

tsv-utils之注释文件关联和描述: links

一、tsv-utils links介绍

功能描述：

`tsv-utils links` 在 `associate` 的基础上添加了对新的 `value` 的文字描述。

命令行接口：

```
1 $ tsv-utils links
2
3 Usage: tsv-utils links <feature-map> <links> <links-definition>
```

二、使用场景实例及其用法

示例演示

示例文件：

KEGG 功能能注释，`ko` 等级关联至更高级别，比如 `Module`

示例文件： `ko.txt`, `ko-module.txt`

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

```
1 #seqid ko
2 TRINITY_g100334_i1_1      K00370
3 TRINITY_g100485_i1_1      K00260
4 TRINITY_g101100_i1_1      K00262
5 TRINITY_g101136_i1_1      K01455
6 TRINITY_g102554_i1_1      K00370
```

```
1 $ cat ko-pathway.txt | head -n6
```

```
1 K00001 map00010
2 K00002 map00010
3 K00016 map00010
4 K00114 map00010
5 K00121 map00010
6 K00128 map00010
```

```
1 $ cat pathway-definition.txt | head -n6
```

1	map00010	Glycolysis / Gluconeogenesis
2	map00020	Citrate cycle (TCA cycle)
3	map00030	Pentose phosphate pathway
4	map00040	Pentose and glucuronate interconversions
5	map00051	Fructose and mannose metabolism
6	map00052	Galactose metabolism

运行命令:

```
1 $ tsv-utils links ko.txt ko-pathway.txt pathway-definition.txt | head -n 6
```

1	TRINITY_g100334_i1_1	map00910	Nitrogen metabolism
2	TRINITY_g100334_i1_1	map01100	Metabolic pathways
3	TRINITY_g100334_i1_1	map01120	Microbial metabolism in diverse environments
4	TRINITY_g100334_i1_1	map02020	Two-component system
5	TRINITY_g100485_i1_1	map00220	Arginine biosynthesis
6	TRINITY_g100485_i1_1	map00250	Alanine, aspartate and glutamate metabolism

注意事项: 文件顺序, 第一个输入文件为转录本的, **Transcript -> ko** 注释文件, 第二个为KEGG的提升分类(**pathway**)映射表(**ko -> pathway**), 第一列为需要关联的主键, 第三个文件为提升等级(**pathway**)的注释信息.

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 7:03:58 PM