



曲阜师范大学

# 学生实验报告

学院 计算机学院

课程 人工智能实训

姓名 彭清元

学号 2020416240

年级 2020 级

专业 计算机科学与技术

# 曲阜师范大学实验报告

学院 计算机学院 年级、专业 20 级计算机科学与技术二班 学号 2020416240

姓名 彭清元 课程名称 人工智能 上课时间 第 1 节 — 第 11 节

实验日期 2023 年 5 月 27 日 星期 六 教师签名 雷玉霞 成绩

## 实验二 知识识别与合并

### 1. 【实训目的】

- (1) 加深对专家系统知识库的建立过程的理解。
- (2) 进一步了解知识的识别，分类及其合并的整体流程。
- (3) 简单实现自然语言到知识的分类处理过程

### 2. 【实验设备】

- (1) 一台装有 windows10 系统的笔记本电脑。
- (2) 集成 JDK 1.8.0\_151 编译环境
- (3) 使用 IntelliJ IDEA 进行编写程序

### 3. 【实验成员】

组长：彭清元

成员：彭清元

### 4. 【实验原理】

#### (1) 获取网上知识

本次数据是不同表述的文本，这些文本数据可以是关于学校、电影、书籍、音乐之类数据对象，同时获取对同一个对象不同描述的多个样本进行汇总，作为一个整体的数据集

#### (2) 对知识进行分析、识别

处理同一对象不同描述或者不同属性的时候，在对象中提取出该对象的属性、标志性词语，将这些对象分别进行识别出来，形成一段标准的文字段。再分析识别的过程中，将专有名词比如：“学校名称”、“年龄”、“创建时间”等进行分析识别，最后加以分割，对带有强烈的主谓宾关系的表述，形如“...的...是”这类语言，根据句子中的关键词进行分割，最终达到知识的分析和识别效果

#### (3) 对分析完成的知识进行合并

将多个样本的同一对象中，相同的属性或者相关的描述进行分析和识别完毕之后，接下来就是需要完成知识合并操作，将相同属性或者相关性较大的数据进行合并操作，最终呈现出来一个对象的同一个属性有多个值，进一步完善了这个对象的完整性

#### (4) 使用可视化界面呈现

使用可视化界面，从知识获取到知识的分析识别，最终到知识的合并，整个流程以可视化的形式体现，通过点击按钮可以获得不同的文件内容，对文本内容依次进行数据分析，将不同的文本内容分析的数据结果再返回到页面上展示，最后将各自分析结果进行合并汇总

## 5. 【设计思路】

### (1) 获取知识

这里选取的样本是电视剧《芸汐传》的相关信息，首先在网上获取关于《芸汐传》的多个样本相关信息，将这些信息处理成 json 格式数据，将这些样本形成一个小的数据集，为后续实现知识识别和分析提供了数据样本。

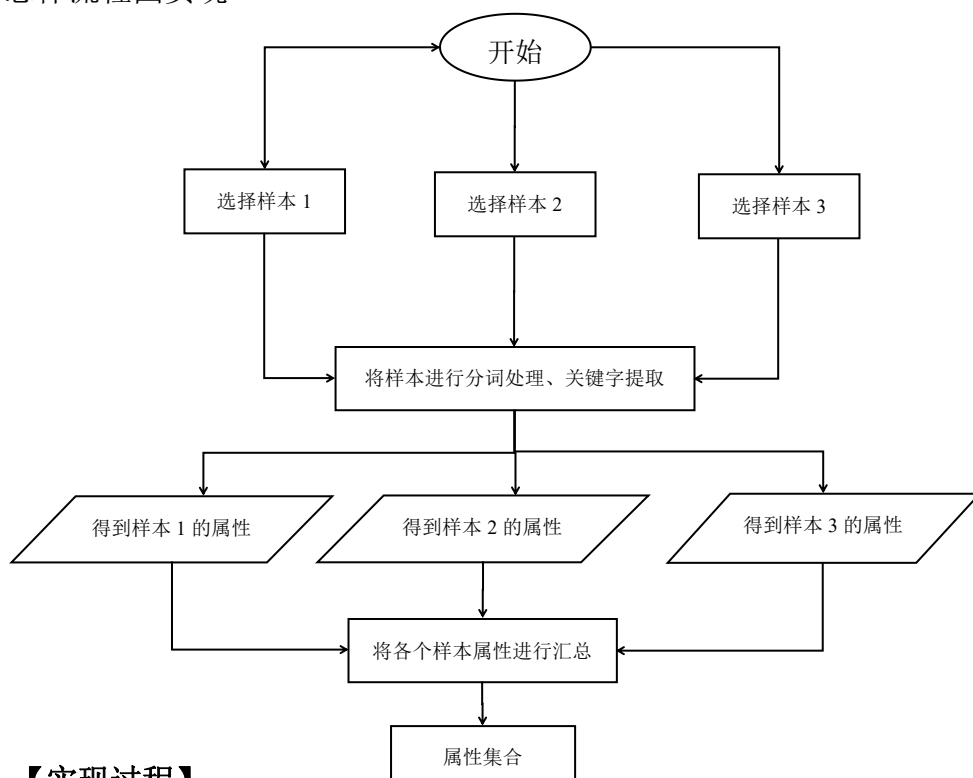
### (2) 分析数据集

利用上述获取到的数据集，从其中识别出，导演、主演、发布时间、名称、编剧、类型、监制、别名、阅读量、讨论量、简介、等关键字，将这些关键字作为识别标志符号，进行分割处理，在分词过程中使用 split 对文本进行分词处理，对每一句话使用 substring 进行检索

### (3) 可视化设计

使用 HTML 设计可视化界面，将数据进行双向绑定，使用 axios 获取样本数据，将样本数据传送给 JavaScript 处理，将处理完毕的数据再返回给 HTML，进行渲染，当所有样本处理完成之后，最后将分析的几个样本进行合并

总体流程图实现



## 6. 【实现过程】

### (1) 实现数据获取

将实验数据存入样本 1.md，样本 2.md，样本 3.md，样本 4.md 文件进行保存，通过读取文件进行数据获取，将获取的数据转成 Json 格式，存入到 data.json 文件中去。

样本 1 的数据如下所示

《芸汐传》是由林健龙 刘镇明 导演的一部国产剧，芸汐传在大陆首映，并在2018年第一时间上映，主要演员有鞠婧祎 张哲瀚 米热 林思意 王佑硕 卢星宇 许佳琪 等，樱花动漫为大家提供芸汐传全集完整版免费在线观看，支持手机、平板、电脑多终端高清免费播放，芸汐传现更新至已完结。

将文本数据读取出来之后转成 json 格式的数据，使用 Node 库中 fs 文件，将文件从样本中获取出来，转成 json 格式之后，在依次存入到 yangben 对象中，最后使用 fs 将转化后的数据存入文件中去，具体实现代码：

```
/**
 * 将 样本从文本中读取出来 转化成 Json 格式 存入 Json 文件中
 * @type {}{}
 */

var yangben = {}
index = {
  "1": "one",
  "2": "two",
  "3": "three",
  "4": "four",
}

const fs = require("fs")
const {writeFile} = require("fs");

function WriteData(number, path) {
  fs.readFile(path, 'utf8', (err, data) => {
    if (err) console.log('读取失败')
    else {
      yangben[index[number]] = {
        message: data
      }

      let len = Object.keys(yangben).length
      if (len == 4) {
        // console.log(yangben)
        const palce = './config.json';
        writeFile(palce, JSON.stringify(yangben, null), (error) => {
          if (error) {
            console.log('An error has occurred ', error);
            return;
          }
          console.log('Data written successfully to disk');
        });
      }
    }
  });
}
```

```

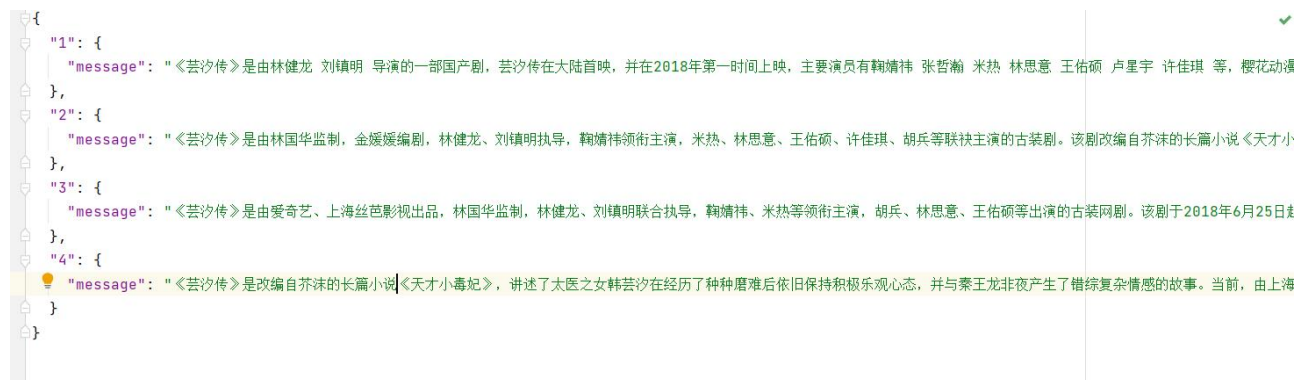
    }
  })
}

function SendData() {
  for (let i = 1; i <= 4; i++) {
    WriteData(i, `样本${i}.md`)
  }
}

SendData()

