

# tsv-utils之对数值矩阵排序/选择: rank

## 一、tsv-utils rank介绍

功能描述:

`tsv-utils rank` 对给定数值表进行排序, 并按照数值大小排序, 输出指定数据的行, 其余行可以归并到 `others` 分类或者指定分类。

命令行接口:

```
1 $ tsv-utils rank
2
3 Usage: tsv-utils rank [options] <tsv>
4 Options: -r INT top catalogs.
5           -m      merge remain to others catalog.
6           -e STR extend specified string to Others catalog.
```

可选参数:

```
1 -r 整数 显示指定的行数;
2 -m      合并剩下的行到一起;
3 -e      指定合并分类名称;
```

## 二、使用场景实例及其用法

使用场景经典案例:

1. 数据可视化, 比如柱状图, 显示有限分类特征。

示例演示:

示例文件: `genus.freqs.txt`

```
1 $ cat genus.freqs.txt | head -n 6
```

1	#level	A-1	A-2	B-1	B-2	C-1	C-2	
2	Parvibaculum		0.0003635		0.0002874		5.318e-05	0
		0.004845	0.004368					
3	Agromyces	0	0	0	0		0.009585	0.0114
4	Pigmentiphaga	0.000109			0.0001642		0.002765	0.001841
		5.209e-05	0					
5	Mangrovibacterium		0	0		0.002499	0.001785	0
		0						
6	Pseudonocardia	0.00189	0.001683		0	0	0	0

运行命令: 设置 `-r` 参数, 指定显示的行数, 设置 `-m` 参数, 将剩余的合并在一起, 指定合并列分类名称 `-e others`

```
1 $ tsv-utils rank -r 10 -m -e 'Others' genus.freqs.txt
```

1	#level	A-1	A-2	B-1	B-2	C-1	C-2	
2	Above_genus		0.627	0.5986	0.3409	0.3196	0.4205	0.4281
3	Alishewanella	0	0		0.1083	0.1044	0.0001042	2.992e-05
4	Arcobacter	0	0		0.01345	0.02711	0.09533	0.1081
5	Halomonas		0.008215		0.01835	0.08402	0.1065	0.002709
								0.002064
6	Flavobacterium	0	0		0.07387	0.07369	5.209e-05	0.0001496
7	Thauera	0	0	0.01899	0.01607	0.06673	0.04613	
8	Roseomonas	0	0		0.008615		0.008926	0.06111
								0.06366
9	Pseudomonas		0.003053		0.003531		0.05398	0.05729
								0.003334
								0.002752
10	Rhizobium		0.002036		0.002586		0.02484	0.03157
								0.05163
								0.05095
11	others	0.3597	0.377	0.273	0.2548	0.2985	0.298	

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 7:21:56 PM