

# 廈門大學



## 信息学院软件工程系

### 《JAVA 程序设计》实验报告

#### 实验九

姓名：任宇

学号：33920212204567

学院：信息学院

专业：软件工程

完成时间：2024-04-24

## 一、实验目的及要求

- 熟悉字符串及正则表达式

## 二、实验题目及实现过程

### 1. 题目一：

编写程序完成：

- 1) 输出以下新闻片段中出现的单词（每个单词只输出一次）。
- 2) 输出以下新闻片段中包含 the 的句子。

#### （一）实验环境

集成开发环境：IDEA Community Edition 2022.3.2

JDK 版本：JDK17

#### （二）实现过程

对于本题，只设计一个 `NewsProcessor` 类，其中包含两个方法：`printUniqueWords`（输出新闻中所有单词）和 `printSentencesContainingThe`（输出新闻中包含 the 的所有句子）。程序实现的重点是相应的正则表达式，分析如下：

- `printUniqueWords` ----- `text.toLowerCase().split("\\W+")`;

`\\W` 是一个预定义的字符类，它匹配任何非字母数字字符，包括下划线在内的字母（a-z 和 A-Z）、数字（0-9）以外的所有字符。在 Java 字符串中，反斜杠需要被转义，因此写作 `\\W`。

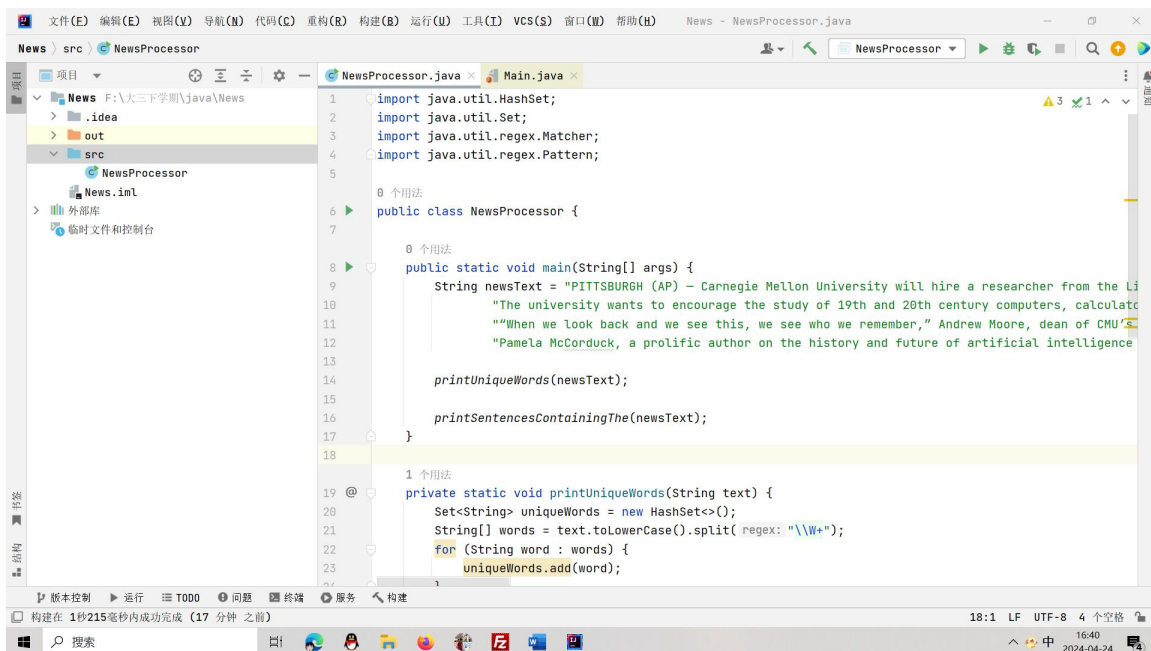
- `printSentencesContainingThe` ----- `Pattern.compile("[^.!?]+(?:[.!?](?!\"?)?\\s|$)[^.!?]*[.!?]?[\"?](?=[\\s$])", Pattern.MULTILINE)`;

