tsv-utils之多列合并：collapse

一、tsv-utils collapse介绍

功能描述：

`tsv-utils collapse` 将指定的多列元素合并为一列，并可以指定合并元素之间的分隔符；

命令行接口：

```
$ tsv-utils collapse

Usage: tsv-utils collapse [options] <tsv>
Options:
  -f  fields to collapse. Comma separated, ie: 1,2,3.
  -d  delim.
```

可选参数：

```
-f   指定要合并连续的列，如1,2,3;
-d   指定元素之间的分隔符;
```

二、使用场景实例及其用法(更换使用用例)

示例演示

示例文件： `megablast.txt，SILVA_123_SSURef.txt`

```
$ cat  blastn.txt | head -n6
```

```
ZOTU_1  JQ088343.1.1479 100.000 404     0       0       1       404     355     758     0.0     747
ZOTU_2  JF138741.1.1356 100.000 429     0       0       1       429     323     751     0.0     793
ZOTU_3  LN568336.1.1321 98.765  405     4       1       1       404     298     702     0.0     719
ZOTU_4  FJ672447.1.1441 99.764  424     1       0       1       424     355     778     0.0     778
ZOTU_5  EU542474.1.1489 100.000 424     0       0       1       424     353     776     0.0     784
ZOTU_6  JQ072613.1.1343 100.000 404     0       0       1       404     270     673     0.0     747
```

```
$ cat SILVA_123_SSURef.txt | head -n6
```

```
X92134.1.1483 Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Pseudomonas;bacterium 52N3
EF442993.1.1422 Bacteria;Proteobacteria;Deltaproteobacteria;Desulfobacterales;Desulfobulbaceae;Desulfobulbus;Desulfobulbus sp. DSM 2033
JQ972882.1.1517 Bacteria;Actinobacteria;Actinobacteria;Streptosporangiales;Thermomonosporaceae;Actinomadura;Actinomadura bangladeshensis
FJ799133.1.1480 Bacteria;Tenericutes;Mollicutes;Acholeplasmatales;Acholeplasmataceae;Acholeplasma;bacterium enrichment culture clone BA75
AY554420.1.1452 Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Proteiniphilum;Bacteroides sp. 22C
AY741401.1.1512 Bacteria;Proteobacteria;Gammaproteobacteria;NKB5;Legionella-like amoebal pathogen HT99
```

运行命令：

准备 `annotation.txt` 文件

```
$ cut -f 1,2  megablast.txt | tsv-utils annotation -c 2 SILVA_123_SSURef.txt - | tee annotation.txt | head -n6
```

```
ZOTU_1  JQ088343.1.1479 Bacteria;Proteobacteria;Epsilonproteobacteria;Campylobacterales;Campylobacteraceae;Arcobacter;uncultured bacterium
ZOTU_2  JF138741.1.1356 Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;Halomonadaceae;Halomonas;uncultured bacterium
ZOTU_3  LN568336.1.1321 Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Acetobacteraceae;Roseomonas;uncultured bacterium
ZOTU_4  FJ672447.1.1441 Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Proteiniphilum;uncultured bacterium
ZOTU_5  EU542474.1.1489 Bacteria;Bacteroidetes;Bacteroidetes vadinHA17;uncultured bacterium
ZOTU_6  JQ072613.1.1343 Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;uncultured bacterium
```

合并第二列和第三列, 使用 `-f2` 参数.

```
$ tsv-utils  collapse -f2 -d':'  annotation.txt  | head -n 6
```

```
ZOTU_1 JQ088343.1.1479:Bacteria;Proteobacteria;Epsilonproteobacteria;Campylobacterales;Campylobacteraceae;Arcobacter;uncultured bacterium
ZOTU_2 JF138741.1.1356:Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;Halomonadaceae;Halomonas;uncultured bacterium
ZOTU_3 LN568336.1.1321:Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Acetobacteraceae;Roseomonas;uncultured bacterium
ZOTU_4 FJ672447.1.1441:Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Proteiniphilum;uncultured bacterium
ZOTU_5  EU542474.1.1489:Bacteria;Bacteroidetes;Bacteroidetes vadinHA17;uncultured bacterium
ZOTU_6 JQ072613.1.1343:Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium;uncultured bacterium
```

注意事项: 当前实现只支持单字符分隔符.

Last Update: 8/30/2020 7:21:56 PM