# 实验04 sed实验

班级：数据科学与大数据一班

学号：202026203005

姓名：张华

用户名：s13

**一、实验目的**

1. 练习使用正则表达式匹配字符串

2. 练习使用sed完成文本内容的转换

**二、实验要求**

1. 填写实验报告，请将关键命令及其结果进行截图(请确保截图中的文字清晰可见)

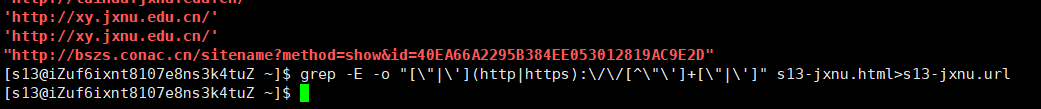
2. 导出为pdf文件，文件名为用户名-姓名-lab04.pdf，在规定截止时间之前上传作业）

3. 以下步骤中**所有s01请换成你自己的用户名。**

**三、实验步骤**

1. 复制学校首页的网页源代码并保存为s01-jxnu.html文件。

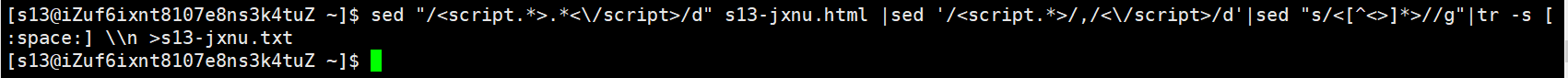
（1）提取网页中所有的绝对URL(可能在一对单引号或双引号中)到文件s01-jxnu.url中。



文本

描述已自动生成

（2）用一条管道命令去除该页面的所有javascript代码、网页标签以及多余的空白符和空行，提取出页面的文本内容并保存为s01-jxnu.txt文件。



（3）用less查看s01-jxnu.txt文件的内容。

文本

中度可信度描述已自动生成

2. 现有文件begin-end.txt的内容如下：

$ cat begin-end.txt

good begin

Begin working on that

research you wanted

to, do not let it end

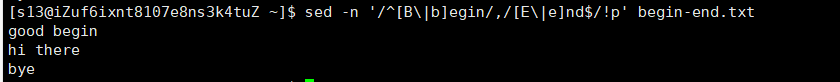
hi there

begin and strive to

reach the End

bye

（1）打印文件begin-end.txt中所有由行首为begin的行到行尾为end行所构成的行组（忽略大小写）之外的行。



（2）将文件begin-end.txt中包含长度为5的单词的行存入文件five.txt，同时将包含长度为6的单词的行存入文件six.txt。

文本

描述已自动生成

图形用户界面, 文本, 应用程序

描述已自动生成

（3）将文件hello.txt的第3-5行替换为文件begin-end.txt的第2-4行。文件hello.txt的内容如下：

$ cat hello.txt

Hello World

How are you

This game is good

Today is sunny

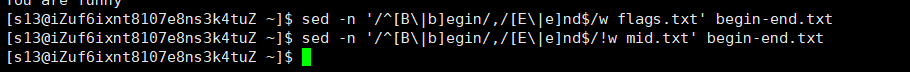
12345

You are funny

文本

描述已自动生成

（4）将文件begin-end.txt中所有包含begin或end（不区分大小写）的行写入文件flags.txt，其他行则写入文件mid.txt。



图形用户界面, 文本, 应用程序, 聊天或短信

描述已自动生成

文本

中度可信度描述已自动生成

4. 转换文本文件内容，并利用groff从纯文本生成带图表和公式的pdf论文。

（1）修改虚拟机网络设置，使虚拟机能够正常访问外网。

（2）将实验文件mypaper.txt上传到虚拟机。

（3）执行以下命令安装groff软件包

sudo apt install groff

（4）请编写sed脚本s01.sed，并执行命令

sed -f s01.sed mypaper.txt >s01-paper.ms

生成符合groff语法的s01-paper.ms文件

转换要求如下表（**左边单行内容一般转换成右边的多行，请注意换行**）

|  |  |
| --- | --- |
| **转换前内容** | **转换后变成** |
| Wuji Zhang | 你的姓名的汉语拼音 |
| <title>…content…</title> | .TL  …content… |
| <author>…content…</author> | .AU  …content… |
| <institution>…content…</institution> | .AI  …content… |
| <abstraction>…content…</abstraction> | .AB  …content…  .AE |
| <i>…content…</i>…other content… | .I …content…  …other content…  （注意：.I后面要留一个空格） |
| <h1>…content…</h1> | .NH  …content… |
| <p>  …content…  </p> | .PP  …content… |
| <foot>  …content…  </foot> | .FS  …content…  .FE |
| <li> \*  …content…  </li> | .IP \*  …content… |
| <quote>  …content…  </quote> | .QP  …content… |
| <table>  …content…  </table> | .TS  …content…  .TE |
| <equation> (1.s01)  …content…  </equation> | .EQ (1.s01)  …content…  .EN |
| <picture>  …content…  </picture> | .PS  …content…  .PE |
| <pre> C  …content…  </pre> | .DS C  …content…  .DE |
| **注意：存在一行转换成多行的情况！请严格按上表进行转换，包括空格和换行！** | |

（此处请贴上s01.sed的代码）

图片包含 表格

描述已自动生成

文本

描述已自动生成

（5）执行groff命令生成pdf文件

groff -t -e -p -ms s01-paper.ms | ps2pdf - s01-paper.pdf

（6）用pdf阅读器打开s01-paper.pdf，并欣赏你生成的论文。

（此处请贴上pdf阅读器中显示的s01-paper.pdf的内容）

文本

描述已自动生成

图示

描述已自动生成