```

将数据转成 Json 格式，转化成功之后的数据显示



```

{
  "1": {
    "message": "《芸汐传》是由林健龙、刘镇明 导演的一部国产剧，芸汐传在大陆首映，并在2018年第一时间上映，主要演员有鞠婧祎、张哲瀚、米热、林思意、王佑硕、卢星宇、许佳琪 等，樱花动漫"
  },
  "2": {
    "message": "《芸汐传》是由林国华监制，金媛媛编剧，林健龙、刘镇明执导，鞠婧祎领衔主演，米热、林思意、王佑硕、许佳琪、胡兵等联袂主演的古装剧。该剧改编自芥沫的长篇小说《天才小毒妃》"
  },
  "3": {
    "message": "《芸汐传》是由爱奇艺、上海丝芭影视出品，林国华监制，林健龙、刘镇明联合执导，鞠婧祎、米热等领衔主演，胡兵、林思意、王佑硕等出演的古装网剧。该剧于2018年6月25日"
  },
  "4": {
    "message": "《芸汐传》是改编自芥沫的长篇小说《天才小毒妃》，讲述了太医之女韩芸汐在经历了种种磨难后依旧保持积极乐观心态，并与秦王龙非夜产生了错综复杂情感的故事。当前，由上海"
  }
}

```

## (2) 获取关键字

将我们需要得到的属性，从文本中获取出来，首先将每一类属性的相近词、同义词进行汇总，再将每一个属性的关键字进行分割处理，存储起来，为后续判断提供条件。部分代码展示

```

Type = data.data.Types.split(" ")
Director = data.data.Director.split(" ")
Starring = data.data.Starring.split(" ")
Release = data.data.Release.split(" ")
screenwriter = data.data.screenwriter.split(" ")
producer = data.data.producer.split(" ")
alias = data.data.alias.split(" ")
reads = data.data.reads.split(" ")
discovers = data.data.discovers.split(" ")
Info = data.data.Info.split(" ")

```

## (3) 获取关键字属性

将 Json 中的样本数据提取出来之后，将整个 Json 数据作为一个字符串处理，将字符串按照 “、”、“ ”、“。” 使用正则表达式进行分割，将一个文本数据分割形成若干个子句，原本数据显示

《芸汐传》是由爱奇艺、上海丝芭影视出品，林国华监制，林健龙、刘镇明联合执导，鞠婧祎、米热等领衔主演，胡兵、林思意、王佑硕等出演的古装网剧。该剧于 2018 年 6 月 25 日起在爱奇艺播出，并于 2020 年 2 月 12 日起在湖南卫视播出。该剧改编自作者芥沫的小说《天

才小毒妃》。《芸汐传》截至收官，爱奇艺播放量突破 30 亿，微博主话题阅读量 10 亿，讨论量达 226 万。2019 年，该剧获得 2019 微博电视剧大赏人气剧集第 4 名。

将文本处理之后的数据显示

```
0: "《芸汐传》是由爱奇艺、上海丝芭影视出品"
1: "林国华监制"
2: "林健龙、刘镇明联合执导"
3: "鞠婧祎、米热等领衔主演"
4: "胡兵、林思意、王佑硕等出演的古装网剧"
5: "该剧于2018年6月25日起在爱奇艺播出"
6: "并于2020年2月12日起在湖南卫视播出"
7: "该剧改编自作者芥沫的小说《天才小毒妃》"
8: "《芸汐传》截至收官"
9: "爱奇艺播放量突破30亿"
10: "微博主话题阅读量10亿"
11: "讨论量达226万"
12: "2019年"
13: "该剧获得2019微博电视剧大赏人气剧集第4名"
14: ""
```

以上述数据为例，具体说明获取整个样本所有属性的值的过程，比如获取名称整个属性，常常是伴随着“该作品”、“《》”等标志信息，获取时间属性，常常使用“于...日”进行搭配，因此根据这些关键字进行分割获取对应的值，使用双层循环对文本分割后所有的子句进行遍历，访问所有属性的关键字进行一一匹配，将获取的值以字典的形式存入到对象中，返回回去

```
let obj = {
  导演: Dir == null ? "" : Dir,
  主演: Sta == null ? "" : Sta,
  发布时间: Rel == null ? "" : Rel,
  名称: nam == null ? "" : nam, // "编剧": scr == null ? "" : scr,
  类型: Tys == null ? "" : Tys,
  监制: pro == null ? "" : pro,
  别名: al == null ? "" : al,
  阅读量: rea == null ? "" : rea,
  讨论量: dis == null ? "" : dis
}
```

将属性数据存入对象之后显示

```
▼ Object
  ▶ 主演: Array(2)
  ▶ 别名: Array(1)
  ▶ 发布时间: Array(2)
  ▶ 名称: Array(1)
  ▶ 导演: Array(1)
  ▶ 监制: Array(1)
  ▶ 类型: Array(1)
  ▶ 讨论量: Array(1)
  ▶ 阅读量: Array(1)
```

然后将同样的处理方法进行处理另外一个文本，再不做优化的情况下可以发现有一些数据是会存在缺失或者多余的现象，这是因为文本对象的属性不足导致的

```
{
  "导演": [
    [
      "林健龙",
      "刘镇明联合"
    ]
  ]
}
```

```

],
"主演": [
    "鞠婧祎、米热等"
],
"发布时间": [
    "2018 年 6 月 25 日",
    "2020 年 2 月 12 日"
],
"名称": [
    "《芸汐传》"
],
"编剧": [],
"类型": [
    "古装"
],
"监制": [
    "国华"
],
"别名": [
    "《天才小毒妃》"
]
}

```

因此，对属性集进一步优化，在属性集中加入相关性较大或者表述相近的词语，对关键字的检索添加多个判断，保证数据准确性

#### (4) 将多文本属性合并

使用 `for in` 语句对每一个属性的值进行遍历，将四个文本中的属性的值先进行解构，解构完成之后使用 `push` 函数加入到对应的属性当中

```

for (it in string2) {

    if (it == item) {

        obj[item].push(...string2[it])

    }

}

```

考虑到 `obj[item]` 中存在重复数据，因此使用 `forEach` 语句对合并完成之后的数据进行过滤

```

arr = {};

Sta.forEach(item => arr[item] = item);

Sta = Object.values(arr);

```

#### (5) 可视化设计

一开始点击按钮之后使用 axios 将数据从 Json 文件中调出来,使用 v-model 指令将数据进行绑定,将传过来的 Json 字符串传入 *Process1* 函数进行处理,将处理完成的数据返回,使用插值表达式进行渲染展示

```
<div>

    <button @click="btn1">导入文本 1</button>

    <div>

        <textarea>{{msg1}}</textarea>

    </div>

</div>
```

