

# linux中下载pubid

```
1 curl -s "https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?
  db=pubmed&term=cotton+fiber&retmax=1000&retstart=0&retmode=xml"
  >pmids_batch_1.xml
2
3 curl -s "https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?
  db=pubmed&term=cotton+fiber&retmax=1000&retstart=1000&retmode=xml"
  >pmids_batch_2.xml
4
5 curl -s "https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?
  db=pubmed&term=cotton+fiber&retmax=1000&retstart=2000&retmode=xml"
  >pmids_batch_3.xml
6
7 curl -s "https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?
  db=pubmed&term=cotton+fiber&retmax=1000&retstart=3000&retmode=xml"
  >pmids_batch_4.xml
8
```

```
[zihwang@mn02 bionlp_course]$ curl -s "https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?
db=pubmed&term=cotton+fiber&retmax=1000&retstart=0&retmode=xml" >pmids_batch_1.xml
[zihwang@mn02 bionlp_course]$ ll
total 0
-rw-r--r-- 1 zihwang LLi 18702 Oct 15 10:41 pmids_batch_1.xml
[zihwang@mn02 bionlp_course]$ less pmids_batch_1.xml
[zihwang@mn02 bionlp_course]$ curl -s "https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?
db=pubmed&term=cotton+fiber&retmax=1000&retstart=1000&retmode=xml" >pmids_batch_2.xml
[zihwang@mn02 bionlp_course]$ curl -s "https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?
db=pubmed&term=cotton+fiber&retmax=1000&retstart=2000&retmode=xml" >pmids_batch_3.xml
[zihwang@mn02 bionlp_course]$ curl -s "https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?
db=pubmed&term=cotton+fiber&retmax=1000&retstart=3000&retmode=xml" >pmids_batch_4.xml
[zihwang@mn02 bionlp_course]$
```

```
-rw-r--r-- 1 zihwang LLi 18702 Oct 15 10:41 pmids_batch_1.xml
-rw-r--r-- 1 zihwang LLi 18705 Oct 15 10:42 pmids_batch_2.xml
-rw-r--r-- 1 zihwang LLi 18705 Oct 15 10:42 pmids_batch_3.xml
-rw-r--r-- 1 zihwang LLi 18705 Oct 15 10:42 pmids_batch_4.xml
```

## 处理pubmids文件提取出id

```
1 less pmids_batch_1.xml | awk -F "<" '{print $2}' | sed 's/ID\\>///g' | awk -F
  ">" '{print $2}' >pmids_1.xml
2 less pmids_batch_2.xml | awk -F "<" '{print $2}' | sed 's/ID\\>///g' | awk -F
  ">" '{print $2}' >pmids_2.xml
```

```

3 less pmids_batch_3.xml | awk -F "<" '{print $2}' | sed 's/ID\\>\\/g' | awk -F
">" '{print $2}' >pmids_3.xml
4 less pmids_batch_4.xml | awk -F "<" '{print $2}' | sed 's/ID\\>\\/g' | awk -F
">" '{print $2}' >pmids_4.xml
5 less pmids_batch_5.xml | awk -F "<" '{print $2}' | sed 's/ID\\>\\/g' | awk -F
">" '{print $2}' >pmids_5.xml

```

```

[zihwang@mn02 bionlp_course]$ less pmids_batch_1.xml | awk -F "<" '{print $2}' | sed 's/ID\\>\\/g' |
awk -F ">" '{print $2}' >pmids_1.xml
[zihwang@mn02 bionlp_course]$ ll
total 640
-rw-r--r-- 1 zihwang LLi 9004 Oct 15 10:53 pmids_1.xml
-rw-r--r-- 1 zihwang LLi 18702 Oct 15 10:41 pmids_batch_1.xml
-rw-r--r-- 1 zihwang LLi 18705 Oct 15 10:42 pmids_batch_2.xml
-rw-r--r-- 1 zihwang LLi 18705 Oct 15 10:42 pmids_batch_3.xml
-rw-r--r-- 1 zihwang LLi 18705 Oct 15 10:42 pmids_batch_4.xml
-rw-r--r-- 1 zihwang LLi 8916 Oct 15 10:49 pmids_batch_5.xml
[zihwang@mn02 bionlp_course]$ less pmids_batch_2.xml | awk -F "<" '{print $2}' | sed 's/ID\\>\\/g' |
awk -F ">" '{print $2}' >pmids_2.xml
[zihwang@mn02 bionlp_course]$ less pmids_batch_3.xml | awk -F "<" '{print $2}' | sed 's/ID\\>\\/g' |
awk -F ">" '{print $2}' >pmids_3.xml
[zihwang@mn02 bionlp_course]$ less pmids_batch_4.xml | awk -F "<" '{print $2}' | sed 's/ID\\>\\/g' |
awk -F ">" '{print $2}' >pmids_4.xml
[zihwang@mn02 bionlp_course]$ less pmids_batch_5.xml | awk -F "<" '{print $2}' | sed 's/ID\\>\\/g' |
awk -F ">" '{print $2}' >pmids_5.xml

```

## 去掉文件中的空行

```

1 sed -i '/^$/d' pmids_1.xml
2 sed -i '/^$/d' pmids_2.xml
3 sed -i '/^$/d' pmids_3.xml
4 sed -i '/^$/d' pmids_4.xml
5 sed -i '/^$/d' pmids_5.xml

```

```

[zihwang@mn02 bionlp_course]$ sed -i '/^$/d' pmids_1.xml
[zihwang@mn02 bionlp_course]$ less pmids_1.xml
[zihwang@mn02 bionlp_course]$ less pmids_1.xml | wc -l
1000
[zihwang@mn02 bionlp_course]$ sed -i '/^$/d' pmids_2.xml
[zihwang@mn02 bionlp_course]$ sed -i '/^$/d' pmids_3.xml
[zihwang@mn02 bionlp_course]$ sed -i '/^$/d' pmids_4.xml
[zihwang@mn02 bionlp_course]$ sed -i '/^$/d' pmids_5.xml

```

## 合并id文件

```

1 cat pmids_1.xml pmids_2.xml pmids_3.xml pmids_4.xml pmids_5.xml >pmids.xml

```

```

[zihwang@mn02 bionlp_course]$ cat pmids_1.xml pmids_2.xml pmids_3.xml pmids_4.xml pmids_5.xml >pmi
ds.xml

```

