

CS463: Natural Language Processing

Assignment-1: Auto Filling Text

Instructions:

- 1- Students will form teams of **3** students (**Can be from different groups**).
- 2- Deadline of submission is **Friday, March 24th at 11:55 pm.**
- 3- Submission will be on google classroom.
- 4- No late submission is allowed.
- 5- No submission through e-mails.
- 6- File Naming (team ids) → No grade for wrong ids and missing ids.
- 7- **In case of Cheating, you will get a **negative grade** whether you give the code to someone, take the code from someone/internet, or even send it to someone for any reason.**
- 8- You have to write clean code and follow a good coding style including choosing meaningful variable names.

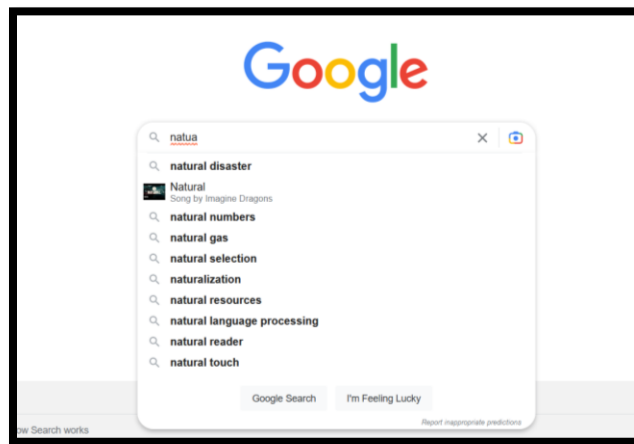


Figure 1: Example of Assignment output

Task Details:

Write a python program to build an auto filling program using n-gram models. The dataset used for this assignment is “news articles corpus”, you can create a sample of 200,000 words as input of the assignment.

<https://www.kaggle.com/datasets/sbhatti/news-articles-corpus>

Using Write a program to build a GUI auto filling program using trigram model. The dataset used for this assignment is “news articles corpus” 800k article, you can create a sample of 200,000 words as input for the assignment. Submit your code .py or .ipynb and pdf showing different inputs trials in your GUI (as in figure 1).