

# Puzzle on Grasshoppers



## Assignment

- Read a file of Grasshopper species (*Orthoptera*).
- Determine statistics of the data.

## Description

This exercise is about Grasshopper and Cricket species (*Orthoptera*) in Europe. The text file 'orthoptera.txt' contains information on at least 1000 species and subspecies. Species are listed as follows:

Acrididae; Acanthacris ruficornis (Fabricius, 1787); Acanthacris;; ruficornis;; (Fabricius, 1787)

This line shows us the following information for this species:

- Family: *Acrididae*
- Species/Subspecies: *Acanthacris ruficornis*
- Genus: *Acanthacris*
- Species name: *ruficornis*
- Author and Year: Fabricius, 1787

Note that there is a possible confusion between the fields SPECIES/SUBSPECIES and SPECIES NAME! The first one, of course, is the full scientific name of the species. The latter is part of the full name.

The file must be read into a python data structure and then processed to obtain some statistic information about the species.

## Assignment

Pick up the file 'orthoptera.txt' from Blackboard.

Use notepad to take a look at the file. Make a plan for reading and processing the file. What data structures are you going to use to store all the information in the file? Do you use one list, or do you use multiple lists? Do you use dictionaries? Where would dictionaries come in handy? Do you make a dictionary of lists or do you make a list of dictionaries? The proper choice for the data structure may make any processing later a lot easier!

List the names of the fields and describe their meaning.

Which of the fields are "atomic" *i.e.* have only a single value?

Are there any fields you would like to split? (hint: also take a look at the questions below). How would you split these fields?

Do all the values have a uniform shape or are there things that we have to take into account?

**Use Python** to answer the following questions:

How many species and subspecies are listed?

How many distinct *Orthoptera* families are known inside Europe?

What is the family with most species?

How many species, not counting the subspecies, are listed?

What is the name of the genus has most subspecies?

How many species are described by Linnaeus? Did he ever publish with someone else?

Did Linnaeus describe any subspecies?

During which period did Linnaeus describe all these species?

How many other scientists were active during the active years of Linnaeus?

How many species and subspecies were described by a Dutch *Orthoptera* specialist called Fer Willemsen? Who are the authors that Fer Willemsen worked with?

What is the name of the species most recently described? When was that?

### Advanced

In what way can the value of a SPECIES/SUBSPECIES field be constructed from other fields?

Are there records in the list for which the information in the SPECIES/SUBSPECIES field does **not** correspond with the other fields?

Plot a histogram of the year and the number of species described in that year. In which century were most *Orthoptera* species described?

How many *Orthoptera* species occur in the Netherlands/your country?

### Hints

Structure your Python code in logical blocks!

The following functions/methods/keywords may come in handy:

open, close, readline, split, strip, append, set, sorted, max, len, zip, in, ...  
(pylab): plot, hist