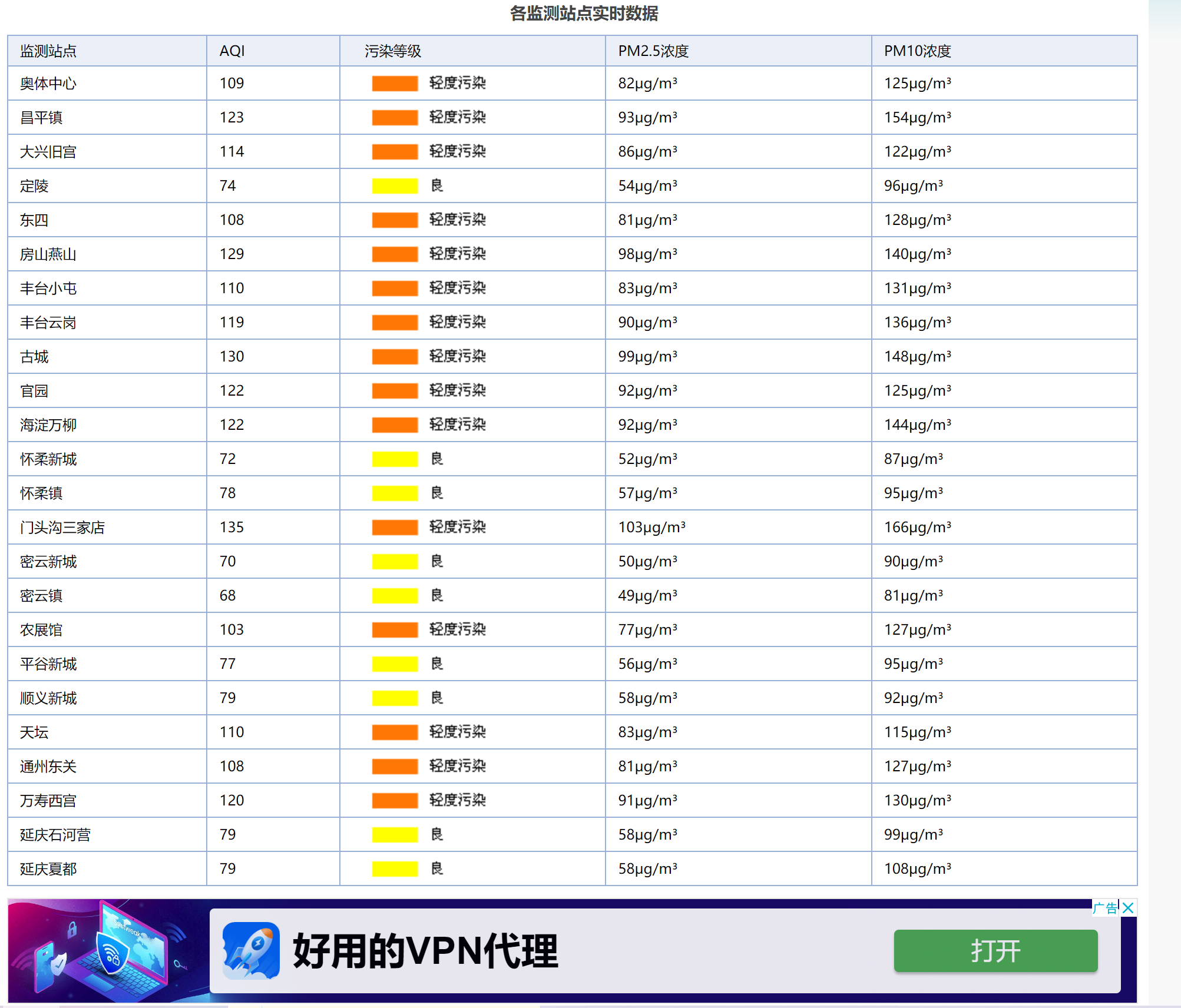
# 编程作业1--正则表达式应用

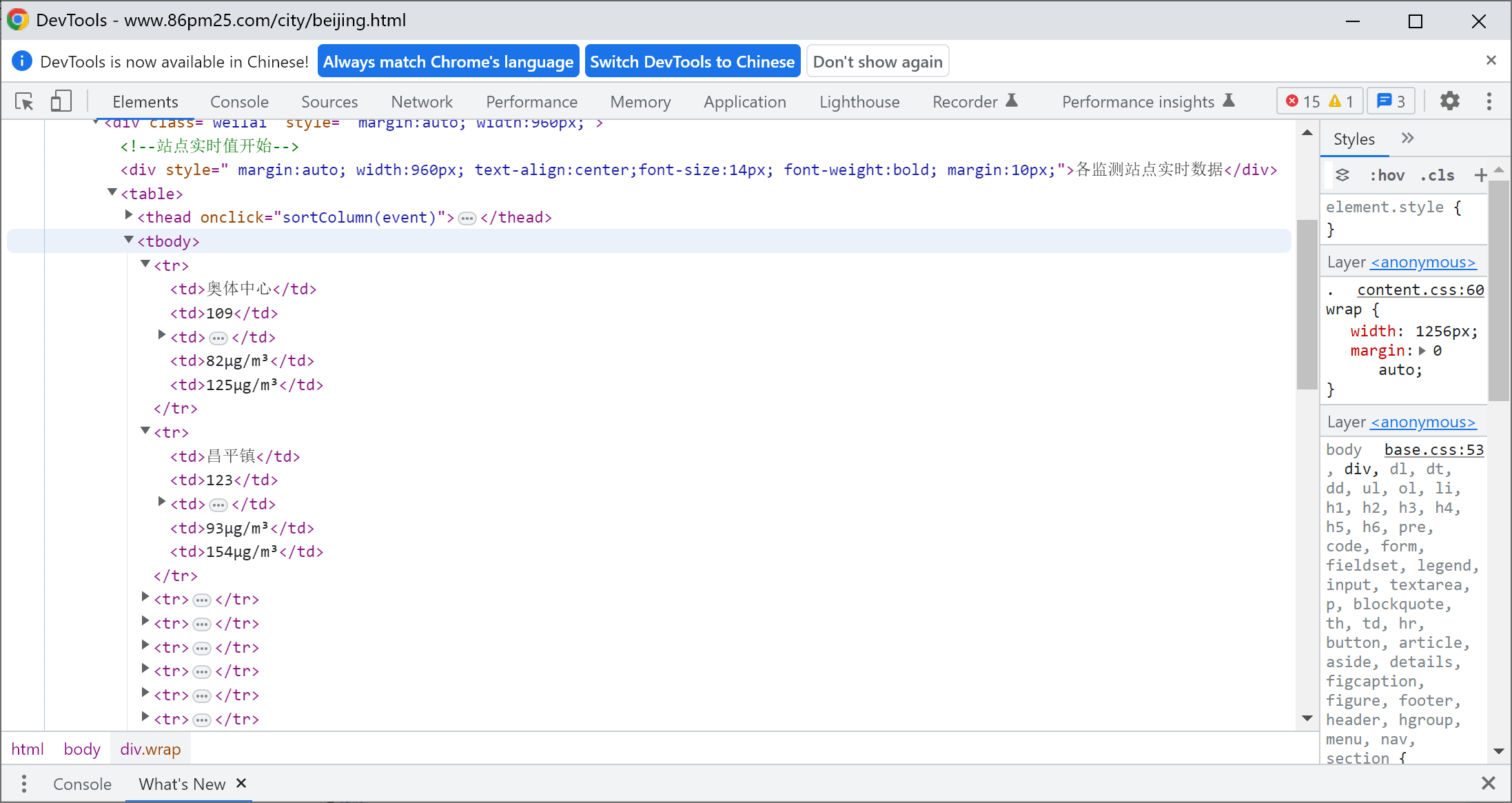
**一、实验步骤**

