多文件组合操作： agg

一、tsv-utils agg 表汇总操作

`tsv-utils agg` 对多个制表符文件按照指定列进行合并，可以指定主键（可以组合多列），以及合并的多列；

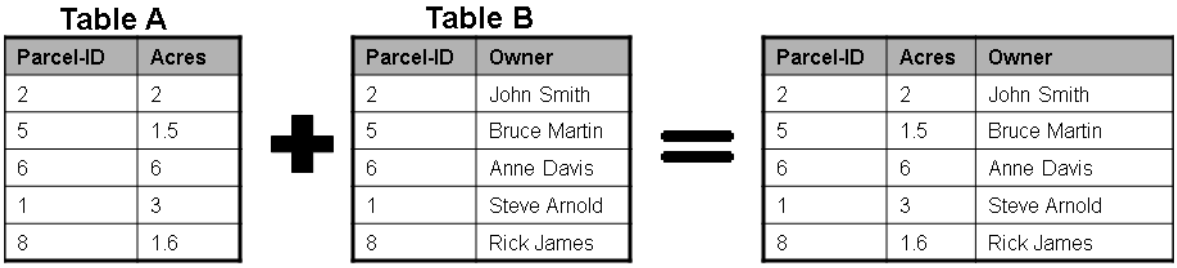

Table A

Parcel-ID	Acres
2	2
5	1.5
6	6
1	3
8	1.6

Table B

Parcel-ID	Owner
2	John Smith
5	Bruce Martin
6	Anne Davis
1	Steve Arnold
8	Rick James

Parcel-ID	Acres	Owner
2	2	John Smith
5	1.5	Bruce Martin
6	6	Anne Davis
1	3	Steve Arnold
8	1.6	Rick James

命令行接口：

```
$ tsv-utils agg

Usage: tsv-utils agg [options] [label:text ...]
Options:
  -k STR  the keys fields pattern: 1:2:3, default: [1];
  -t STR  the titles for keys: key_1:key_2:key_3, default: [catalog];
  -c INT  the target column default: [2];
  -p CHAR placehold for missing value: default ['0'];
  -i      ignore the head line;
  -v      print version number
```

可选参数：

```
-k 指定（组合） 列作为合并的主 key ，默认第一列；
-t 指定（组合） 列的表头，单列默认catalog，如果指定多列作为key，默认使用key_1， key_2 \
   描述表头，可以显示指定. -k的个数要和-t的个数对应；
-c 汇总目标列，默认使用第二列；
-p 占位符， 填充缺失数据
-i 忽略标题行
```

二、使用场景实例及其用法

经典使用场景：

`tsv-utils agg` 使用场景更多，比如多做样本数据分析是，单样本分析完，需要对分组样本进行如表数值汇总操作，比如：执行完非冗余基因集合丰度定量，需要合并成多样本表丰度表，这样可以使用tsv-utils agg轻松完成。

示例演示：

示例文件： `C11-1.genes.txt  C11-2.genes.txt`

该文件是Salmon（combine-lab.github.io/salmon/）对`RNA-seq` 数据的定量结果。

```
$cat C11-1.genes.txt | head -n 6
```

```
#gene_id        length  counts  tpm
TrG1209W        849     0       0
TrG1207W        708     5       4.76657
TrG1206C        1416    1       0.36881
TrG1204W        1119    1       0.496646
TrG1203C        1554    1       0.329413
```

```
$cat C11-2.genes.txt | head -n 6
```

```
#gene_id        length  counts  tpm
TrG1209W        849     0       0
TrG1207W        708     2       1.94121
TrG1206C        1416    1       0.365554
TrG1204W        1119    0       0
TrG1203C        1554    3       0.977823
```

合并基因的丰度信息（每个文件的第三列）, 使用`第一列`作为主键，归并`第四列`, 空值使用数值`0`填充。

```
$ tsv-utils agg  -k 1 -c 4  -p 0  C11-1:C11-1.genes.txt  C11-2:C11-
2.genes.txt | head -n 6

#catalog        C11-1   C11-2
TrG1209W        0       0
TrG1207W        4.76657 1.94121
TrG1206C        0.36881 0.365554
TrG1204W        0.496646        0
TrG1203C        0.329413        0.977823
```


Last Update: Friday, August 28, 2020