GENAI ASSIGNMENT

6th Semester, Academic Year 2026

Date:27-01-2026

Name: Dhruv Thakur	SRN:PES2UG23CS175	Section C

Project Title

Smart Resume Parser using Named Entity Recognition (NER)

Problem Statement

To build an AI system that automatically extracts important information such as Name, University, and Company from unstructured resume text using Natural Language Processing techniques and pre-trained Transformer models.

Abstract

This project implements a Smart Resume Parser using Hugging Face's BERT-based Named Entity Recognition (NER) pipeline. The system takes raw resume text as input and identifies key entities such as person names and organizations. These entities are then mapped to structured fields like candidate name, educational institution, and previous company. The objective of the project is to demonstrate the practical application of NLP pipelines for information extraction, which is widely used in recruitment and HR automation systems.

Short Documentation (What I understood and what I built)

In this project, I learned how pre-trained Transformer models can be used for real-world information extraction tasks. I used the Hugging Face Transformers library and loaded a BERT-based NER model using the pipeline interface. The model detects entities such as PER (Person) and ORG (Organization) from unstructured text.

The resume text is passed through the NER pipeline, which returns a list of detected entities along with their labels. Logical rules are applied to select the first PER entity as the candidate's name. ORG entities are further filtered using keyword matching (such as "University" or "College") to classify them as either educational institutions or companies.

Finally, the extracted information is displayed in a structured format containing:

- Name
- University
- Company

This system demonstrates how NLP can be used to automate resume screening and information extraction in recruitment systems.

sample output

```
name = None
universities = []
companies = []

for e in entities:
    if e["entity_group"] == "PER" and name is None:
        name = e["word"]
    elif e["entity_group"] == "ORG":
        if "university" in e["word"].lower() or "college" in e["word"].lower():
            universities.append(e["word"])
        else:
            companies.append(e["word"])

print("Name:", name)
print("University:", universities)
print("Company:", companies)
```

```
Name: Dhruv Thakur
University: ['PES University']
Company: ['Council of Scientific and Industrial Research', 'CSIR', '##V', 'NLP', 'Computer Vision']
```

```
result = {
    "Name": name,
    "University": universities[0] if universities else "Not Found",
    "Company": companies[0] if companies else "Not Found"
}

result
```

```
{'Name': 'Dhruv Thakur',
 'University': 'PES University',
 'Company': 'Council of Scientific and Industrial Research'}
```

github link:

https://github.com/dhruvv154/PES2UG23CS175_Genai_Labs/tree/main/UNIT-1/PROJECT