1、从因特网上搜索相关Web网页，找到含有北京PM2.5的网站如下：

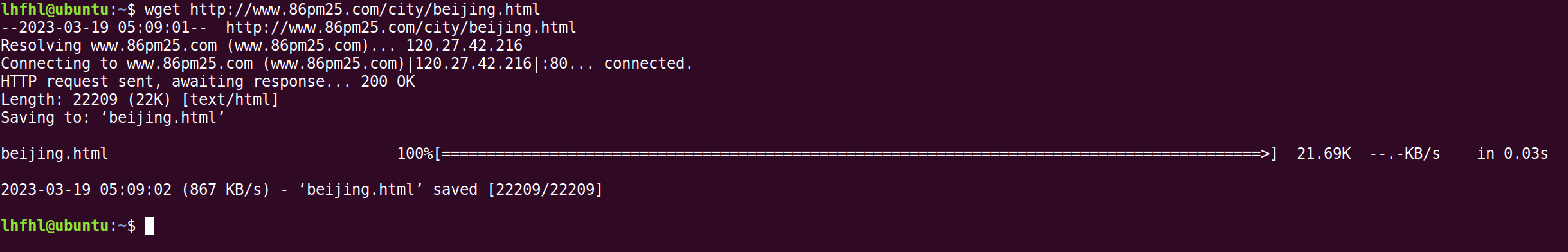
<http://www.86pm25.com/city/beijing.html>，查看网页如下：



