

中山大学 软件学院 作业安排	
实验课程名称	编译原理（13 级软件工程 电政、嵌软、通软）
授课教师	万海
作业安排时间	2015 年 9 月 30 日 第二周
作业提交时间	作业一：纸质版在国庆后上课时（2015 年 10 月 10 日 星期六补星期三的课），由学委收集并提交给任课老师。 作业二：2015 年 10 月 7 日 18:00 前提交到 FTP 服务器

### 作业一

作业要求	<p>一、下列正则表达式定义了什么语言（用尽可能简短的自然语言描述）？</p> <ol style="list-style-type: none"> <li><math>((\epsilon a)b^*)^*</math></li> <li><math>((\epsilon a)b)^*</math></li> <li><math>b^*(ab^*ab^*)^*</math></li> <li><math>c^*a(a c)^*b(a b c)^* \mid c^*b(b c)^*a(a b c)^*</math></li> <li><math>(00 11)^*((01 10)(00 11)^*(01 10)(00 11)^*)^*</math></li> </ol> <p>二、按以下自然语言描述，写出正则表达式</p> <ol style="list-style-type: none"> <li>写出一个和第一题第五小题等价的且更简短的正则表达式</li> <li>利用第一题第五小题的结果，写出“由偶数个 0 和奇数个 1 构成的所有只含 0 和 1 的字符串”</li> </ol> <p>三、非空数字串只由 0-9 这十个数字组成，回答：</p> <ol style="list-style-type: none"> <li>设 <math>no\_0-8</math> 表示不含 0-8 的无重复数字的串，试用 <math>no\_0-8</math> 表示 <math>no\_0-7</math></li> <li>试用 <math>no\_0-1</math> 表示 <math>no\_0</math></li> <li>利用前两问的结果，试表示所有相邻数字都不相同的非空数字串。</li> </ol>
提交说明	作业提交纸质版。

### 作业二

作业要求	<p>Java 程序设计：</p> <p>为 UnLinker.java 文件中的 UnLinker 类实现成员函数 String clean(String text)。</p> <p>函数 clean 的功能是：屏蔽字符串参数 text 中的网页链接信息，并返回屏蔽后的结果；如果无需屏蔽，则返回原来的字符串。每个网页链接从左至右依次分成前缀，域名和后缀三部分，各部分定义如下：</p> <ol style="list-style-type: none"> <li>前缀是以下三种情况的一种： <ul style="list-style-type: none"> <li>http://</li> <li>http://www.</li> <li>www.</li> </ul> </li> <li>域名包含一个或多个字符，每个字符是一个英文字母（大小写均可），或者是一个数字（0 到 9），或者是一个句点（.）。</li> <li>后缀是以下五种情况的一种： <ul style="list-style-type: none"> <li>.com</li> <li>.org</li> <li>.edu</li> <li>.info</li> <li>.tv</li> </ul> </li> </ol>
------	--

除了上述的规定外，参数 `text` 可以含有其它的字符。另外需要注意的是，每个网页链接都会在满足上述条件的前提下尽可能地左右延伸，例如，当

`text="espihttp://www.tv.org.superwww.cali.comaladocious"` 时，其中包含的网页链接为 <http://www.tv.org.superwww.cali.com>，而不是 <http://www.tv.org> 和 [www.cali.com](http://www.cali.com)。

在屏蔽网页链接时，每个识别到的链接必须被一个字符串“OMIT”和紧跟一个数字替换，例如 `text` 中的第一个链接被“OMIT1”替换，第二个链接被“OMIT2”替换，依此类推。如下表例子所示。

参数 <code>text</code>	<code>clean (String text)</code> 的返回值
<code>espihttp://www.tv.org.superwww.cali.comaladocious</code>	<code>espiOMIT1aladocious</code>
<code>check www.foo.com 4 www.foo.com www.scoopz.com</code>	<code>check OMIT1 4 OMIT2 OMIT3</code>
<code>check www.foo.com 4 www.foo.comwww.scoopz.com</code>	<code>check OMIT1 4 OMIT2</code>
<code>check www.foo.com 4 www.foo.comhttp://scoopz.com</code>	<code>check OMIT1 4 OMIT2OMIT3</code>
<code>http://411.com goodz 4 www.733t.com, 2http://.com</code>	<code>OMIT1 goodz 4 OMIT2, 2OMIT3</code>

说明：

- 1) 字符串 `text` 的长度在 1 到 50 之间；
- 2) 字符串 `text` 只含有字母、数字、空格以及逗号 (,)、句点 (.)、冒号 (:) 和斜线 (/)。

提交说明

- 1) 只须实现 `clean` 函数，注意不要修改 `UnLinker.java` 文件中其它部分的内容（如果修改了，请在提交之前改回来），也不要修改 `clean` 的函数名，参数和返回类型；
- 2) `UnLinker.in` 文件是整个程序的输入文件，在本地机器上调试或测试程序时，可以修改 `UnLinker.in` 中的内容（用 Editplus 或 Notepad 都可以打开修改）；
- 3) 当你以命令行的方式运行程序时，可以从屏幕看到输出的结果，例如，当 `UnLinker.in` 内容为：

```
espihttp://www.tv.org.superwww.cali.comaladocious
check www.foo.com 4 www.foo.com www.scoopz.com
check www.foo.com 4 www.foo.comwww.scoopz.com
check www.foo.com 4 www.foo.comhttp://scoopz.com
http://411.com goodz 4 www.733t.com, 2http://.com

时，运行 java UnLinker，屏幕应输出

espiOMIT1aladocious
check OMIT1 4 OMIT2 OMIT3
check OMIT1 4 OMIT2
check OMIT1 4 OMIT2OMIT3
OMIT1 goodz 4 OMIT2, 2OMIT3
```

以上内容按“学号\_姓名\_实验 1.rar”打包后，提交到 FTP。

**13 级嵌软、通软、电政分别提交到各自方向所在目录：**

`ftp://my.ss.sysu.edu.cn/~wh`

再访问以下目录：

`/homework_upload/13 级 编译原理/13 级嵌软（通软、电政）/Week1（2015 年 10 月 7 日 星期三 1800 截止）/`