

---

# Fourth Lab

Língua Natural – MEIC-A and MEIC-T

N-grams

---

Luísa Coheur

`luisa.coheur@inesc-id.pt`

2019

---

## Goal:

- Conduct a proper experiment;
- Play with n-grams (and check, how Unix can help (again)).
- Make a small experiment in text generation.

1. Take some time analysing the three files that are given to you.
  - The first is `alice.txt`, a text file that was already used in the first lab.
  - The second is `n-grams.sh` (to execute: `./n-grams.sh`). It calculates 2-grams, 3-grams and 4-grams from `alice.txt`. It also shows the 10 most frequent n-grams.
  - The third file is `n-grams.py` (to execute: `python3 n-grams.py`). Based on the n-grams frequency previously calculated, it allows you to:
    - calculate the probability of a sentence;
    - generate sentences (given a trigger word).
2. Play with the given programs. Try to improve them. Could the generator only use the most frequent sequences? What do you need to do to use trigrams instead of bigrams? And what about the smoothing techniques you have studied?