- `[^.!?]+`：这部分匹配一个或多个不是句号（.）、问号（?）、或感叹号（!）的任何字符，会继续匹配直到碰到可能的句子结束符号为止。

- `(?:[!~?](?!["'\\]\s|$)[^!~?])*`: 匹配后跟一个不直接结束句子的标点符号, 识别中间的缩写和引号内的标点。
- `[!~?]["'"]?(?=\s|$)`: 匹配句子的结尾, 可能包括一个句号、问号或感叹号, 后面可能跟有一个引号。这部分使用`?`表示前面的字符(标点或引号)是可选的。`(?=\s|$)`是一个正向前瞻, 确保这个标点后面是空格或字符串的结束。

### (三) 过程截图

#### a) 全屏截图:



#### b) 主要代码:

##### Main 方法:

```

8  public static void main(String[] args) {
9      String newsText = "PITTSBURGH (AP) — Carnegie Mellon University will hire a researcher from the
10         "The university wants to encourage the study of 19th and 20th century computers, calcula
11         "When we look back and we see this, we see who we remember," Andrew Moore, dean of CMU'
12         "Pamela McCorduck, a prolific author on the history and future of artificial intelligenc
13
14      printUniqueWords(newsText);
15      printSentencesContainingThe(newsText);
16  }
  
```

printUniqueWords 方法:

```

17 @ private static void printUniqueWords(String text) {
18     Set<String> uniqueWords = new HashSet<>();
19     String[] words = text.toLowerCase().split( regex: "\\W+");
20     for (String word : words) {
21         uniqueWords.add(word);
22     }
23
24     System.out.println("新闻中不重复的单词:");
25     uniqueWords.forEach(System.out::println);
26 }

```

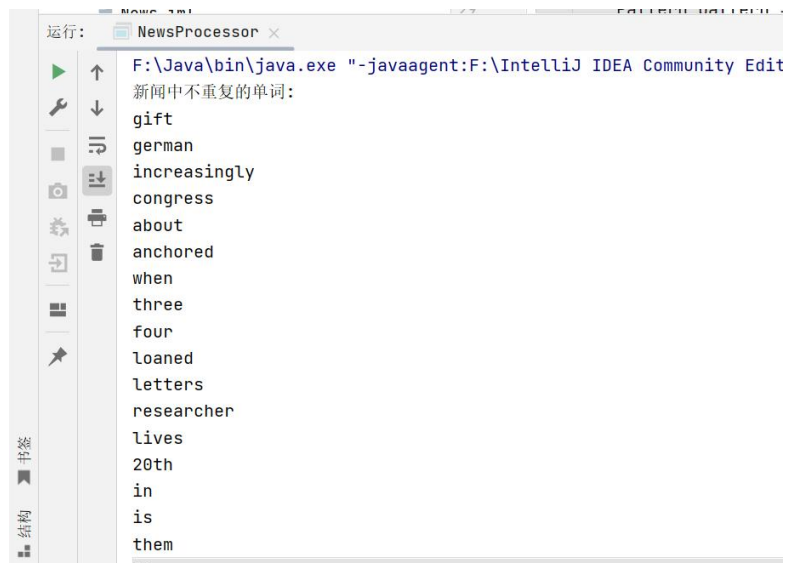
printSentencesContainingThe 方法:

```

28 private static void printSentencesContainingThe(String text) {
29     Pattern pattern = Pattern.compile( regex: "[^.!?]*(?:[.!?](?!['\"]?\\s|$)[^.!?])*[.!?]?['\"]");
30     Matcher matcher = pattern.matcher(text);
31
32     System.out.println("新闻中包含单词'the'的句子:");
33     int count=1;
34     while (matcher.find()) {
35         String sentence = matcher.group();
36         if (sentence.toLowerCase().contains(" the ")) {
37             System.out.println(count+++" "+sentence.trim());
38             count++;
39         }
40     }
41 }
42 }

```

c) 运行结果:



```

运行: NewsProcessor x
F:\Java\bin\java.exe "-javaagent:F:\IntelliJ IDEA Community Edit
新闻中不重复的单词:
gift
german
increasingly
congress
about
anchored
when
three
four
loaned
letters
researcher
lives
20th
in
is
them

