

```
#!/usr/bin/perl env
```

```
# prints tsv as:
```

```
# species name\tmass\tlifespan
```

```
# original table had double rows for some species.
```

```
# output table only includes the second row
```

```
# Atanasov should have made up his mind about the mass of Eudyptes chrysocome.
```

```
use v5.10;
```

```
use strict;
```

```
use warnings;
```

```
use Data::Dumper;
```

```
# hash table as:
```

```
# species name => { MASS => value, LIFESPAN => value }
```

```
my %table;
```

```
my $current_species = "";
```

```
my $column_counter = 0;
```

```
while(<>){
```

```
    chomp;
```

```
    # remove leading whitespace
```

```
    s/^\s*//;
```

```
    # skip the Order lines
```

```
    next if m/^\d{0,2}[gd]er/;
```

```
    # grab the species name, set it as key for the table, and go to next line
```

```
    if ( m/^\d+\s+(\w+\s?\w+)/ ){
```

```
        $current_species = $1;
```

```
        $table{ $current_species } = {};
```

```
        $column_counter = 0;
```

```
        next;
```

```
    }
```

```
    # append fragmented species name to current_species and update the table
```

```
    if ( m/^\s+([a-z]+)/ ){
```

```
        my $fragment .= $1;
```

```
        my $ref = $table{$current_species};
```

```
        delete $table{$current_species};
```

```
        $table{$current_species . ' ' . $fragment} = $ref;
```

```
        next;
```

```
    }
```

```
    # grab mass
```

```
    if ( $column_counter == 0 ){
```

```
        $table{$current_species}{MASS} = $_;
    }
    # grab lifespan
    if ($column_counter == 2) {
        $table{$current_species}{LIFESPAN} = $_;
    }

    $column_counter++;
}

# print Dumper(\%table);

foreach my $species (sort keys %table) {
    my $mass = $table{$species}{MASS};
    my $lifespan = $table{$species}{LIFESPAN};

    $species =~ s/_/_/;
    say "$species\t$mass\t$lifespan";
}
```