

# tsv-utils之根据指定文件进行文件关联: associate

## 一、tsv-utils associate介绍

功能描述:

`tsv-utils associate` 根据字典: `key/value` 元素, 将指定的制表符文件的列进行关联, 将 `key/value`的`value` 元素转移至主键, 形成 `key/values` 对, 对于重复的对于关系形成一对多的关系, 比如 `gene:ko` 的注释关系, 可以通过 `ko:pathway` 字典, 构建 `gene:pathway` 的一对多关系。

命令行接口:

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

## 二、使用场景实例及其用法

示例演示

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-module.txt | head -n6
```

```
1 K10944 Ammonia monooxygenase
2 K10945 Ammonia monooxygenase
3 K10946 Ammonia monooxygenase
4 K00926 Carbamate kinase
5 K01948 Carbamoyl-phosphate synthase
6 K01672 Carbonic anhydrase
```

运行命令:

```
1 $ tsv-utils associate ko.txt ko-module.txt | head -n6
```

1	TRINITY_g100334_i1_1	Nitrite oxidoreductase
2	TRINITY_g100334_i1_1	Nitrite oxidoreductase
3	TRINITY_g100334_i1_1	Nitrite oxidoreductase
4	TRINITY_g100485_i1_1	glutamate dehydrogenase
5	TRINITY_g101100_i1_1	glutamate dehydrogenase
6	TRINITY_g101136_i1_1	Formamidase

**注意事项：** 当前输入和输出只支持两列文件， 并根据第二列进行关联。

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: 8/31/2020 5:24:15 PM