```

```
with
artificial
calculators
machines
```

新闻中包含单词 'the' 的句子:

```
1:PITTSBURGH (AP) - Carnegie Mellon University will hire a researcher from the Library of Congress to help it decode a collection that includes t
2:The university wants to encourage the study of 19th and 20th century computers, calculators, encryption machines and other materials related to
3:"When we look back and we see this, we see who we remember," Andrew Moore, dean of CMU's School of Computer Science, said, adding his students
4:"We see people who took technology to save lives and save the world."
Pamela McCorduck, a prolific author on the history and future of artificial intelligence and the widow of Joseph Traub, a renowned computer scier
5:The collection, anchored by a three-rotor and four-rotor Enigma machine, is on display in the Fine and Rare Book Room in CMU's Hunt Library in
6:The gift makes CMU one of a few institutions in the United States with Enigma machines.
```

进程已结束,退出代码0

## 2. 题目二:

用正则表达式对用户输入的用户名、密码、邮箱进行判断,若不满足输入要求则提示出错类型。

- 用户名要求: 不能为空, 只能由字母、数字和\_组成, 第一位不能为数字。
- 密码要求: 不能为空, 密码长度至少 8 位, 由字母、数字、下划线组成。
- 邮箱要求: 不能为空, 需包含 " @ " 符号。 " @ " 符号后需要出现多个由 "." 分割的词。

### (一) 实验环境

集成开发环境: IDEA Community Edition 2022.3.2

JDK 版本: JDK17

### (二) 实现过程

程序只包含一个 Login 类, 用于处理用户输入的用户名、密码和邮箱的验证。设计 validateUsername、validatePassword 和 validateEmail 这三个方法用以处理用户名的验证、密码的验证和邮箱的验证, 并且使用 try-catch 结构来处理输入错误, 在发现错误时捕捉异常并提示用户重新输入。正则表达式按照题目的要求书写, 分别如下所示:

- validateUsername ----- `^[A-Za-z_][A-Za-z0-9_]*$`

^匹配输入字符串的开始, [A-Za-z\_]代表首字符必须是字母或下划线, 而 [A-Za-z0-9\_]\*说明后续字符可以是字母、数字或下划线, \*表示零个或多个。\$则匹配输入字符串的结束。

- validatePassword ----- ^\w{8,}\$

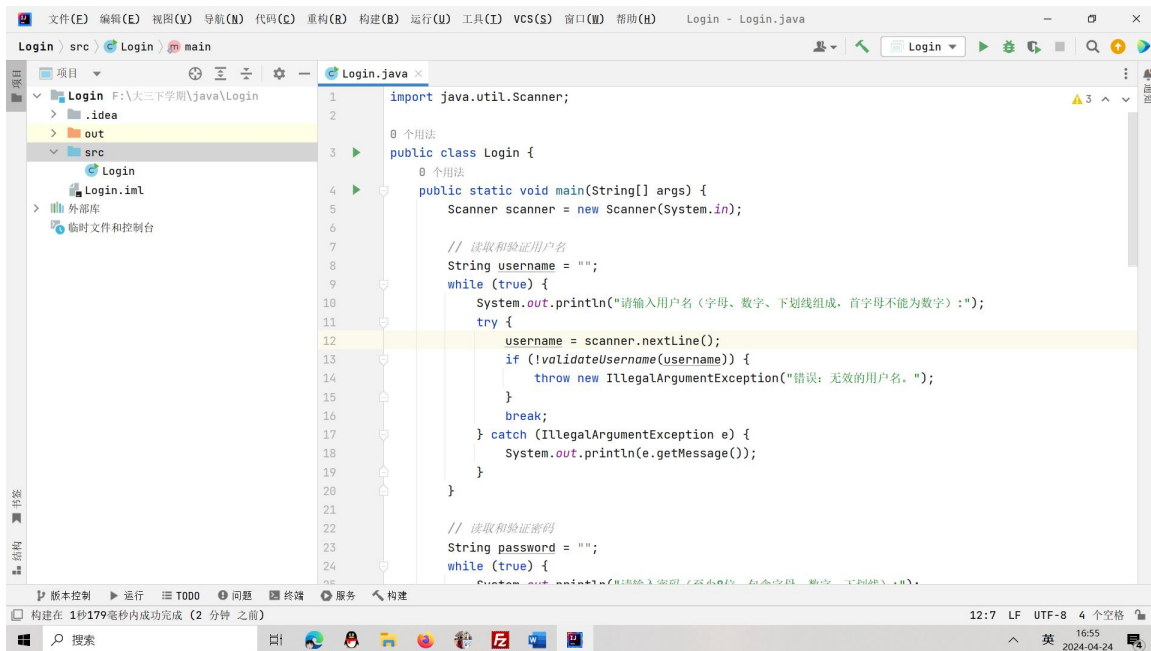
^匹配输入字符串的开始，接着\\w 匹配任何字母、数字或下划线。{8,}代表至少出现 8 次，最后\$匹配输入字符串的结束。

