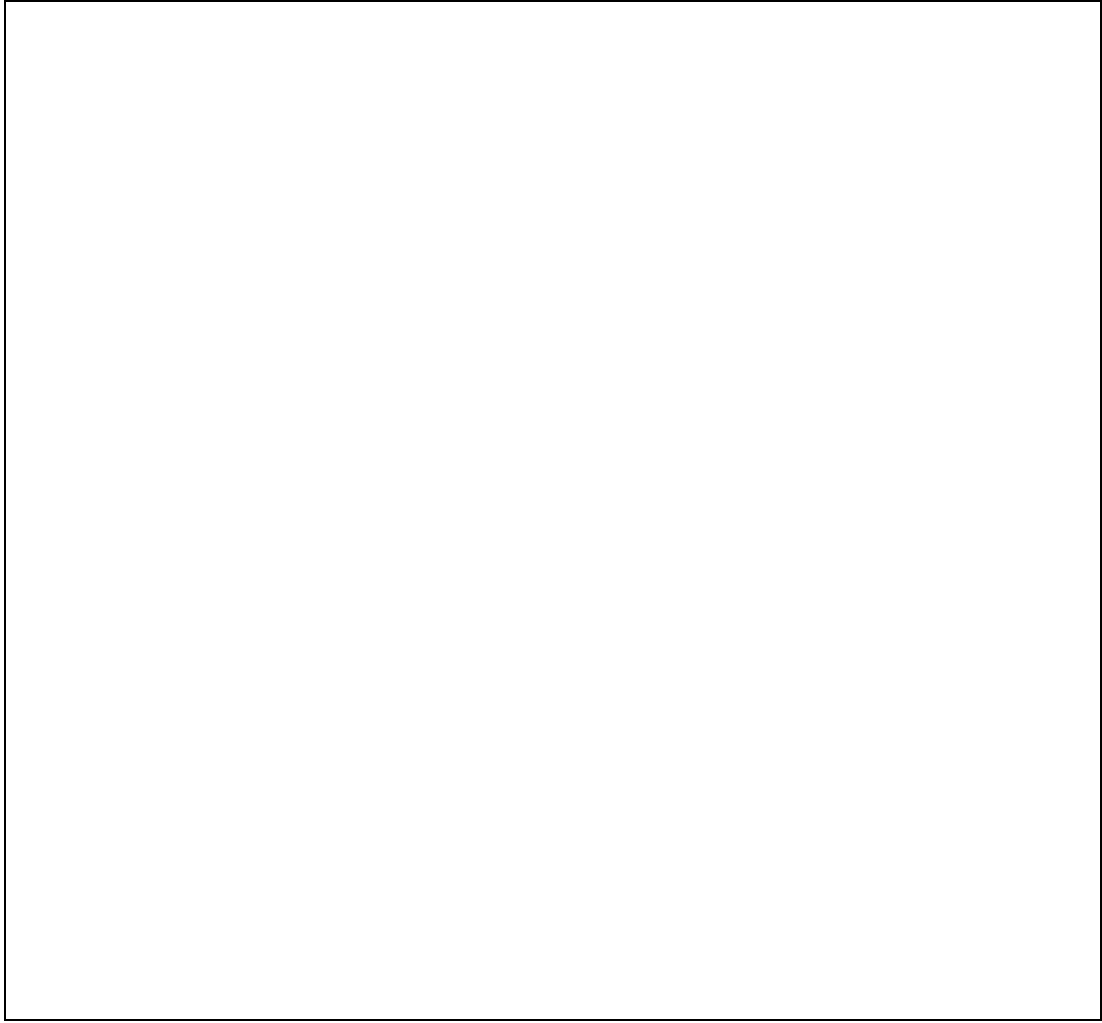
++

其他（例如感想、建议等等）。

1.所给的 txt 文件中，部分数据格式不对，缺少空字符，例如 java 检测到下列异常：

这是在统计平均分的时候抛出的异常，原因是 5（分数）后面没有用空字符隔开

深圳大学学生实验报告用纸

指导教师批阅意见：

成绩评定：

指导教师签字：

2019 年　　　月　　　日

备注：

注：1、报告内的项目或内容设置，可根据实际情况加以调整和补充。

2、教师批改学生实验报告时间应在学生提交实验报告时间后 10 日内。