从控制台可以看到我们需要的地点和PM2.5数据在tr和td标签中，



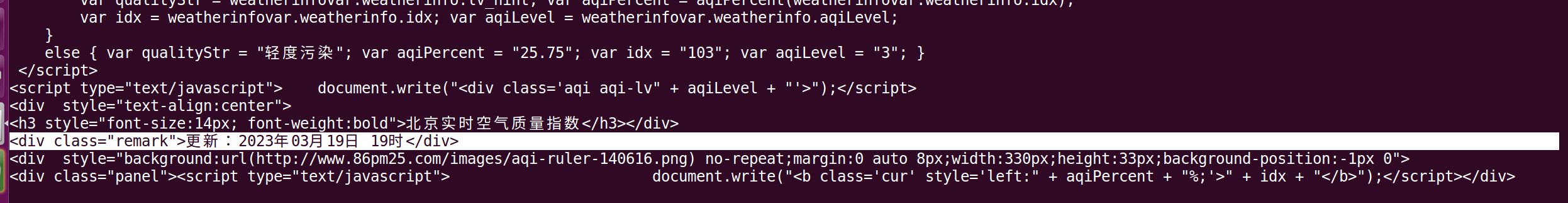
2、使用命令wegt下载网页：

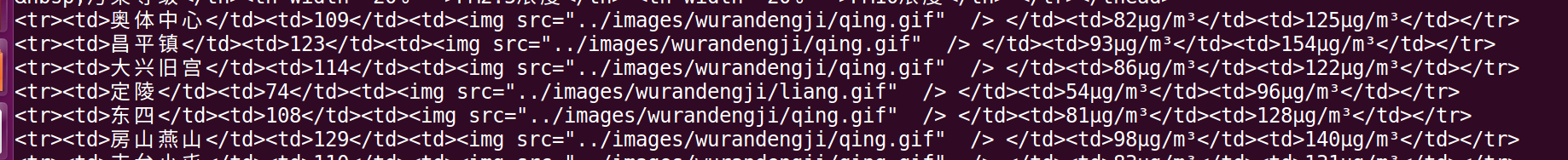


3、使用cat命令查看下载的文件：



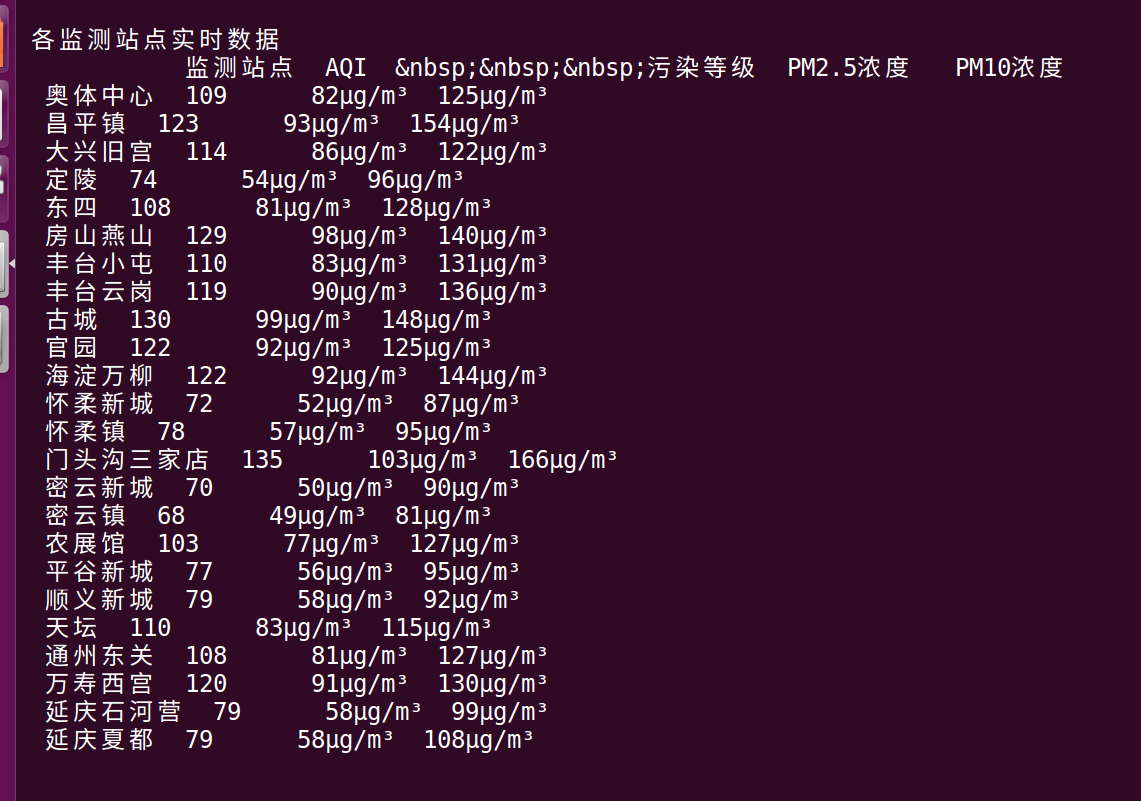
4、查看该文件，发现<div class="remark">标签和tr以及td标签里的数据是我们需要的：