#### 可视化代码展示

```
<!DOCTYPE html>
<html lang="en">
  <head>
    <meta charset="UTF-8">
    <title>知识获取和合并</title>
  </head>

  <style>
    header {
      width: 1200px;
      margin: 10px auto;
      color: darkorange;
      text-align: center;
    }

    .container {
      width: 1200px;
      height: 100px;
      margin: 10px auto;
      display: flex;
      flex-direction: row;
      /*background-color: red;*/
      justify-content: space-around;
      /*padding: 10px 10px;*/
    }
  </style>
</html>
```



```
.left {  
  flex: 5;  
  display: flex;  
  justify-content: space-around;  
  flex-wrap: wrap;  
}
```

```
.right {  
  flex: 5;  
  display: flex;  
  justify-content: space-around;  
  flex-wrap: wrap;  
}
```

```
.left div {  
  /*width: 220px;*/  
}
```

```
.right {  
  flex: 2;  
}
```

```
.left textarea {  
  width: 200px;  
  height: 250px;  
  outline: none;  
  resize: none  
}
```

```
.right textarea {  
  width: 200px;  
  height: 500px;  
  outline: none;  
  resize: none  
}
```

```
button {  
  margin-top: 20px;  
  margin-bottom: 20px;  
}
```

```
</style>
```

```
<body>
```

```
<div id="App">
```

```
<header><h1>欢迎使用由 20 级计科 2 班彭清元设计的知识获取和合并的可视化  
展示</h1></header>
```

```
<div class="container">
```

```
<div class="left">
  <div>
    <button @click="btn1">导入文本 1</button>
    <div>
      <textarea>{{msg1}}</textarea>
    </div>
  </div>
  <div>
    <button @click="btn2">导入文本 2</button>
    <div>
      <textarea>{{msg2}}</textarea>
    </div>
  </div>
  <div>
    <button @click="btn3">导入文本 3</button>
    <div>
      <textarea>{{msg3}}</textarea>
    </div>
  </div>
  <div>
    <button @click="btn4">导入文本 4</button>
    <div>
      <textarea>{{msg4}}</textarea>
    </div>
  </div>
  <div>
    <button @click="btn5">文本 1 信息</button>
    <div>
      <textarea>{{msg5}}</textarea>
    </div>
  </div>
  <div>
    <button @click="btn6">文本 2 信息</button>
    <div>
      <textarea>{{msg6}}</textarea>
    </div>
  </div>
  <div>
    <button @click="btn7">文本 3 信息</button>
    <div>
      <textarea>{{msg7}}</textarea>
    </div>
  </div>
  <div>
    <button @click="btn8">文本 4 信息</button>
    <div>
```

```

        <textarea>{msg8}</textarea>
    </div>
</div>

</div>
<div class="right">
    <div>
        <button @click="btn9">合并结果</button>
    </div>
    <textarea>{msg9}</textarea>
</div>
</div>
</div>
</div>
</body>
</html>

```

## (6) 可视化界面展示

将页面分为数据导入、数据信息分析和数据合并三大块，整体布局如下图所示

欢迎使用由20级计科2班彭清元设计的知识获取和合并的可视化展示

导入文本1

《芸汐传》是由林健龙、刘镇明导演的一部国产剧。芸汐传在大陆首映，并在2018年第一时间上映，主要演员有鞠婧祎、张哲瀚、米热、林思意、王佑硕、卢星宇、许佳琪等。樱花动漫为大家提供芸汐传全集完整版免费在线观看，支持手机、平板、电脑多终端高清免费播放，芸汐传现更新至已完结。

文本1信息

```

{
  "导演": [
    "林健龙",
    "刘镇明"
  ],
  "主演": [
    "鞠婧祎",
    "张哲瀚",
    "米热",
    "林思意",
    "王佑硕",
    "卢星宇",
    "许佳琪"
  ],
  "发布时间": [
    "2018年"
  ]
}

```

导入文本2

《芸汐传》是由林国华监制，金媛媛编剧，林健龙、刘镇明执导，鞠婧祎领衔主演，米热、林思意、王佑硕、许佳琪、胡兵等联袂主演的古装剧。该剧改编自芥沫的长篇小说《天才小毒妃》。该剧于2018年6月25日在爱奇艺播出，2020年2月12日在湖南卫视上星播出，2020年7月24日《芸汐传》再次在湖南卫视播出。

文本2信息

```

{
  "导演": [
    "林健龙",
    "刘镇明"
  ],
  "主演": [
    "鞠婧祎",
    "米热",
    "林思意",
    "王佑硕",
    "许佳琪",
    "胡兵"
  ],
  "发布时间": [
    "2018年6月25日",
    "2020年2月12日"
  ]
}

```

导入文本3

《芸汐传》是由爱奇艺、上海丝芭影视出品，林国华监制，林健龙、刘镇明联合执导，鞠婧祎、米热等领衔主演，胡兵、林思意、王佑硕等出演的古装网剧。该剧于2018年6月25日起在爱奇艺播出，并于2020年2月12日起在湖南卫视播出。该剧改编自作者芥沫的小说《天才小毒妃》。《芸汐传》截至收官，爱奇艺播放量突破30亿，微博话题阅读量10亿，讨论量达226万。2019年，该剧获得2019微博电视剧大赏人气剧集第4名。

文本3信息

```

{
  "导演": [
    "林健龙",
    "刘镇明联合"
  ],
  "主演": [
    "鞠婧祎",
    "米热等"
  ],
  "发布时间": [
    "2018年6月25日",
    "2020年2月12日"
  ],
  "名称": [
    "《芸汐传》"
  ],
  "别名": [
    "《天才小毒妃》"
  ]
}

```

导入文本4

《芸汐传》是改编自芥沫的长篇小说《天才小毒妃》，讲述了太医之女韩芸汐在经历了种种磨难后依旧保持积极乐观心态，并与秦王龙非夜产生了错综复杂情感的故事。当前，由上海丝芭影视出品的电视剧《芸汐传》，正在横店影视城秦王宫、春秋(唐)园及外景拍摄基地有条不紊取景拍摄之中。

文本4信息

```

{
  "名称": [
    "《芸汐传》"
  ],
  "别名": [
    "《天才小毒妃》"
  ],
  "简介": [
    "太医之女韩芸汐在经历了种种磨难后依旧保持积极乐观心态，并与秦王龙非夜产生了错综复杂情感的故事"
  ]
}

```

合并结果

```

{
  "导演": [
    "林健龙",
    "刘镇明",
    "林健龙",
    "刘镇明联合"
  ],
  "主演": [
    "鞠婧祎",
    "张哲瀚",
    "米热",
    "林思意",
    "王佑硕",
    "卢星宇",
    "许佳琪",
    "胡兵",
    "鞠婧祎",
    "米热等"
  ],
  "发布时间": [
    "2018年",
    "2018年6月25日",
    "2020年2月12日",
    "2020年7月24日"
  ],
  "名称": [
    "《芸汐传》"
  ],
  "别名": [
    "《天才小毒妃》"
  ],
  "编剧": [
    "金媛媛"
  ],
  "类型": [
    "古装"
  ],
  "监制": [
    "林国华"
  ]
}

```

## 7. 实验结果

通过点击导入文本1的按钮，文本1信息从json文件中导入进来

欢迎使用由20级计科2班彭清元设计的知识获取和合并的可视化展示

导入文本1

《芸汐传》是由林健龙、刘镇明导演的一部国产剧，芸汐传在大陆首映，并在2018年第一时间上映，主要演员有鞠婧祎、张哲瀚、米热、林思意、王佑硕、卢星宇、许佳琪等。樱花动漫为大家提供芸汐传全集完整版免费在线观看，支持手机、平板、电脑多终端高清免费播放，芸汐传更新至已完结。

导入文本2

导入文本3

导入文本4

合并结果

文本1信息

文本2信息

文本3信息

文本4信息

点击文本 1 信息就可以获取当前文本 1，通过文本处理算法处理好了之后返回的信息

欢迎使用由20级计科2班彭清元

导入文本1

《芸汐传》是由林健龙、刘镇明导演的一部国产剧，芸汐传在大陆首映，并在2018年第一时间上映，主要演员有鞠婧祎、张哲瀚、米热、林思意、王佑硕、卢星宇、许佳琪等。樱花动漫为大家提供芸汐传全集完整版免费在线观看，支持手机、平板、电脑多终端高清免费播放，芸汐传更新至已完结。

导入文本2

导入文本3

文本1信息

```
[
  "导演": [
    [
      "林健龙",
      "刘镇明"
    ]
  ],
  "主演": [
    "鞠婧祎",
    "张哲瀚",
    "米热",
    "林思意",
    "王佑硕",
    "卢星宇",
    "许佳琪"
  ],
  "发布时间": [
    "2018年"
  ]
]
```