- validateEmail ----- ^[@\s]+@[^\s]+\.[^\s]+(?:\.[^\s]+)\*\$

^匹配输入字符串的开始。[@\s]则是要求至少一个非@和非空白字符。@是字符 "@" 必须出现一次。[@\s]+代表"@"之后至少一个非@和非空白符。\\.代表点号.必须出现一次，用来分隔域名。[^\s]+则是点号之后至少一个非@和非空白字符。最后，(?:\.[^\s]+)\*是一个非捕获组，匹配零个或多个由点号分隔的其他域名部分。\$匹配输入字符串的结束。

### (三) 过程截图

#### a) 全屏截图：



#### b) 主要代码：

Main 类主要代码（以用户名为例，另外两个类似）：

```

4 public static void main(String[] args) {
5     Scanner scanner = new Scanner(System.in);
6
7     // 读取和验证用户名
8     String username = "";
9     while (true) {
10        System.out.println("请输入用户名（字母、数字、下划线组成，首字母不能为数字）");
11        try {
12            username = scanner.nextLine();
13            if (!validateUsername(username)) {
14                throw new IllegalArgumentException("错误：无效的用户名。");
15            }
16            break;
17        } catch (IllegalArgumentException e) {
18            System.out.println(e.getMessage());
19        }
20    }

```

三个正则表达式匹配检查方法：

```

56 @ private static boolean validateUsername(String username) {
57     String usernamePattern = "^[A-Za-z_][A-Za-z0-9_]*$";
58     return username.matches(usernamePattern);
59 }
60
61 1 个用法
62 @ private static boolean validatePassword(String password) {
63     String passwordPattern = "^[\\w]{8,}$";
64     return password.matches(passwordPattern);
65 }
66
67 1 个用法
68 @ private static boolean validateEmail(String email) {
69     String emailPattern = "^[^@\\s]+@[^@\\s]+\\.([^@\\s]+(?:\\.\\s+[^@\\s]+)*)$";
70     return email.matches(emailPattern);
71 }

```

c) 运行结果：

运行： Login x

```

F:\Java\bin\java.exe "-javaagent:F:\IntelliJ IDEA Community E
请输入用户名（字母、数字、下划线组成，首字母不能为数字）：
123
错误：无效的用户名。
请输入用户名（字母、数字、下划线组成，首字母不能为数字）：
_12_ab_a1
请输入密码（至少8位，包含字母、数字、下划线）：
121a
错误：无效的密码。
请输入密码（至少8位，包含字母、数字、下划线）：
..12354637
错误：无效的密码。
请输入密码（至少8位，包含字母、数字、下划线）：
renyu0424
请输入邮箱地址：
249748231@qq
错误：无效的邮箱地址。
请输入邮箱地址：
23182ren@qq.wx.com.cn
所有输入都已成功验证。
进程已结束，退出代码0

```

### 三、实验总结与心得记录

通过这次实验，我深入理解了 Java 中字符串处理和正则表达式的应用。在编写用户名、密码和邮箱验证功能时，我实践了各种正则表达式，学习了如何精确控制输入格式，确保用户输入的数据符合预期的规范。实验过程中，我也遇到了一些挑战，特别是在处理复杂的正则表达式时，需要精确地定义每个字符和符号的意义，这要求我不断测试和调整表达式以达到目标效果。