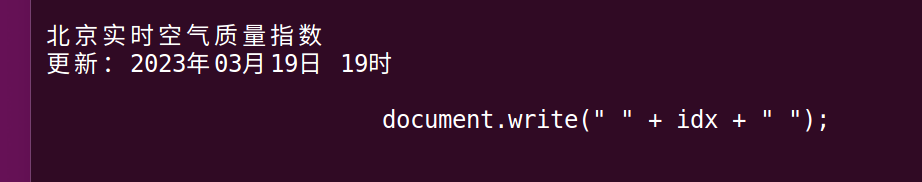


5、使用命令cat beijing.html | sed -e 's/<[^<>]\*>/ /g'将标签替换成空格





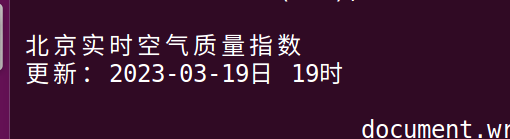
6、上述处理后的日期数据如下：



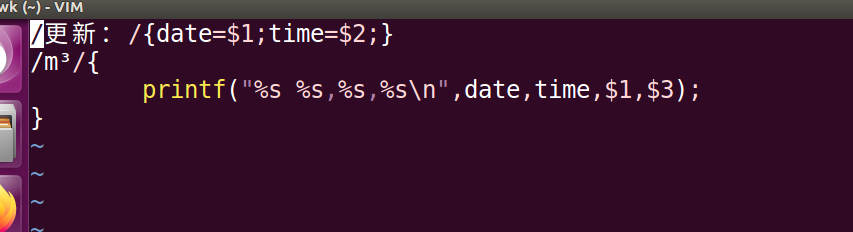
使用如下命令：

cat beijing.html | sed -e 's/<[^<>]\*>/ /g' -e 's/[年月]/-/g'