文本2信息

文本3信息

通过点击导入文本 2 的按钮，文本 2 信息从 json 文件中导入进来

导入文本1	导入文本2	导入文本3	导入文本4
<p>《芸汐传》是由林健龙、刘镇明导演的一部国产剧，芸汐传在大陆首映，并在2018年第一时间上映，主要演员有鞠婧祎、张哲瀚、米热、林思意、王佑硕、卢星宇、许佳琪等，樱花动漫为大家提供芸汐传全集完整版免费在线观看，支持手机、平板、电脑多终端高清免费播放，芸汐传现更新至已完结。</p>	<p>《芸汐传》是由林国华监制，金媛媛编剧，林健龙、刘镇明执导，鞠婧祎领衔主演，米热、林思意、王佑硕、许佳琪、胡兵等联袂主演的古装剧。该剧改编自芥沫的长篇小说《天才小毒妃》。该剧于2018年6月25日在爱奇艺播出，2020年2月12日在湖南卫视上星播出，2020年7月24日《芸汐传》再次在湖南卫视播出。</p>		
文本1信息	文本2信息	文本3信息	文本4信息
<pre>{   "导演": [     "林健龙",     "刘镇明"   ],   "主演": [     "鞠婧祎",     "张哲瀚",     "米热",     "林思意",     "王佑硕",     "卢星宇",     "许佳琪"   ],   "发布时间": [     "2018年"</pre>			

点击文本 2 信息就可以获取当前文本 2，通过文本处理算法处理好了之后返回的信息

导入文本1	导入文本2	导入文本3	导入文本4
<p>《芸汐传》是由林健龙、刘镇明导演的一部国产剧，芸汐传在大陆首映，并在2018年第一时间上映，主要演员有鞠婧祎、张哲瀚、米热、林思意、王佑硕、卢星宇、许佳琪等，樱花动漫为大家提供芸汐传全集完整版免费在线观看，支持手机、平板、电脑多终端高清免费播放，芸汐传现更新至已完结。</p>	<p>《芸汐传》是由林国华监制，金媛媛编剧，林健龙、刘镇明执导，鞠婧祎领衔主演，米热、林思意、王佑硕、许佳琪、胡兵等联袂主演的古装剧。该剧改编自芥沫的长篇小说《天才小毒妃》。该剧于2018年6月25日在爱奇艺播出，2020年2月12日在湖南卫视上星播出，2020年7月24日《芸汐传》再次在湖南卫视播出。</p>		
文本1信息	文本2信息	文本3信息	文本4信息
<pre>{   "导演": [     "林健龙",     "刘镇明"   ],   "主演": [     "鞠婧祎",     "张哲瀚",     "米热",     "林思意",     "王佑硕",     "卢星宇",     "许佳琪"   ],   "发布时间": [     "2018年"</pre>	<pre>{   "导演": [     "林健龙",     "刘镇明"   ],   "主演": [     "鞠婧祎",     "米热",     "林思意",     "王佑硕",     "许佳琪",     "胡兵"   ],   "发布时间": [     "2018年6月25日",     "2020年2月12日"</pre>		

通过点击导入文本 3 的按钮，文本 3 信息从 json 文件中导入进来

导入文本1

《芸汐传》是由林健龙、刘镇明导演的一部国产剧，芸汐传在大陆首映，并在2018年第一时间上映，主要演员有鞠婧祎、张哲瀚、米热、林思意、王佑硕、卢星宇、许佳琪等，樱花动漫为大家提供芸汐传全集完整版免费在线观看，支持手机、平板、电脑多终端高清免费播放，芸汐传现更新至已完结。

导入文本2

《芸汐传》是由林国华监制，金媛媛编剧，林健龙、刘镇明执导，鞠婧祎领衔主演，米热、林思意、王佑硕、许佳琪、胡兵等联袂主演的古装剧。该剧改编自芥沫的长篇小说《天才小毒妃》。该剧于2018年6月25日在爱奇艺播出，2020年2月12日在湖南卫视上星播出，2020年7月24日《芸汐传》再次在湖南卫视播出。

导入文本3

《芸汐传》是由爱奇艺、上海丝芭影视出品，林国华监制，林健龙、刘镇明联合执导，鞠婧祎、米热等领衔主演，胡兵、林思意、王佑硕等出演的古装网剧。该剧于2018年6月25日起在爱奇艺播出，并于2020年2月12日起在湖南卫视播出。该剧改编自作者芥沫的小说《天才小毒妃》。《芸汐传》截至收官，爱奇艺播放量突破30亿，微博主话题阅读量10亿，讨论量达226万。2019年，该剧获得2019微博电视剧大赏人气剧集第4名。

导入文本4

文本1信息

```
{
  "导演": [
    "林健龙",
    "刘镇明"
  ],
  "主演": [
    "鞠婧祎",
    "张哲瀚",
    "米热",
    "林思意",
    "王佑硕",
    "卢星宇",
    "许佳琪"
  ],
  "发布时间": [
    "2018年"
  ]
}
```

文本2信息

```
{
  "导演": [
    "林健龙",
    "刘镇明"
  ],
  "主演": [
    "鞠婧祎",
    "米热",
    "林思意",
    "王佑硕",
    "许佳琪",
    "胡兵"
  ],
  "发布时间": [
    "2018年6月25日",
    "2020年2月12日"
  ]
}
```

文本3信息

文本4信息

点击文本 3 信息就可以获取当前文本 3，通过文本处理算法处理好了之后返回的信息

导入文本1

《芸汐传》是由林健龙、刘镇明导演的一部国产剧，芸汐传在大陆首映，并在2018年第一时间上映，主要演员有鞠婧祎、张哲瀚、米热、林思意、王佑硕、卢星宇、许佳琪等，樱花动漫为大家提供芸汐传全集完整版免费在线观看，支持手机、平板、电脑多终端高清免费播放，芸汐传现更新至已完结。

导入文本2

《芸汐传》是由林国华监制，金媛媛编剧，林健龙、刘镇明执导，鞠婧祎领衔主演，米热、林思意、王佑硕、许佳琪、胡兵等联袂主演的古装剧。该剧改编自芥沫的长篇小说《天才小毒妃》。该剧于2018年6月25日在爱奇艺播出，2020年2月12日在湖南卫视上星播出，2020年7月24日《芸汐传》再次在湖南卫视播出。

导入文本3

《芸汐传》是由爱奇艺、上海丝芭影视出品，林国华监制，林健龙、刘镇明联合执导，鞠婧祎、米热等领衔主演，胡兵、林思意、王佑硕等出演的古装网剧。该剧于2018年6月25日起在爱奇艺播出，并于2020年2月12日起在湖南卫视播出。该剧改编自作者芥沫的小说《天才小毒妃》。《芸汐传》截至收官，爱奇艺播放量突破30亿，微博主话题阅读量10亿，讨论量达226万。2019年，该剧获得2019微博电视剧大赏人气剧集第4名。

导入文本4

文本1信息

```
{
  "导演": [
    "林健龙",
    "刘镇明"
  ],
  "主演": [
    "鞠婧祎",
    "张哲瀚",
    "米热",
    "林思意",
    "王佑硕",
    "卢星宇",
    "许佳琪"
  ],
  "发布时间": [
    "2018年"
  ]
}
```

文本2信息

```
{
  "导演": [
    "林健龙",
    "刘镇明"
  ],
  "主演": [
    "鞠婧祎",
    "米热",
    "林思意",
    "王佑硕",
    "许佳琪",
    "胡兵"
  ],
  "发布时间": [
    "2018年6月25日",
    "2020年2月12日"
  ]
}
```

文本3信息

```
{
  "导演": [
    "林健龙",
    "刘镇明"
  ],
  "主演": [
    "鞠婧祎",
    "米热"
  ],
  "发布时间": [
    "2018年6月25日",
    "2020年2月12日"
  ],
  "名称": [
    "《芸汐传》"
  ]
}
```

文本4信息

通过点击导入文本 4 的按钮，文本 4 信息从 json 文件中导入进来



导入文本1

《芸汐传》是由林健龙、刘镇明导演的一部国产剧，芸汐传在大陆首映，并在2018年第一时间上映，主要演员有鞠婧祎、张哲瀚、米热、林思意、王佑硕、卢星宇、许佳琪等，樱花动漫为大家提供芸汐传全集完整版免费在线观看，支持手机、平板、电脑多终端高清免费播放，芸汐传现更新至已完结。

导入文本2

《芸汐传》是由林国华监制，金媛媛编剧，林健龙、刘镇明执导，鞠婧祎领衔主演，米热、林思意、王佑硕、许佳琪、胡兵等联袂主演的古装剧。该剧改编自芥沫的长篇小说《天才小毒妃》。该剧于2018年6月25日在爱奇艺播出，2020年2月12日在湖南卫视上星播出，2020年7月24日《芸汐传》再次在湖南卫视播出。

