

tsv-utils之指定的文件列进行注释: subcolumn

一、tsv-utils subcolumn 介绍

功能描述:

`tsv-utils subcolumn` 根据指定的文件或者字符串, 提取对应的列, 并合并新的文件。

命令行接口:

```
1 $ tsv-utils subcolumn
2
3 Usage:tsv-utils subcolumn <tsv> <id|list>
4
5 Options:
6   -k      query using key id.
7   -r      reorder by specified id order.
```

可选参数:

```
1 -k      选取目标列的id;
2 -r      若添加该参数, 则按指定的顺序对列进行排序;
```

二、使用场景实例及其用法

经典使用案例

16S扩增子数据分析, 抽取部分样本信息重新构建 `ZOTU` 表;

示例演示

示例文件: `zotu_table.txt`, `columns.txt`

```
1 $ cat zotu_table.txt | head -n6
```

```
1 #OTU ID A-1      A-2      B-1      B-2      C-1      C-2
2 ZOTU_1  0        0        87       278      1829     3608
3 ZOTU_2  223      447      1268     1583     52       69
4 ZOTU_3  0        0        162      159      1116     2021
5 ZOTU_4  0        0        99       50       1250     2172
6 ZOTU_5  0        0        1        8        1216     2143
```

```
1 $ cat columns.txt | head
```

```
1 #OTU ID
2 A-2
3 A-1
```

运行命令:

参数使用1: 设置 `-k` 参数, 抽取指定id的列。

```
1 | $ tsv-utils subcolumn -k zotu_table.txt "#OTU ID,A-1,A-2" | head -n6
```

1	#OTU ID	A-1	A-2
2	ZOTU_1	0	0
3	ZOTU_2	223	447
4	ZOTU_3	0	0
5	ZOTU_4	0	0
6	ZOTU_5	0	0

注意事项: 多列抽取, 在列id之间以 `,` 作为分隔符。

参数使用2: 设置 `-r` 参数, 可以按照指定的顺序对列进行排序。

```
1 | $ tsv-utils subcolumn -k -r zotu_table.txt "#OTU ID,A-2,A-1" | head -n6
```

1	#OTU ID	A-2	A-1
2	ZOTU_1	0	0
3	ZOTU_2	447	223
4	ZOTU_3	0	0
5	ZOTU_4	0	0
6	ZOTU_5	0	0

参数使用3: 使用文件列表, 设置 `-r` 参数, 可以按照指定的顺序对列进行排序。

```
1 | $tsv-utils subcolumn -r zotu_table.txt columns.txt | head -n 6
```

1	#OTU ID	A-2	A-1
2	ZOTU_1	0	0
3	ZOTU_2	447	223
4	ZOTU_3	0	0
5	ZOTU_4	0	0
6	ZOTU_5	0	0

注意事项: 如果从Windows编辑的文件, 记得 `dos2unix`

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: Friday, August 28, 2020