将**“年”**和**“月”**替换成所需格式的**“-”，如下：**



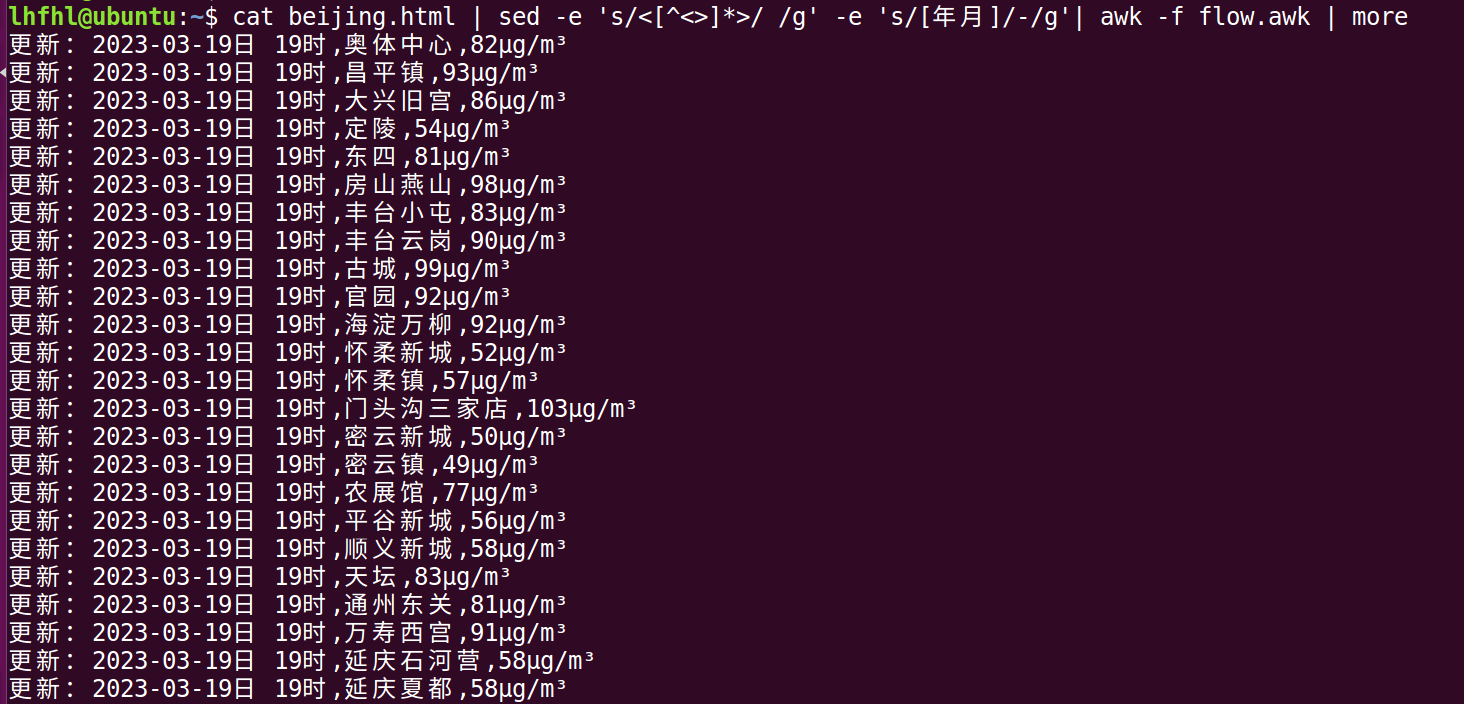
1. 为了筛选出时间和地点以及PM2.5的数据，编写flow.awk文件如下，其功能是将含有**“更新：”**的一行中的第一个和第二个数据分别赋值给date和time，将含有**m³**的一行中第一个数据**地点**和第三个数据**PM2.5**，使用**printf**语句打印：