导入文本3

《芸汐传》是由爱奇艺、上海丝芭影视出品，林国华监制，林健龙、刘镇明联合执导，鞠婧祎、米热等领衔主演，胡兵、林思意、王佑硕等出演的古装网剧。该剧于2018年6月25日起在爱奇艺播出，并于2020年2月12日起在湖南卫视播出。该剧改编自作者芥沫的小说《天才小毒妃》。《芸汐传》截至收官，爱奇艺播放量突破30亿，微博话题阅读量10亿，讨论量达226万。2019年，该剧获得2019微博电视剧大赏人气剧集第4名。

导入文本4

《芸汐传》是改编自芥沫的长篇小说《天才小毒妃》，讲述了太医之女韩芸汐在经历了种种磨难后依旧保持积极乐观心态，并与秦王龙非夜产生了错综复杂情感的故事。当前，由上海丝芭影视出品的电视剧《芸汐传》，正在横店影视城秦王宫、春秋(唐)园及外景拍摄基地有条不紊取景拍摄之中。

文本1信息

```
{
  "导演": [
    "林健龙",
    "刘镇明"
  ],
  "主演": [
    "鞠婧祎",
    "张哲瀚",
    "米热",
    "林思意",
    "王佑硕",
    "卢星宇",
    "许佳琪"
  ],
  "发布时间": [
    "2018年"
```

文本2信息

```
{
  "导演": [
    "林健龙",
    "刘镇明"
  ],
  "主演": [
    "鞠婧祎",
    "米热",
    "林思意",
    "王佑硕",
    "许佳琪",
    "胡兵"
  ],
  "发布时间": [
    "2018年6月25日",
    "2020年2月12日"
```

文本3信息

```
{
  "导演": [
    "林健龙",
    "刘镇明"
  ],
  "主演": [
    "鞠婧祎",
    "米热"
  ],
  "发布时间": [
    "2018年6月25日",
    "2020年2月12日"
  ],
  "名称": [
    "芸汐传"
  ]
}
```

文本4信息

点击文本4信息就可以获取当前文本4，通过文本处理算法处理好了之后返回的信息

导入文本1

《芸汐传》是由林健龙、刘镇明导演的一部国产剧，芸汐传在大陆首映，并在2018年第一时间上映，主要演员有鞠婧祎、张哲瀚、米热、林思意、王佑硕、卢星宇、许佳琪等，樱花动漫为大家提供芸汐传全集完整版免费在线观看，支持手机、平板、电脑多终端高清免费播放，芸汐传现更新至已完结。

导入文本2

《芸汐传》是由林国华监制，金媛媛编剧，林健龙、刘镇明执导，鞠婧祎领衔主演，米热、林思意、王佑硕、许佳琪、胡兵等联袂主演的古装剧。该剧改编自芥沫的长篇小说《天才小毒妃》。该剧于2018年6月25日在爱奇艺播出，2020年2月12日在湖南卫视上星播出，2020年7月24日《芸汐传》再次在湖南卫视播出。

导入文本3

《芸汐传》是由爱奇艺、上海丝芭影视出品，林国华监制，林健龙、刘镇明联合执导，鞠婧祎、米热等领衔主演，胡兵、林思意、王佑硕等出演的古装网剧。该剧于2018年6月25日起在爱奇艺播出，并于2020年2月12日起在湖南卫视播出。该剧改编自作者芥沫的小说《天才小毒妃》。《芸汐传》截至收官，爱奇艺播放量突破30亿，微博话题阅读量10亿，讨论量达226万。2019年，该剧获得2019微博电视剧大赏人气剧集第4名。

导入文本4

《芸汐传》是改编自芥沫的长篇小说《天才小毒妃》，讲述了太医之女韩芸汐在经历了种种磨难后依旧保持积极乐观心态，并与秦王龙非夜产生了错综复杂情感的故事。当前，由上海丝芭影视出品的电视剧《芸汐传》，正在横店影视城秦王宫、春秋(唐)园及外景拍摄基地有条不紊取景拍摄之中。

文本1信息

```
{
  "导演": [
    "林健龙",
    "刘镇明"
  ],
  "主演": [
    "鞠婧祎",
    "张哲瀚",
    "米热",
    "林思意",
    "王佑硕",
    "卢星宇",
    "许佳琪"
  ],
  "发布时间": [
    "2018年"
```

文本2信息

```
{
  "导演": [
    "林健龙",
    "刘镇明"
  ],
  "主演": [
    "鞠婧祎",
    "米热",
    "林思意",
    "王佑硕",
    "许佳琪",
    "胡兵"
  ],
  "发布时间": [
    "2018年6月25日",
    "2020年2月12日"
```

文本3信息

```
{
  "导演": [
    "林健龙",
    "刘镇明"
  ],
  "主演": [
    "鞠婧祎",
    "米热"
  ],
  "发布时间": [
    "2018年6月25日",
    "2020年2月12日"
  ],
  "名称": [
    "芸汐传"
  ]
}
```

文本4信息

```
{
  "名称": [
    "芸汐传"
  ],
  "别名": [
    "天才小毒妃"
  ],
  "简介": [
    "太医之女韩芸汐在经历了种种磨难后依旧保持积极乐观心态，并与秦王龙非夜产生了错综复杂情感的故事"
  ]
}
```

点击合并按钮，后台调用 *merge* 方法，将四个处理好的文本信息进行合并，呈现至可视化界面

## 欢迎使用由20级计科2班彭清元设计的知识获取和合并的可视化展示

导入文本1

《芸汐传》是由林健龙、刘镇明导演的一部国产剧，芸汐传在大陆首映，并在2018年第一时间上映，主要演员有鞠婧祎、张哲瀚、米热、林思意、王佑硕、许佳琪、胡兵等。该剧改编自米热、林思意、王佑硕、许佳琪、胡兵等联袂主演的古装小说《天才小毒妃》。该剧于2018年6月25日起在爱奇艺播出，并于2020年2月12日起在湖南卫视播出。该剧改编自作者芥沫的小说《天才小毒妃》。《芸汐传》截至收官，爱奇艺播放量突破30亿，微博话题阅读量10亿，讨论量达226万。2019年，该剧获得2019微博电视剧大赏人气剧集第4名。

导入文本2

《芸汐传》是由林国华监制，金媛媛编剧，林健龙、刘镇明执导，鞠婧祎领衔主演，米热、林思意、王佑硕、许佳琪、胡兵等联袂主演的古装剧。该剧改编自芥沫的长篇小说《天才小毒妃》。该剧于2018年6月25日起在爱奇艺播出，2020年2月12日起在湖南卫视播出。2020年7月24日《芸汐传》再次在湖南卫视播出。

导入文本3

《芸汐传》是由爱奇艺、上海丝芭影视出品，林国华监制，林健龙、刘镇明联合执导，鞠婧祎、米热等领衔主演，胡兵、林思意、王佑硕等出演的古装网剧。该剧于2018年6月25日起在爱奇艺播出，并于2020年2月12日起在湖南卫视播出。该剧改编自作者芥沫的小说《天才小毒妃》。《芸汐传》截至收官，爱奇艺播放量突破30亿，微博话题阅读量10亿，讨论量达226万。2019年，该剧获得2019微博电视剧大赏人气剧集第4名。

导入文本4

《芸汐传》是改编自芥沫的长篇小说《天才小毒妃》，讲述了太医之女韩芸汐在经历了种种磨难后依旧保持积极乐观心态，并与秦王龙非夜产生了错综复杂情感的故事。当前，由上海丝芭影视出品的电视剧《芸汐传》，正在横店影视城秦王宫、春秋(唐)园及外景拍摄基地有条不紊取景拍摄之中。

合并结果

```
{
  "导演": [
    "林健龙",
    "刘镇明"
  ],
  "主演": [
    "鞠婧祎",
    "张哲瀚",
    "米热",
    "林思意",
    "王佑硕",
    "卢星宇",
    "许佳琪",
    "胡兵"
  ],
  "发布时间": [
    "2018年",
    "2018年6月25日",
    "2020年2月12日",
    "2020年7月24日"
  ],
  "名称": [
    "《芸汐传》"
  ],
  "编剧": [
    "金媛媛"
  ],
  "类型": [
    "古装"
  ],
  "监制": [
    "林国华"
  ],
  "别名": [
    "《天才小毒妃》"
  ]
}
```

文本1信息

```
{
  "导演": [
    "林健龙",
    "刘镇明"
  ],
  "主演": [
    "鞠婧祎",
    "张哲瀚",
    "米热",
    "林思意",
    "王佑硕",
    "卢星宇",
    "许佳琪"
  ],
  "发布时间": [
    "2018年"
  ]
}
```

