

网址

<https://www.aqistudy.cn/>

网页数据展示

	监测点	AQI	质量等级	PM2.5	PM10	CO	NO2	O3	SO2	首要污染物
1	京东南区域点(185	中度污染	32	320	0.20	9	85	3	PM10
2	京东南区域点(185	中度污染	32	320	0.20	9	85	3	PM10
3	亦庄开发区(*)	173	中度污染	27	295	0.30	13	63	3	PM10
4	门头沟龙泉镇(137	轻度污染	19	224	0.20	9	82	3	PM10
5	大兴黄村镇(*)	137	轻度污染	28	224	0.10	22	54	3	PM10
6	平谷镇(*)	136	轻度污染	22	222	0.50	16	65	1	PM10
7	房山良乡(*)	131	轻度污染	28	212	0.50	16	64	3	PM10
8	通州东关	130	轻度污染	22	210	0.30	10	68	3	PM10
9	密云镇	118	轻度污染	20	185	0.20	8	74	3	PM10
10	密云新城	116	轻度污染	18	181	0.20	6	68	3	PM10
11	顺义新城	110	轻度污染	21	170	0.20	5	75	3	PM10
12	大兴旧宫	109	轻度污染	26	168	0.30	13	67	3	PM10

数据处理

筛选城市

```
#编辑data.awk文件 获得关键数据 时间+城市+浓度
/dateSpan"/ {print $0} #时间
/title=";/ && !/title=";";/ {print $0} #地点+浓度
```

去除冗余字符

```
sed -e 's/<span title=";\n/g' #分行
sed -e '/具体监测点/d' #删除第一行无关数据
```

获取关键信息地点+浓度

```
sed -e 's/^.*>\(.*\)<[/]span>.*datagrid-cell-c2-  
pm2_5";>[0-9][0-9]*.*$/\1/g'
```

提取时间

```
sed -e 's/^.*>.*<[/]span>$/time \1/'
```

格式化输出

```
#out.awk  
/time/ {t=$2} 3获取时间  
/!time/ {printf("%s,%s,%s",t,$1,$2)} #格式化输出
```

总语句

```
cat data.html | awk -f data.awk | sed 's/<span  
title=&quot;;\n/g' | sed -e '/具体监测点/d' | sed -e  
's/^.*>\(.*\)<[/]span>.*datagrid-cell-c2-pm2_5";>\([0-  
9][0-9]*\).*$/\1 \2/g' | sed -e 's/^.*>\(.*\)<[/]span>$/time \1/' | awk -f out.awk > out.csv
```

csv效果

	out.awk	6.csv	out.csv
			out.csv
1			2022-03-25,京东南区域点(*),32
2			2022-03-25,京东南区域点(*),32
3			2022-03-25,亦庄开发区(*),27
4			2022-03-25,门头沟龙泉镇(*),19
5			2022-03-25,大兴黄村镇(*),28
6			2022-03-25,平谷镇(*),22
7			2022-03-25,房山良乡(*),28
8			2022-03-25,通州东关,22
9			2022-03-25,密云镇,20
10			2022-03-25,密云新城,18
11			2022-03-25,顺义新城,21
12			2022-03-25,大兴旧宫,26
13			2022-03-25,延庆镇(*),52
14			2022-03-25,延庆石河营,47
15			2022-03-25,延庆夏都,52
16			2022-03-25,丰台小屯,26
17			2022-03-25,朝阳农展馆,25
18			2022-03-25,农展馆,25
19			2022-03-25,东四,24
20			2022-03-25,东城东四,24
21			2022-03-25,海淀万柳,30
22			2022-03-25,海淀万柳,30
23			2022-03-25,怀柔镇,18
24			2022-03-25,平谷新城,21
25			2022-03-25,西城官园,30
26			2022-03-25,官园,30
27			2022-03-25,石景山古城,19
28			2022-03-25,古城,19
29			2022-03-25,东城天坛,24
30			2022-03-25,天坛,24
31			2022-03-25,怀柔新城,20
32			2022-03-25,门头沟三家店,20
33			2022-03-25,朝阳奥体中心,25
34			2022-03-25,奥体中心,25
35			2022-03-25,定陵,12
36			2022-03-25,丰台云岗,19
37			2022-03-25,昌平镇,14
38			2022-03-25,房山燕山,19