使用如下命令：

cat beijing.html | sed -e 's/<[^<>]\*>/ /g' -e 's/[年月]/-/g'| awk -f flow.awk | more

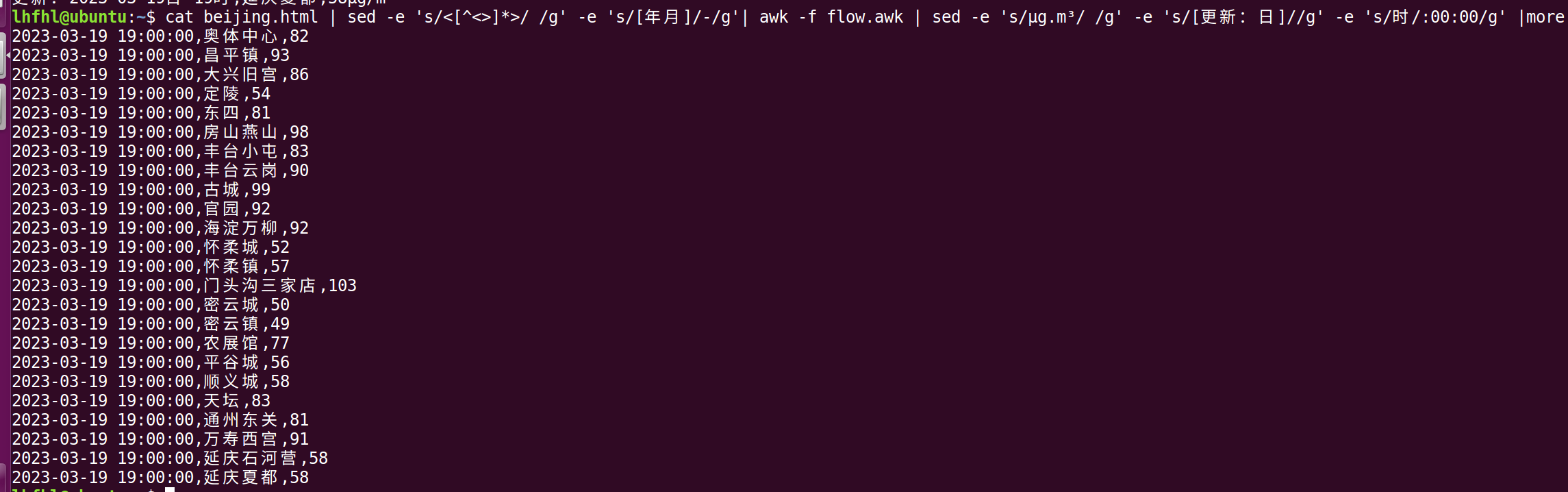
可得如下数据：



8、为了满足格式要求，需要将上图中**“更新：”**替换成空，将**“时”**替换成**“:00:00”**，将“**ug/m³**”替换成空，将**“日”**替换成空，则，最终的命令如下：

cat beijing.html | sed -e 's/<[^<>]\*>/ /g' -e 's/[年月]/-/g'| awk -f flow.awk | sed -e 's/μg.m³/ /g' -e 's/[更新：日]//g' -e 's/时/:00:00/g' |more

执行该命令后，可得满足条件的数据，如下图：



**二、实验总结**

本次实验中的多种命令使用起来并不熟悉，用的时候感觉很难**，**后来通过查找资料和观看视频逐渐熟练了起来，经历了这个过程，我对这些命令有了更深的体会，同时通过本次实验也让我对正则表达式有了更好的掌握，收获了很多。