文本2信息

```
{
  "导演": [
    "林健龙",
    "刘镇明"
  ],
  "主演": [
    "鞠婧祎",
    "米热",
    "林思意",
    "王佑硕",
    "许佳琪",
    "胡兵"
  ],
  "发布时间": [
    "2018年6月25日",
    "2020年2月12日"
  ]
}
```

文本3信息

```
{
  "导演": [
    "林健龙",
    "刘镇明"
  ],
  "主演": [
    "鞠婧祎",
    "米热"
  ],
  "发布时间": [
    "2018年6月25日",
    "2020年2月12日"
  ],
  "名称": [
    "《芸汐传》"
  ]
}
```

文本4信息

```
{
  "名称": [
    "《芸汐传》"
  ],
  "别名": [
    "《天才小毒妃》"
  ],
  "简介": [
    "太医之女韩芸汐在经历了种种磨难后依旧保持积极乐观心态，并与秦王龙非夜产生了错综复杂情感的故事"
  ]
}
```

## 8. 实验心得

本次实验考察自然语言处理中的知识提取和知识合并，在做实验二的时候也是第一次接触这个自然语言处理，对自然语言处理也不是很了解，因此一步步进行解析，到最后顺利完成这个知识的提取合并过程。首先是在网上找到关于《芸汐传》的不同叙述的样本，将这些样本作为一个简单的训练集，将样本转成 Json 格式对象的形式，如果每次处理整个 Json 数据，发现难度过高，而且分辨度过低，难以找到正确的分割地方，因此使用 *split* 对整个 Json 文本进行分割处理，变成每一小段文字，利用 *indexOf* 函数对关键字进行检索，当找到了对应的关键词之后对相对应的文本进行处理，当使用现有库中的关键词对文件进行检索，发现当关键词过少时，对样本数据适应性不高，因此在训练的时候哦反复加入和当前关键字相关的词语或者表达，提高这类关键字的对文本检测的适应性，最后以数据可视化的形式展示出来，通过本次实验，让我明白了文本识别和合并的大致过程，明白了数据集和属性的关键词在数量上越多，对文本检测识别的效果更好，深刻理解到了，大数据的使用好处。

## 9. 【附录】

前端页面代码：

```
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="UTF-8">
```



```
<title>知识获取和合并</title>
</head>

<style>
  header {
    width: 1200px;
    margin: 10px auto;
    color: darkorange;
    text-align: center;
  }

  .container {
    width: 1200px;
    height: 100px;
    margin: 10px auto;
    display: flex;
    flex-direction: row;
    /*background-color: red;*/
    justify-content: space-around;
    /*padding: 10px 10px;*/
  }

  .left {
    flex: 5;
    display: flex;
    justify-content: space-around;
    flex-wrap: wrap;
  }

  .right {
    flex: 5;
    display: flex;
    justify-content: space-around;
    flex-wrap: wrap;
  }

  .left div {
    /*width: 220px;*/
  }

  .right {
    flex: 2;
  }

  .left textarea {
    width: 200px;
```

```

        height: 250px;
        outline: none;
        resize: none
    }

    .right textarea {
        width: 200px;
        height: 500px;
        outline: none;
        resize: none
    }

    button {
        margin-top: 20px;
        margin-bottom: 20px;
    }
</style>
<body>
    <div id="App">
        <header><h1>欢迎使用由 20 级计科 2 班彭清元设计的知识获取和合并的可视化
展示</h1></header>
        <div class="container">
            <div class="left">
                <div>
                    <button @click="btn1">导入文本 1</button>
                    <div>
                        <textarea>{{msg1}}</textarea>
                    </div>
                </div>
                <div>
                    <button @click="btn2">导入文本 2</button>
                    <div>
                        <textarea>{{msg2}}</textarea>
                    </div>
                </div>
                <div>
                    <button @click="btn3">导入文本 3</button>
                    <div>
                        <textarea>{{msg3}}</textarea>
                    </div>
                </div>
                <div>
                    <button @click="btn4">导入文本 4</button>
                    <div>
                        <textarea>{{msg4}}</textarea>
                    </div>
                </div>
            </div>
        </div>
    </div>

```

```

    </div>
    <div>
        <button @click="btn5">文本 1 信息</button>
        <div>
            <textarea>{{msg5}}</textarea>
        </div>
    </div>
    <div>
        <button @click="btn6">文本 2 信息</button>
        <div>
            <textarea>{{msg6}}</textarea>
        </div>
    </div>
    <div>
        <button @click="btn7">文本 3 信息</button>
        <div>
            <textarea>{{msg7}}</textarea>
        </div>
    </div>
    <div>
        <button @click="btn8">文本 4 信息</button>
        <div>
            <textarea>{{msg8}}</textarea>
        </div>
    </div>

</div>
<div class="right">
    <div>
        <button @click="btn9">合并结果</button>
        <div>
            <textarea>{{msg9}}</textarea>
        </div>
    </div>
</div>
</div>

<script src="../axios.js"></script>
<script src="../vue.js"></script>
<script src="知识获取.js"></script>
<script src="test.js"></script>

<script>

    const vm = new Vue({

```

```

el: "#App",
data: {
    msg1: "",
    msg2: "",
    msg3: "",
    msg4: "",
    msg5: "",
    msg6: "",
    msg7: "",
    msg8: "",
    msg9: "",
    list: ""
},
created() {
    this.getdata()
},
methods: {
    async getdata() {
        this.list = await getData()
    },
    async btn1() {
        this.msg1 = this.list.one.message
    },
    btn2() {
        this.msg2 = this.list.two.message
    },
    btn3() {
        this.msg3 = this.list.three.message
    },
    btn4() {
        this.msg4 = this.list.four.message
    },
    btn5() {
        let string = this.msg1
        let message = Process1(string)
        this.msg5 = message
    },
    btn6() {
        let string = this.msg2
        let message = Process2(string)
        this.msg6 = message
    },
    btn7() {
        let string = this.msg3
        let message = Process3(string)
        this.msg7 = message
    }
}

```

```

        console.log(message)
    },
    btn8() {
        let string = this.msg4
        let message = Process4(string)
        this.msg8 = message
    },
    btn9() {
        this.msg9 = merge(this.msg5, this.msg6, this.msg7,
this.msg8)
    }
    })
</script>
</body>
</html>

```

知识提取和合并代码:

```

    list = {}
    Type = []
    Director = []
    Starring = []
    Release = []
    screenwriter = []
    producer = []
    firm = []
    FilmingLocations = []
    name = []
    alias = []

```

```

async function getData() {

    /**
     * 将数据从 json 文件中获取出来
     * @type {*}
     */
    list = await axios.get("config.json")

    return list.data
}

```

```

async function init() {

    let data = await axios.get("data1.json")

    /**

```

```

    * 数据集的相关属性
    * @type {*|string[]}
    */
    Type = data.data.Types.split(" ")
    Director = data.data.Director.split(" ")
    Starring = data.data.Starring.split(" ")
    Release = data.data.Release.split(" ")
    screenwriter = data.data.screenwriter.split(" ")
    producer = data.data.producer.split(" ")
    alias = data.data.alias.split(" ")
    reads = data.data.reads.split(" ")
    discovers = data.data.discovers.split(" ")
    Info = data.data.Info.split(" ")
}

/**
 * 《芸汐传》是由林健龙 刘镇明导演的一部国产剧，
 * 芸汐传在大陆首映，
 * 并在 2018 第一时间上映，
 * 主要演员有鞠婧祎 张哲瀚 米热 林思意 王佑硕 卢星宇 许佳 等主演，
 * 樱花动漫为大家提供芸汐传全集完整版免费在线观看，
 * 支持手机、平板、电脑多终端高清免费播放，
 * 芸汐传现更新至已完结。
 * @param string
 * @constructor
 */

function Process1(string) {

    console.log(string)

    let str = string.split(", ")

    /**
     * 判断 导演
     */

    Dir = []
    Sta = []
    Rel = []
    nam = []
    scr = []
    Tys = []

    for (let i = 0; i < str.length; i++) {

