TP n°3 Unix

samuel.ortion@universite-paris-saclay.fr

Foreword

Wind turbines have an impact on biodiversity inducing an increase in bird and bat mortality. The main cause of this mortality is not the collision between the organism and the rotor blade but rather because of the high pressure induced by the rotation of the wind turbine.

The DREAL of the "Hauts de France" region deposits on data.gouv.fr a database that is updated regularly counting the dead corpse of bird and bats close to the wind turbine [1].

The objective of this practical exercise is to analyse some of these data using Unix commands.

1 Questions

1. How many corpse was discovered during the year 2020?
2. Count the number of corpse discovered per species. Which species was the most commonly found? Use the command `uniq -c`.
3. Use an `awk` script to answer the same question [2]
4. Two main taxonomic groups are affected by wind turbines: bats and birds. In the file "bird.txt" you have the list of bird species referenced in the database. Using this list, extract the list of other taxa (i.e., non bird taxa).
5. Based on the list of birds and bats, and using the commands `sort`, `join`, `cut` and `uniq` check that we found more bird corpses than bat corpses.
6. Looking at the content of the file `bird.txt`, maybe you discovered that the binomial names of the species were not always formatted with the genus name and species name as the standard *Genus species* format, with the first letter of the genus name in uppercase.

   Using an awk script, fix the formatting of all binomial names.

   Hint Don't forget the `toupper` and `substr` command.

[1] Data source: `https://www.data.gouv.fr/datasets/cadavres-doiseaux-et-de-chiropteres-retrouves-au-pied-des-eoli` (visited on 7 novembre 2025).

[2] In the GNU Awk implementation, we have the function `asorti` which allows to sort a table. This might be useful.