

tsv-utils 之多文件组合操作：join

一、tsv-utils join 表合并操作

功能描述：

`tsv-utils join` 合并多个制表符文件，主键为第一列，其余列合并在一个文件的不同列；

命令行接口：

```
1 $ tsv-utils join
2
3 Usage: tsv-utils join [options] [text ...]
4 Options:
5   -p CHAR  placeholder for missing value, default: ['-'];
```

可选参数：

```
1 -p 占位符，数据缺失时使用占位符代替， 数值型可以使用'0'， 文本型使用'-'
```

二、使用场景实例及其用法

使用场景经典案例：

- 16S扩增子分析：组合不同 `16S` 多样性实验的结果（比如不同文章的不同16S分区的物种分类信息），不推荐使用Closed-OTU方式构建新的OTU表，最合适的方式是单独构建OTU表，然后合并不同水平的分类结果。
- 16S扩增子分析：使用 `USEARCH otutab` 对每个文件单独构建 `ZOTU` 表，合并单个构建的 `ZOTU` 表。

示例演示：

示例文件： `A-1.zotu_table.txt`, `A-2.zotu_table.txt`, `B-1.zotu_table.txt`, `B-2.zotu_table.txt`

```
1 $ cat A-1.zotu_table.txt | head -n 6
```

```
1 #OTU ID A-1
2 ZOTU_24 550
3 ZOTU_113      136
4 ZOTU_52 261
5 ZOTU_173      88
6 ZOTU_201      58
```

运行命令：， 缺失的数值使用 `0` 补全

```
1 $ tsv-utils join -p 0 A-1.zotu_table.txt A-2.zotu_table.txt B-
  1.zotu_table.txt B-2.zotu_table.txt | head -n 6
```

1	#catalog	A-1	A-2	B-1	B-2
2	ZOTU_24 550	547	1	0	
3	ZOTU_113	136	98	0	0
4	ZOTU_52 261	279	0	0	
5	ZOTU_173	88	65	0	0
6	ZOTU_201	58	41	0	0

约定： 文件必须使用'#'作为表注释， 这个是很好的习惯， 而且后续所有程序都会默认忽略掉以'#'开头的注释行

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 6:20:48 PM