

- For the N-gram Models (exercise #2) do I need to compute all 26 rows by hand?

It is possible to do all 26, even 650 rows, only with bash commands, you need to use all your bash skills and this is an excellent exercise! And there is several way to perform it.

I encourage you to make the bash script that completely automatize the statistics generation. This method has the advantage to let you work alone or in groups. (for instance one looks for the command that cleans the unuseful data, another one makes the script that generates the bigrams based on the cleaned file etc.).

Don't forget what we learned in class: decompose as much as possible the difficulties. And above all, make sure, before automation, that you know what to do by hand on a short example!

- For generation of N-gram models. Do I need to know special algorithmic tools like 'for' loop 'while' ?

No.

- For generation of N-gram models. Are you sure it is possible to make operations (like addition, division) in bash?

The short answer: Yes.

Your question is more, 'how you can figure it out by myself?' Remember we did an exercise in class where you had to help Bob finding a command (introduction lecture). Finding the right command on the web is a useful skill and I want you to practice it.

Some hints:

- 1) Define your problem. For instance 'I have a file with numbers. How do I add a value (+1, +25, -100 for instance) to each numbers?'
- 2)Based on your formulation think about keywords, google them. Keywords can be verbs, but as well command names that you think might help this task. For this one I personally tried 'sed regex numbers and add'. If those keywords do not work. You can try with synonyms, or co-hyponyms.
- 3)Often the stackoverflow pages have your answer. Don't always limit yourself to the first answer given. And sometimes, accept to use a new command. For instance for this one, I found a user offering a nice solution with 'perl -pe' command. So I used it.