

tsv-utils之数值型制表符文件标准化： nfilter

一、tsv-utils nfilter 介绍

功能描述：

`tsv-utils nfilter` 对数值表进行过滤行操作，通过指定某一列，将该列大于或者小于设定阈值的行过滤掉；

命令行接口：

```
1 $ tsv-utils nfilter
2
3 Usage: tsv-utils nfilter <tsv> <value>
4
5 Options:
6 -f INT    fileld for compare. default: [2].
7 -l FLAG   print line if: value larger than specified fields.
```

可选参数

```
1 -f 指定 指定某一列用来与阈值进行比较，默认指定为第二列
2 -l 标识符 如果添加该参数则将过滤掉小于阈值的行，不添加则过滤掉大于阈值的行
```

二、使用场景实例及其用法

示例演示

示例文件： `zotu_table_ann.txt`

```
1 $ cat zotu_table_ann.txt |head -n 6
```

```
1 #OTU ID A-1      A-2      B-1      B-2      C-1      C-2      taxonomy
2 ZOTU_1  0         0         87       278      1829     3608
3         d:Bacteria,p:"Proteobacteria",c:Epsilonproteobacteria,o:Campylobacterales,f:C
4         ampylobacteraceae,g:Arcobacter
5 ZOTU_2  223      447      1268     1583     52       69
6         d:Bacteria,p:"Proteobacteria",c:Gammaproteobacteria,o:Oceanospirillales,f:Hal
7         omonadaceae,g:Halomonas
8 ZOTU_3  0         0         162      159      1116     2021
9         d:Bacteria,p:"Proteobacteria",c:Alphaproteobacteria,o:Rhodospirillales,f:Acet
10        obacteraceae,g:Roseomonas
11 ZOTU_4  0         0         99       50       1250     2172
12         d:Bacteria,p:"Bacteroidetes",c:"Bacteroidia",o:"Bacteroidales",f:"Porphyromon
13         adaceae"
14 ZOTU_5  0         0         1         8        1216     2143
15         d:Bacteria,p:"Bacteroidetes"
```

运行命令：

打印出 `zotu_table_ann.txt` 中 A-1 样本 ZOTU reads 数少于 1000 的行。

```
1 $ cut -f1,2,8 zotu_table_ann.txt | tsv-utils nfilter -f2 - 500 | head -n 10
```

```
1 #OTU ID A-1      taxonomy
2 ZOTU_1  0
  d:Bacteria,p:"Proteobacteria",c:Epsilonproteobacteria,o:Campylobacterales,f:
  Campylobacteraceae,g:Arcobacter
3 ZOTU_2  223
  d:Bacteria,p:"Proteobacteria",c:Gammaproteobacteria,o:Oceanospirillales,f:Ha
  lomonadaceae,g:Halomonas
4 ZOTU_3  0
  d:Bacteria,p:"Proteobacteria",c:Alphaproteobacteria,o:Rhodospirillales,f:Ace
  tobacteraceae,g:Roseomonas
5 ZOTU_4  0
  d:Bacteria,p:"Bacteroidetes",c:"Bacteroidia",o:"Bacteroidales",f:"Porphyromo
  nadaceae"
6 ZOTU_5  0      d:Bacteria,p:"Bacteroidetes"
7 ZOTU_6  1
  d:Bacteria,p:"Proteobacteria",c:Alphaproteobacteria,o:Rhizobiales,f:Rhizobia
  ceae
8 ZOTU_7  0
  d:Bacteria,p:"Bacteroidetes",c:"Bacteroidia",o:"Bacteroidales",f:"Porphyromo
  nadaceae"
9 ZOTU_8  0
  d:Bacteria,p:"Proteobacteria",c:Gammaproteobacteria,o:Alteromonadales,f:Alte
  romonadaceae,g:Alishewanella
10 ZOTU_9  0
  d:Bacteria,p:"Bacteroidetes",c:Flavobacteriia,o:"Flavobacteriales",f:Flavoba
  cteriaceae,g:Flavobacterium
```

打印出 `zotu_table_ann.txt` 中 A-1 样本 ZOTU reads 数少于 10000 的行, 使用参数 `-l`;

```
1 $ cut -f1,2,8 zotu_table_ann.txt | tsv-utils nfilter -l -f2 - 500 |
  head -n 10
```

```

1 #OTU ID A-1      taxonomy
2 ZOTU_16 860
  d:Bacteria,p:"Proteobacteria",c:Alphaproteobacteria,o:Rhodospirillales
3 ZOTU_17 829
  d:Bacteria,p:"Bacteroidetes",c:Sphingobacteriia,o:"Sphingobacteriales",f:Chit
  tinophagaceae
4 ZOTU_19 779
  d:Bacteria,p:"Bacteroidetes",c:Sphingobacteriia,o:"Sphingobacteriales",f:Chit
  tinophagaceae,g:Parafilimonas,s:Parafilimonas_terrae
5 ZOTU_20 694
  d:Bacteria,p:"Proteobacteria",c:Alphaproteobacteria,o:Rhizobiales
6 ZOTU_21 724
  d:Bacteria,p:"Proteobacteria",c:Gammaproteobacteria,o:Xanthomonadales,f:Xant
  homonadaceae
7 ZOTU_23 603      d:Bacteria,p:"Acidobacteria",c:Acidobacteria_Gp3,g:Gp3
8 ZOTU_24 550
  d:Bacteria,p:"Proteobacteria",c:Alphaproteobacteria,o:Rhizobiales,f:Phylloba
  cteriaceae
9 ZOTU_26 607
  d:Bacteria,p:"Proteobacteria",c:Alphaproteobacteria,o:Sphingomonadales,f:Ery
  throbacteraceae
10 ZOTU_28 550
  d:Bacteria,p:"Actinobacteria",c:Actinobacteria,o:Actinomycetales,f:Nocardioi
  daceae,g:Aeromicrobium

```

本文材料为 **BASE (Biostack Applied bioinformatic SEies) 课程 Linux Command Line Tools for Life Scientists** 材料， 版权归 上海逻捷信息科技有限公司 所有。

Last Update: 8/30/2020 8:15:15 PM