****

**CSCE-5290**

**NATURAL LANGUAGE PROCESSING**

**Faculty:- Professor Sayed Shah**

**Project Title:- Name Entity recognition and sequence classification.**

**Team Mates:**

**1)Rajesh Nemani**

**2)Sameer Syed**

**3) Aishwarya Beemanapalli**

**4)Aravind Chennuri**

What is Name Entity recognition (NER)?

Name Entity recognition is one of the applications of Natural Language Processing that which can process and comprehend a lot of unstructured human language. additionally referred to as entity extraction, entity chunking, and entity identification The initial stage in providing answers, getting information, and subject modeling is NER extraction. Depending on the requirements of the application, many NER implementation models exist.

Why Name Entity Recognition?

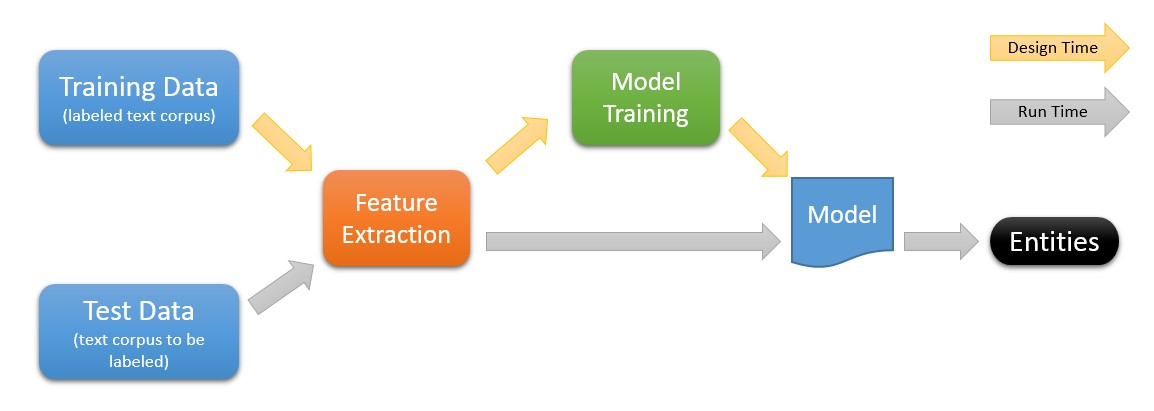
An NER system can find entity elements in raw data and identify the category to which the element belongs. The computer reads the statement and underlines any significant entity components. Depending on the task, NER could be allocated different sensitive entities.

NER is used in many different fields, including news and media, search engines, and content recommendations.

Examples of Named Entity Recognition Use Cases:

* Information Extraction Systems
* Question-Answer Systems
* Machine Translation Systems
* Automatic Summarizing Systems
* Semantic Annotation

**Workflow of NER**:



**Aim and Objectives**:

* Apply feature-based supervised machine learning techniques to the categorization of documents and tokens.
* Examine HMMs and CRFs as sequence classification methods.
* Learn the specifics of named entity recognition, document categorization, and word classification.
* Utilize pre-trained word embeddings and assess their value for the NER task.

**Implementation**:

Named Entity Recognition (NER) is a job that requires identifying names of people, places, organizations, and dates in free text. Investigate the dataset, the categories, and the mean, standard deviation, minimum, and maximum. To examine the dataset, use the data frame and the panda’s package. create a classifier for this dataset using the describe function. NER consists of two subtasks: labeling the expressions and determining the boundaries of such expressions the open and close brackets with tags such as PER, LOC or ORG. Using the data's BIO encoding, this sequence labeling challenge is transformed into a classification task.

**References:**

“<https://blog.vsoftconsulting.com/blog/understanding-named-entity-recognition-pre-trained-models>”

“<https://learn.microsoft.com/en-us/archive/blogs/machinelearning/machine-learning-and-text-analytics>”

“<https://github.com/shaniklein/NLP-Projects>”