```

```

        for (let j = 0; j < Director.length; j++) {
            if (str[i].indexOf(Director[j]) !== -1) {
                Dir.push(str[i].substring(str[i].indexOf("是由") + 2,
str[i].indexOf("导演")).trim().split(" "))
            }
        }
    }

    let temp = []

    for (let i = 0; i < str.length; i++) {
        for (let j = 0; j < Starring.length; j++) {
            if (str[i].indexOf(Starring[j]) !== -1 && Starring[j] == "主要演员") {
                temp.push(str[i].substring(str[i].indexOf("主要演员有") + 5,
str[i].indexOf("等")).trim().split(" "))
            }
        }
    }

    console.log(temp)
    for (let i = 0; i < temp.length; i++) Sta.push(...temp[i])

    for (let i = 0; i < str.length; i++) {
        for (let j = 0; j < Release.length; j++) {
            if (str[i].indexOf(Release[j]) !== -1) {
                Rel.push(str[i].substring(str[i].indexOf("在") + 1,
str[i].indexOf("第一时间")))
            }
        }
    }

    for (let i = 0; i < str.length; i++) {
        if (str[i].indexOf("是") !== -1) {
            console.log(str[i].indexOf("是"))
            nam.push(str[i].substring(0, str[i].indexOf("是")))
        }
    }

    for (let i = 0; i < str.length; i++) {
        for (let j = 0; j < screenwriter.length; j++) {
            if (str[i].indexOf(screenwriter[j]) !== -1) {
                console.log(screenwriter[j])
                scr.push(str[i].substring(str[i].indexOf(",") + 1,

```

```

str[i].indexOf("编剧")).trim().split(" "))
    }
}

for (let i = 0; i < str.length; i++) {
    for (let j = 0; j < Type.length; j++) {
        if (str[i].indexOf(Type[j]) !== -1) {
            console.log(Type[j])
            Tys.push(Type[j])
        }
    }
}

let obj = {
    导演: Dir == null ? "" : Dir,
    主演: Sta == null ? "" : Sta,
    发布时间: Rel == null ? "" : Rel,
    名称: nam == null ? "" : nam,
}

return obj;
}

function Process2(string) {

    console.log(string)

    let str = string.split(", ")

    /**
     * 判断 导演
     */

    Dir = []
    Sta = []
    Rel = []
    nam = []
    scr = []
    Tys = []
    pro = []
    al = []

    for (let i = 0; i < str.length; i++) {
        for (let j = 0; j < Director.length; j++) {

```



```

        if (str[i].indexOf(Director[j]) != -1) {
            Dir.push(str[i].substring(0, str[i].indexOf("执导
"))) ).trim().split(", "))
        }
    }
}

let temp = []

for (let i = 0; i < str.length; i++) {
    for (let j = 0; j < Starring.length; j++) {
        if (str[i].indexOf(Starring[j]) != -1 && Starring[j] == "领衔主演") {
            temp.push(str[i].substring(0, str[i].indexOf("领衔主演
"))) ).trim().split(" "))
        }

        if (str[i].indexOf(Starring[j]) != -1 && Starring[j] == "联袂主演") {
            temp.push(str[i].substring(0, str[i].indexOf("等
"))) ).trim().split(", "))
        }
    }
}

for (let i = 0; i < temp.length; i++) Sta.push(...temp[i])

for (let i = 0; i < str.length; i++) {
    for (let j = 0; j < Release.length; j++) {
        if (str[i].indexOf(Release[j]) != -1) {
            Rel.push(str[i].substring(str[i].indexOf("于") + 1,
str[i].indexOf("日") + 1))
        }
    }
}

for (let i = 0; i < str.length; i++) {
    if (str[i].indexOf("是") != -1) {
        // console.log(str[i].indexOf("是"))
        nam.push(str[i].substring(0, str[i].indexOf("是")))
    }
}

for (let i = 0; i < str.length; i++) {
    for (let j = 0; j < screenwriter.length; j++) {
        if (str[i].indexOf(screenwriter[j]) != -1) {

```

```

        scr.push(str[i].substring(0, str[i].indexOf("编剧")))
    }
}

for (let i = 0; i < str.length; i++) {
    for (let j = 0; j < producer.length; j++) {
        if (str[i].indexOf(producer[j]) !== -1) {
            pro.push(str[i].substring(str[i].indexOf("是由") + 2,
str[i].indexOf("监制")))
        }
    }
}

for (let i = 0; i < str.length; i++) {
    for (let j = 0; j < Type.length; j++) {
        if (str[i].indexOf(Type[j]) !== -1) {
            console.log(Type[j])
            Tys.push(Type[j])
        }
    }
}

for (let i = 0; i < str.length; i++) {
    for (let j = 0; j < alias.length; j++) {
        if (str[i].indexOf(alias[j]) !== -1) {
            al.push(str[i].substring(str[i].indexOf("《"), str[i].indexOf("》
") + 1))
        }
    }
}

let obj = {
    导演: Dir == null ? "" : Dir,
    主演: Sta == null ? "" : Sta,
    发布时间: Rel == null ? "" : Rel,
    名称: nam == null ? "" : nam,
    编剧: scr == null ? "" : scr,
    类型: Tys == null ? "" : Tys,
    监制: pro == null ? "" : pro,
    别名: al == null ? "" : al
}

return obj;
}

```

```

function Process3(string) {

    let str = string.trim().split(/, |, |。 |\s+/)

    console.log(str)

    /**
     * 判断 导演
     */

    Dir = []
    Sta = []
    Rel = []
    nam = []
    scr = []
    Tys = []
    pro = []
    al = []
    rea = []
    dis = []

    for (let i = 0; i < str.length; i++) {
        for (let j = 0; j < Director.length; j++) {
            if (str[i].indexOf(Director[j]) != -1) {
                Dir.push(str[i].substring(0, str[i].indexOf("联合执导
")))
            }
        }
    }

    let temp = []

    for (let i = 0; i < str.length; i++) {
        for (let j = 0; j < Starring.length; j++) {
            if (str[i].indexOf(Starring[j]) != -1 && Starring[j] == "领衔主演") {
                temp.push(str[i].substring(0, str[i].indexOf("等领衔主演
")))
            }

            if (str[i].indexOf(Starring[j]) != -1 && Starring[j] == "联袂主演") {
                temp.push(str[i].substring(0, str[i].indexOf("等
")))
            }
        }
    }
}

```

```

}

for (let i = 0; i < temp.length; i++) Sta.push(...temp[i])

for (let i = 0; i < str.length; i++) {
  for (let j = 0; j < Release.length; j++) {
    if (str[i].indexOf(Release[j]) !== -1) {
      // console.log(str[i])
      Rel.push(str[i].substring(str[i].indexOf("于") + 1,
str[i].indexOf("日") + 1))
    }
  }
}

for (let i = 0; i < str.length; i++) {
  if (str[i].indexOf("是") !== -1) {
    // console.log(str[i].indexOf("是"))
    nam.push(str[i].substring(0, str[i].indexOf("是")))
  }
}

// for (let i = 0; i < str.length; i++) {
//   for (let j = 0; j < screenwriter.length; j++) {
//     if (str[i].indexOf(screenwriter[j]) !== -1) {
//       scr.push(str[i].substring(0, str[i].indexOf("编剧")))
//     }
//   }
// }

for (let i = 0; i < str.length; i++) {
  for (let j = 0; j < producer.length; j++) {
    if (str[i].indexOf(producer[j]) !== -1) {
      pro.push(str[i].substring(str[i].indexOf(", "), str[i].indexOf("
监制")))
    }
  }
}

for (let i = 0; i < str.length; i++) {
  for (let j = 0; j < Type.length; j++) {
    if (str[i].indexOf(Type[j]) !== -1) {
      // console.log(Type[j])
      Tys.push(Type[j])
    }
  }
}

```

```

    }
}

for (let i = 0; i < str.length; i++) {
    for (let j = 0; j < alias.length; j++) {
        if (str[i].indexOf(alias[j]) !== -1) {
            al.push(str[i].substring(str[i].indexOf("《"), str[i].indexOf("》
") + 1))
        }
    }
}

for (let i = 0; i < str.length; i++) {
    for (let j = 0; j < reads.length; j++) {
        if (str[i].indexOf(reads[j]) !== -1) {
            rea.push(str[i].substring(str[i].indexOf("阅读量") + 3))
        }
    }
}

for (let i = 0; i < str.length; i++) {
    for (let j = 0; j < discovers.length; j++) {
        if (str[i].indexOf(discover[j]) !== -1) {
            dis.push(str[i].substring(str[i].indexOf("讨论量达") + 4))
        }
    }
}

let obj = {
    导演: Dir == null ? "" : Dir,
    主演: Sta == null ? "" : Sta,
    发布时间: Rel == null ? "" : Rel,
    名称: nam == null ? "" : nam, // "编剧": scr == null ? "" : scr,
    类型: Tys == null ? "" : Tys,
    监制: pro == null ? "" : pro,
    别名: al == null ? "" : al,
    阅读量: rea == null ? "" : rea,
    讨论量: dis == null ? "" : dis
}

return obj;
}

