Ziroh Labs Project Documentation

Project Plan: -

Background: -

Stanford NER is a Java implementation of a Named Entity Recognizer. Named Entity Recognition (NER) labels sequences of words in a text which are the names of things, such as person and company names, or gene and protein names. It comes with well-engineered feature extractors for Named Entity Recognition, and many options for defining feature extractors.

Problem Statement: -

- Identifying Phone Numbers, SSN/Aadhar and Names in Documents and moving them to Zunu drive

Objectives/Goals: -

- To do the project in Maven and work using the Stanford-ner libraries accordingly.

How To Implement: -

1. Download the zipped stanford-ner folder.
2. Unzip the folder to the location you want.
3. Open IntelliJ on your desktop.
4. Click on open file and select the unzipped folder or you can directly drag and drop the file.
5. Go to src ->main->java->NerDemo.java
6. If JDK 20 is not downloaded go to Configure (top right corner) select Download JDK and select JDK 20 then Apply->Ok
7. After Downloading JDK 20 change the location of the serialized classifiers.
8. To access them go to the downloaded folder then click on "Classifiers", then click on any classifier, Right Click-> Copy as Path and paste it in the Serialized Classifier Line.
9. After completing these steps run the program and enter the location of the folder and your files will be displayed accordingly with the implementation of ner.

Conclusion: -

- Hence, we have got our desired output and have successfully implemented the code using stanford-ner packages.

Team Members: -

1. Arnav Deshmukh (PRN: - 1032211848)
2. Shlok Sarda (PRN: - 1032212328)
3. Sayyam Saboo (PRN: - 1032211097)
4. Viyasha Thakkar (PRN: - 1032211595)
5. Preshika Giri (PRN: - 1032221686)