# Java 大作业（数据&具体要求）

## 数据情况

假定用以下指标体系来衡量国家的科研能力：

|  |  |  |
| --- | --- | --- |
| **一级指标** | **二级指标** | **二级指标编号** |
| **研究情况** | **信息化顶级学者人口数量** | **D-111** |
| **信息化高被引文产出量** | **D-112** |
| **应用情况** | **信息化专利申请量** | **D-121** |
| **信息化专利授权量** | **D-122** |

科研能力通过“研究情况”和“应用情况”两个一级指标来体现，这两个一级指标各包含两个二级指标：“研究情况”一级指标包括“信息化顶级学者人口数量”和“信息化高被引文产出量”两个二级指标；“应用情况”一级指标包括“信息化专利申请量”和“信息化专利授权量”两个二级指标。

现给定数据“信息化高被引论文.zip”和“专利数据.xls”：

（1）“信息化高被引论文.zip”包含了在2019、2020、2021每年的某一领域的高被引论文，数据分布在6个TXT文件中，每个文件中数据格式为制表分隔符（即，使用”\t”分隔的数据）。PS.在计算时需要将6个文件一起统计

(2)“专利数据.xls”包含了2019、2020、2021每年的某一领域的专利。

## 具体要求

#### 1.需要根据上述给出的数据（即，“信息化高被引论文.zip”和“专利数据.xls”）分别统计和计算上表中的二级指标和一级指标，统计和计算的方法如下：

1. **D-111**

五个国家（即，中、英、法、美、俄）在2019、2020、2021每年的信息化顶级学者人口数量。请按照某一个国家信息化高被引论文中发表至少两篇论文的学者数量进行统计。在统计时请从文件的“RP”字段中的地址中提取国家信息。为了规范统计，请按照如下规则对国家进行统计：

若该字段的第一个通讯作者地址部分中若包含“China”（中国）、“USA”（美国）、“England”（英国）、“France”（法国）、“Russia”（俄罗斯），则该作者记为对应国家。

例如，对于：

“Xu, HY; Yang, HJ (通讯作者)，China Acad Chinese Med Sci, Inst Chinese Mat Med, Beijing 100700, Peoples R China.; Xu, HY; Huang, LQ (通讯作者)，China Acad Chinese Med Sci, Natl Resource Ctr Chinese Materia Med, Beijing 100700, Peoples R China.; Wang, XJ (通讯作者)，Chinese Acad Sci, Inst Genet & Dev Biol, Key Lab Genet Networks, Beijing 100101, Peoples R China.”

这条记录按照第一个通讯作者“Xu, HY”的地址进行统计，由于““Xu, HY”与“Yang, HJ”同为一个单位，因此这里“Xu, HY”的地址部分为“China Acad Chinese Med Sci, Inst Chinese Mat Med, Beijing 100700, Peoples R China”，由于其中包含“China”，因此可以判断“Xu, HY”来自中国，该作者计数一次。

1. **D-112**

五个国家（即，中、英、法、美、俄）在2019、2020、2021每年的信息化高被引文产出量。请从“信息化高被引论文.zip”文件的“RP”字段中的地址中提取国家信息。为了规范统计，请按照如下规则对国家进行统计：

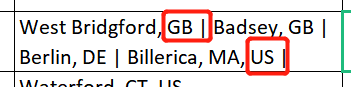
“RP”字段的第一个通讯作者地址部分中若包含“China”（中国）、“USA”（美国）、“England”（英国）、“France”（法国）、“Russia”（俄罗斯），则该作者记为对应国家。具体案例可参考①的举例。

1. **D-121**

五个国家（即，中、英、法、美、俄）在2019、2020、2021每年的信息化专利申请量**。**具体统计从“专利数据.xls”文件的“发明人地址”字段中的地址中提取国家信息，从“申请日”中提取年份信息（如果为空则不统计）。为了规范统计，请按照如下规则对国家进行统计：

该字段的地址部分中若包含“CN”（中国）、“US”（美国）、“GB”（英国）、“FR”（法国）、“RU”（俄罗斯），则该专利属于该国。**注意，一个专利可以同时属于多个国家，也就是说，如果地址中，同时包含“CN”和“US”，则同时增加中国与美国的专利计数+1。**

例如：



注意上图中，若地址同时包含了“GB”与“US”，则同时增加英国与美国的专利计数+1。

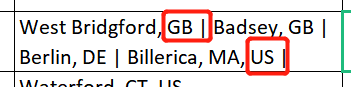
1. **D-122**

五个国家（即，中、英、法、美、俄）在2019、2020、2021每年的信息化专利授权量

具体统计从“专利数据.xls”文件的“发明人地址”字段中的地址中提取国家信息，从“授权日”中提取年份信息（如果为空则不统计）。为了规范统计，请按照如下规则对国家进行统计：

该字段的地址部分中若包含“CN”（中国）、“US”（美国）、“GB”（英国）、“FR”（法国）、“RU”（俄罗斯），则该专利属于该国。**注意，一个专利可以同时属于多个国家，也就是说，如果地址中，同时包含“CN”和“US”，则同时增加中国与美国的专利计数+1。**

例如：



上图中，地址同时包含了“GB”与“US”，则同时增加英国与美国的专利计数+1。

**提示：上述需要进行excel读取的java demo代码和jar包请见“ExcelDemo.zip”文件**

#### 2.上述①-④指标统计完后，需要对其进行MIN-MAX归一化，以获得上述指标的具体得分，该得分即为这些指标的数值，归一化的规则为：



这里计算完归一化的二级指标后，“理论情况”和“应用情况”两个一级指标的计算方式如下所示：

#### 3.自己构建构建数据库将计算完后的数据后的数据进行存储，并在文档中给出数据库设计（ER图、表结构、索引等）。

#### 4.要求使用Java Web实现一个能够可视化这些数据的Web系统（包含数据库、服务端、前端），要求数据能从3中设计的数据库中读出并在服务端处理后放入前端显示，包括并不限于所有国家指标的展示，以及这些指标在2019、2020、2021三年的横向对比。

#### 5.要求Web系统能够实现指标的增加、删除、修改（包括指标本身和指标数值的增删改操作）。