Named Entity Recognition (NER) Enhanced with Streamlit for Intuitive Applications

A Project Report

submitted in partial fulfillment of the requirements

of

Industrial Artificial Intelligence with cloud computing

by

M Ashfaqulla Sharif, 2010229

B N Jaswanth Kumar, 2010208

S Mahammed Zubair, 2010274

G Praveen Kumar, 2010220

Under the Esteemed Guidance of

Abdul Aziz Md

ACKNOWLEDGEMENT

We would like to take this opportunity to express our deep sense of gratitude to all individuals who helped us directly or indirectly during this thesis work.

Firstly, we would like to thank my supervisor, Md Abdul Aziz, for being a great mentor and the best adviser I could ever have. His advice, encouragement and critics are source of innovative ideas, inspiration and causes behind the successful completion of this dissertation. The confidence shown on me by him was the biggest source of inspiration for me. It has been a privilege working with him from last one year. He always helped me during my thesis and many other aspects related to academics. His talks and lessons not only help in thesis work and other activities of college but also make me a good and responsible professional.

M Ashfaqulla Sharif

ABSTRACT

This paper focuses on a tech-savvy task called Named Entity Recognition (NER) using Natural Language Processing (NLP). NER involves spotting and categorizing things like names of people, places, and organizations in text. We dive into various methods, from traditional ones to the latest cool techniques like spaCy,. We compare these methods using standard datasets and metrics like precision and recall. We also look at how tweaking models for specific topics affects performance. Our findings help people figure out the best ways to make computers understand, analyse and extract useful info from text.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

CHAPTER 1

INTRODUCTION

Problem Statement: Developing an accessible and real-time Named Entity Recognition (NER) and analysis solution in Streamlit, while ensuring user-friendly customization, presents a key challenge.

Problem Definition: Named Entity Recognition (NER) is a natural language processing (NLP) task that involves identifying and classifying entities, such as persons, organizations, locations, dates, and other specific categories, within a given text. The primary goal of NER is to extract meaningful information from unstructured text, enabling machines tounderstand and categorize entities for various applications, including information retrieval, question answering, and text summarization.

Expected Outcomes: The outcome of Named Entity Recognition (NER)is the identification and classification of specific entities within a given text, transforming unstructured data into structured information. NER results in a structured representation of entities such as names of people, organizations, locations, dates, and more, providing a clear understanding of the content. And analysis of the structured information as word clouds

CHAPTER 2

LITERATURE SURVEY

CHAPTER 2

LITERATURE SURVEY

2.1. Paper-1

2.1.1.

"Seamless Entity Insight: Streamlined NER and Word Cloud Analysis with Streamlit for User-Friendly Applications"

Brief Introduction of Paper:

Named Entity Recognition (NER) represents a pivotal aspect of Natural Language Processing (NLP), involving the identification and classification of entities within unstructured text, encompassing entities such as names of people, locations, organizations, and more. NER plays a crucial role in enhancing the comprehension of textual data, facilitating applications in information retrieval, question answering, and sentiment analysis. To bring the benefits of NER to a broader audience and enhance the user experience, the integration of NER with Streamlit, a web application framework known for its simplicity and interactivity, is gaining prominence.

The proposed methodology encompasses the system design and modules used in the project. Streamlit serves as the platform for user interaction, facilitating registration and recognition of text input. NLP algorithms, powered by spaCy, parse and summarize the text for NER. BeautifulSoup aids in fetching HTML content from URLs. The chapter outlines the seamless integration of these tools to ensure a cohesive and efficient workflow.

This integration doesn't merely stop at the technical convergence; it extends to the democratization of NER, breaking down barriers for non-technical users to harness the potential of entity recognition. Through a Streamlit-powered interface, users can input text and witness the real-time identification of entities, providing a transparent and engaging

experience. Moreover, this combination empowers developers and data scientists to build and deploy NER applications more efficiently, catering to diverse domains and user needs.

In essence, NER with Streamlit is a synthesis of advanced language processing capabilities and user-centric design. It signifies a shift towards making complex NLP tasks accessible to a wider audience, fostering a more inclusive and interactive landscape for understanding and utilizing the nuances embedded within textual data.

2.1.2. **Techniques used in Paper:**

This section details the implementation process and the results derived from the integration of NER with Streamlit. Leveraging spaCy, BeautifulSoup, Matplotlib, and WordCloud, the project delivers compelling results by seamlessly identifying entities in real-time while offering user-friendly customization. The resource-efficient processing ensures smooth performance within the constraints of web applications, highlighting the versatility and adaptability of the solution

CHAPTER 3

PROPOSED METHODOLOGY

CHAPTER 3

PROPOSED METHODOLOGY

System Design

Registration: Entering the Text Need To Be Summarised, Recognised And Analysed In The Web App Text Widget

Recognition: The Text is parsed and summarised, recognised usng NLP anlorithm.

Analysis: Further, Recognised text is analysed and create a word cloud.

Modules Used

NLP – spaCy

STREAMLIT

BEAUTIFULSOUP

WORDCLOUD

CHAPTER 4

Implementation and Result

CHAPTER 4

IMPLEMENTATION and RESULT

4.1. Results of Named Entity Rcognisation and Analysis.

The integration of Named Entity Recognition (NER) with Streamlit, supported by spaCy, BeautifulSoup, Matplotlib, and WordCloud, yielded promising results. The project aimed to identify entities in real-time while ensuring user-friendly customization. Through Streamlit's intuitive interface, users could input text directly or extract content from URLs with BeautifulSoup's web scraping capabilities. The NLP algorithms, powered by spaCy, efficiently parsed and summarized the text for NER, enhancing textual comprehension.

The project's outcome showcased compelling results, efficiently identifying entities in real-time and accommodating users with varying technical expertise. The system's customization and adaptability allowed fine-tuning for specific domains, ensuring versatility across industries. The resource-efficient processing and adaptability underscored the system's robustness, making it a dynamic tool for entity recognition in various textual contexts.

CHAPTER 5

CONCLUSION

CHAPTER 5

CONCLUSION

ADVANTAGES:

The integration of NER with Streamlit presents a powerful synergy, offering a user-friendly interface that enables real-time interaction and transparency in entity recognition. Streamlit's intuitive design enhances accessibility, allowing users of diverse technical backgrounds to effortlessly engage with the NER application. The scope of integration extends to providing a versatile and accessible solution for entity identification across various textual domains, enhancing the overall experience of working with unstructured text data.

SCOPE:

The scope of integrating Named Entity Recognition (NER) with Streamlit extends to providing a versatile and accessible solution for entity identification in diverse textual domains. This integration aims to empower users with a user-friendly interface, real-time interaction, and customization capabilities, fostering ease of use for individuals across different industries. The scope encompasses applications ranging from legal documents and news articles to healthcare records, offering a dynamic and efficient tool for entity recognition that adapts to specific user needs and enhances the overall experience of working with unstructured text data.

REFERENCES

[1]

Streamlit documentation: https://docs.streamlit.io/

[2]
spaCy documentation: https://spacy.io/

[3]
Matplotlib documentation: https://matplotlib.org/

[4]
WordCloud documentation: https://github.com/amueller/word_cloud

[5]

Video Recording:

https://drive.google.com/file/d/1JwyXCqY53Tu_naBuUNfjjndGpFcabFo7/view?usp=sharing