

# Assignment 3

## Named-Entity Identification

CS626: Speech and Natural Language Processing and the Web

### Problem statement

- Perform Named-Entity Identification using SVM classifier with appropriate feature engineering
- **Dataset:** CoNLL NER Data; <https://paperswithcode.com/dataset/conll-2003> and <https://huggingface.co/datasets/conll2003> (they are same data, but have common and distinct information)
- **Input:** A sentence
- **Output:** Name-No Name tagged for each word in the sentence
- **Example:**
  - *Input: Delhi is the capital of India.*
  - *Output: Delhi\_1 is\_0 the\_0 capital\_0 of\_0 India\_1 .\_0*
- **Important:** You will need to design the features appropriately so that the feature vector is able to distinguish name from no-name. For example, capitalization is a strong feature, though not all-powerful because starting words in a sentence can have capitalization. Think well on feature engineering.
- Create a document which reports the following
  - P, R, F1-scores
  - Perform detailed error analysis
  - Write a short paragraph on your learning.

### Submission instructions

- The submission link will be created on moodle to submit the assignment
- Only one person from the group with the lowest id is supposed to make the submission
- The name of the folder should be <id1\_id2\_id3>\_Assignment2.zip
  - The uncompressed folder should contain code, readme and the slides used for presentation
  - The readme should contain details about the tools, versions, pre-requisites if any, and how to run the code for both approaches.

### Deadline

- No-Hard deadline (Continuous Evaluation). Evaluation date will be announced soon.