```

```

function Process4(string) {

```

```

let str = string.split(", ")

// console.log(str)
/**
 * 判断 导演
 */

Dir = []
Sta = []
Rel = []
nam = []
scr = []
Tys = []
pro = []
al = []
rea = []
dis = []
info = []

for (let i = 0; i < str.length; i++) {
    for (let j = 0; j < Director.length; j++) {
        if (str[i].indexOf(Director[j]) != -1) {
            Dir.push(str[i].substring(0, str[i].indexOf("联合执导
")).trim().split("、"))
        }
    }
}

let temp = []

for (let i = 0; i < str.length; i++) {
    for (let j = 0; j < Starring.length; j++) {
        if (str[i].indexOf(Starring[j]) != -1 && Starring[j] == "领衔主演") {
            temp.push(str[i].substring(0, str[i].indexOf("等领衔主演
")).trim().split("、"))
        }

        if (str[i].indexOf(Starring[j]) != -1 && Starring[j] == "联袂主演") {
            temp.push(str[i].substring(0, str[i].indexOf("等
")).trim().split("、"))
        }
    }
}

```

```

for (let i = 0; i < temp.length; i++) Sta.push(...temp[i])

for (let i = 0; i < str.length; i++) {
  for (let j = 0; j < Release.length; j++) {
    if (str[i].indexOf(Release[j]) !== -1) {
      // console.log(str[i])
      Rel.push(str[i].substring(str[i].indexOf("于") + 1,
str[i].indexOf("日") + 1))
    }
  }
}

for (let i = 0; i < str.length; i++) {
  if (str[i].indexOf("是") !== -1) {
    // console.log(str[i].indexOf("是"))
    nam.push(str[i].substring(0, str[i].indexOf("是")))
  }
}

// for (let i = 0; i < str.length; i++) {
//   for (let j = 0; j < screenwriter.length; j++) {
//     if (str[i].indexOf(screenwriter[j]) !== -1) {
//       scr.push(str[i].substring(0, str[i].indexOf("编剧")))
//     }
//   }
// }

for (let i = 0; i < str.length; i++) {
  for (let j = 0; j < producer.length; j++) {
    if (str[i].indexOf(producer[j]) !== -1) {
      pro.push(str[i].substring(str[i].indexOf(", "), str[i].indexOf("
监制")))
    }
  }
}

for (let i = 0; i < str.length; i++) {
  for (let j = 0; j < Type.length; j++) {
    if (str[i].indexOf(Type[j]) !== -1) {
      // console.log(Type[j])
      Tys.push(Type[j])
    }
  }
}

```

```

    for (let i = 0; i < str.length; i++) {
        for (let j = 0; j < alias.length; j++) {
            if (str[i].indexOf(alias[j]) !== -1) {
                let tempstr = str[i].substring(str[i].indexOf("改编") + 2)
                al.push(tempstr.substring(tempstr.indexOf("《"),
tempstr.indexOf("》") + 1))
            }
        }
    }

    for (let i = 0; i < str.length; i++) {
        for (let j = 0; j < reads.length; j++) {
            if (str[i].indexOf(reads[j]) !== -1) {
                rea.push(str[i].substring(str[i].indexOf("阅读量") + 3))
            }
        }
    }

    for (let i = 0; i < str.length; i++) {
        for (let j = 0; j < discovers.length; j++) {
            if (str[i].indexOf(discoveres[j]) !== -1) {
                dis.push(str[i].substring(str[i].indexOf("讨论量达") + 4))
            }
        }
    }

    // for (let i = 0; i < str.length; i++) {
    //     for (let j = 0; j < Info.length; j++) {
    //         if (string.indexOf(Info[j]) !== -1) {
    //             info.push(string.substring(str[i].indexOf("讲述了") + 3))
    //         }
    //     }
    // }
    // }
    // }

    for (let j = 0; j < Info.length; j++) {
        if (string.indexOf(Info[j]) !== -1) {
            let tString = string.substring(string.indexOf("讲述了") + 3,
string.indexOf("。"))
            info.push(tString)
        }
    }

    let obj = {

```



```

        名称: nam == null ? "" : nam, // "编剧": scr == null ? "" : scr,
        别名: al == null ? "" : al,
        简介: info == null ? "" : info,
    }

    return obj;
}

function merge(string1, string2, string3, string4) {

    Dir = []
    Sta = []
    Rel = []
    nam = []
    scr = []
    Tys = []
    pro = []
    al = []
    rea = []
    dis = []
    info = []

    let obj = {
        导演: Dir, 主演: Sta, 发布时间: Rel, 名称: nam, 编剧: scr, 类型: Tys, 监
制: pro, 别名: al, 阅读量: rea, 讨论量: dis, 简介: info,
    }

    for (item in obj) {

        for (it in string1) {
            if (it == item) {
                obj[item].push(...string1[it])
            }
        }

        for (it in string2) {
            if (it == item) {
                obj[item].push(...string2[it])
            }
        }

        for (it in string3) {
            if (it == item) {
                obj[item].push(...string3[it])
            }
        }
    }
}

```

```

    for (it in string4) {
        if (it == item) {
            obj[item].push(...string4[it])
        }
    }
}

```

```

let arr = {};
Dir.forEach(item => arr[item] = item);
Dir = Object.values(arr);

```

```

temp = Dir
Dir = []

```

```

for (let i = 0; i < temp.length; i++) Dir.push(...temp[i])

```

```

arr = {};
Sta.forEach(item => arr[item] = item);
Sta = Object.values(arr);

```

```

arr = {};
Rel.forEach(item => arr[item] = item);
Rel = Object.values(arr);

```

```

arr = {};
nam.forEach(item => arr[item] = item);
nam = Object.values(arr);

```

```

arr = {};
scr.forEach(item => arr[item] = item);
scr = Object.values(arr);

```

```

arr = {};
Tys.forEach(item => arr[item] = item);
Tys = Object.values(arr);

```

```

arr = {};
pro.forEach(item => arr[item] = item);
pro = Object.values(arr);

```

```

arr = {};
al.forEach(item => arr[item] = item);
al = Object.values(arr);

```

```
arr = {};  
rea.forEach(item => arr[item] = item);  
rea = Object.values(arr);
```

```
arr = {};  
dis.forEach(item => arr[item] = item);  
dis = Object.values(arr);
```

```
arr = {};  
info.forEach(item => arr[item] = item);  
info = Object.values(arr);
```

```
obj.导演 = Dir  
obj.主演 = Sta  
obj.发布时间 = Rel  
obj.名称 = nam  
obj.别名 = al  
obj.监制 = pro  
obj.编剧 = scr  
obj.简介 = info  
obj.讨论量 = dis  
obj.阅读量 = rea  
obj.类型 = Tys
```

```
return obj
```

```
}
```

```
init()
```

文本数据处理代码:

```
/**
```

```
* 将 样本从文本中读取出来 转化成 Json 格式 存入 Json 文件中
```

```
* @type {{}}
```

```
*/
```

```
var yangben = {}
```

```
index = {
```

```
  "1": "one",
```

```
  "2": "two",
```

```
  "3": "three",
```

```
  "4": "four",
```

```
}
```

```
const fs = require("fs")
```

```
const {writeFile} = require("fs");
```

```

function WriteData(number, path) {
  fs.readFile(path, 'utf8', (err, data) => {
    if (err) console.log('读取失败')
    else {
      yangben[index[number]] = {
        message: data
      }

      let len = Object.keys(yangben).length
      if (len == 4) {
        // console.log(yangben)
        const palce = './config.json';
        writeFile(palce, JSON.stringify(yangben, null), (error) => {
          if (error) {
            console.log('An error has occurred ', error);
            return;
          }
          console.log('Data written successfully to disk');
        });
      }
    }
  })
}

function SendData() {
  for (let i = 1; i <= 4; i++) {
    WriteData(i, `样本${i}.md`)
  }
}

SendData()

```

---

实验的过程不是消极的观察，而是积极的、有计划的探测，一个成功的实验需要的是眼光、勇气和毅力。

——曲阜师范大学名誉校长  
诺贝尔奖获得者丁肇中