

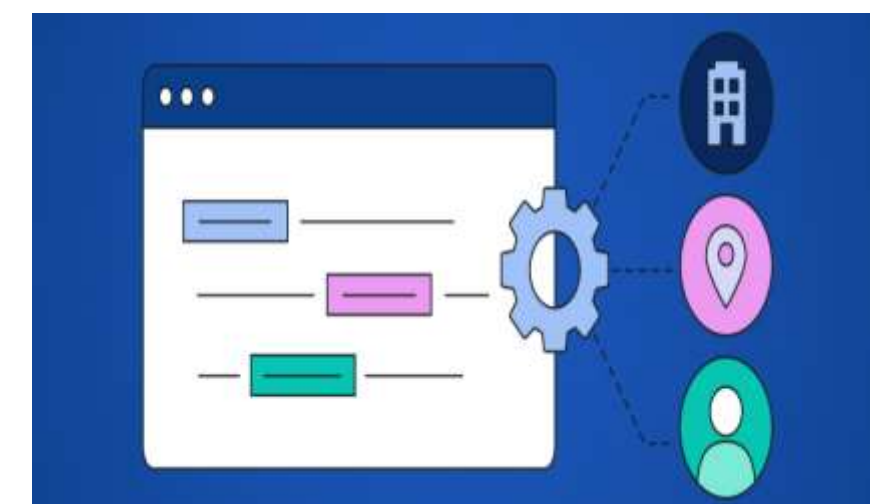
GR-422 Simplified Named Entity Recognition Using Context Free Grammar

Abstract

Named Entity Recognition is a fundamental task in natural language processing. While Spacy is widely used, it occasionally struggles to identify specific person names. CFG is introduced to supplement Spacy's NER capabilities, with the goal of enhancing accuracy in recognizing person names. The scope of this project encompasses the development of CFG rules and their application to a sample text to demonstrate improved NER performance.

Introduction

This project addresses the limitations of Spacy in recognizing specific person names within the context of Named Entity Recognition (NER). By focusing on showcasing Spacy's drawbacks without proposing an integration solution, the project aims to offer a nuanced understanding of challenges and potential refinements in NER processes, introducing the concept of Context-Free Grammar (CFG) as a complementary tool.



Project Goals

The primary goal of this project is to shed light on the limitations of Spacy in recognizing specific person names within the broader landscape of Named Entity Recognition. By strategically showcasing these drawbacks without proposing an integration solution, the project aims to offer a nuanced understanding of challenges within existing NER frameworks. Additionally, the project introduces Context-Free Grammar (CFG) as a potential tool to mitigate Spacy's identified limitations. Through the development of a user-friendly interface, the goal is to facilitate direct user engagement, allowing individuals to input sentences and observe firsthand instances where Spacy encounters challenges in recognizing person names. Ultimately, this project seeks to bridge the gap between theoretical understanding and practical implications, paving the way for informed discussions on refining and advancing entity recognition processes in NLP.

Methods

We analyzed NER shortcomings in Spacy through extensive testing and identify cases where it fails to recognize person names. Context-Free Grammar is employed to define specific grammar rules for person names. We used a diverse dataset for testing, including texts with challenging person name mentions. In this project, text data is first processed to identify 'NAME' entities, which often represent person names, using a Context-Free Grammar.



The CFG rules are applied to the input text to recognize these entities effectively. The results include the successful identification of 'NAME' entities, demonstrating the potential of CFG-based methods in enhancing Named Entity Recognition (NER).

Results

The project's results illuminate the intricacies of Spacy's limitations in recognizing specific person names. Through the user-friendly interface, users can actively engage and witness instances where Spacy falls short in accurate Named Entity Recognition (NER). This direct interaction emphasizes the practical implications of Spacy's drawbacks, fostering a tangible understanding of its challenges in real-world applications. Simultaneously, the introduction of Context-Free Grammar (CFG) demonstrates its effectiveness in addressing the identified limitations, showcasing improved accuracy in recognizing person names.

```
C:\Users\ADMIN\Desktop\NER>python run.py
Hello, I'm Spacy. Enter the text, and I will identify names within it:
Patrick and Grace are studying hard to get good score.
Patrick: name
```

Fig.1 Results of Named Entity Recognition using Spacy.

As users input sentences and observe the divergent outcomes between Spacy and CFG, the results underscore the practical significance of CFG as a complementary tool in refining NER processes. The project's findings not only contribute to a deeper understanding of the challenges within existing NLP frameworks but also pave the way for potential advancements by introducing CFG as a valuable enhancement for enhancing the accuracy of person name recognition. The dual approach of showcasing Spacy's limitations and presenting CFG as a solution provides valuable insights into the nuanced landscape of NER, contributing to ongoing discussions on refining and optimizing entity recognition systems.

```
C:\Users\ADMIN\Desktop\NER>python run.py
Hello, I'm CFG. Enter the text, and I will identify names within it:
Patrick and Grace are studying hard to get good score.
Recognized names:
Patrick
Grace
```

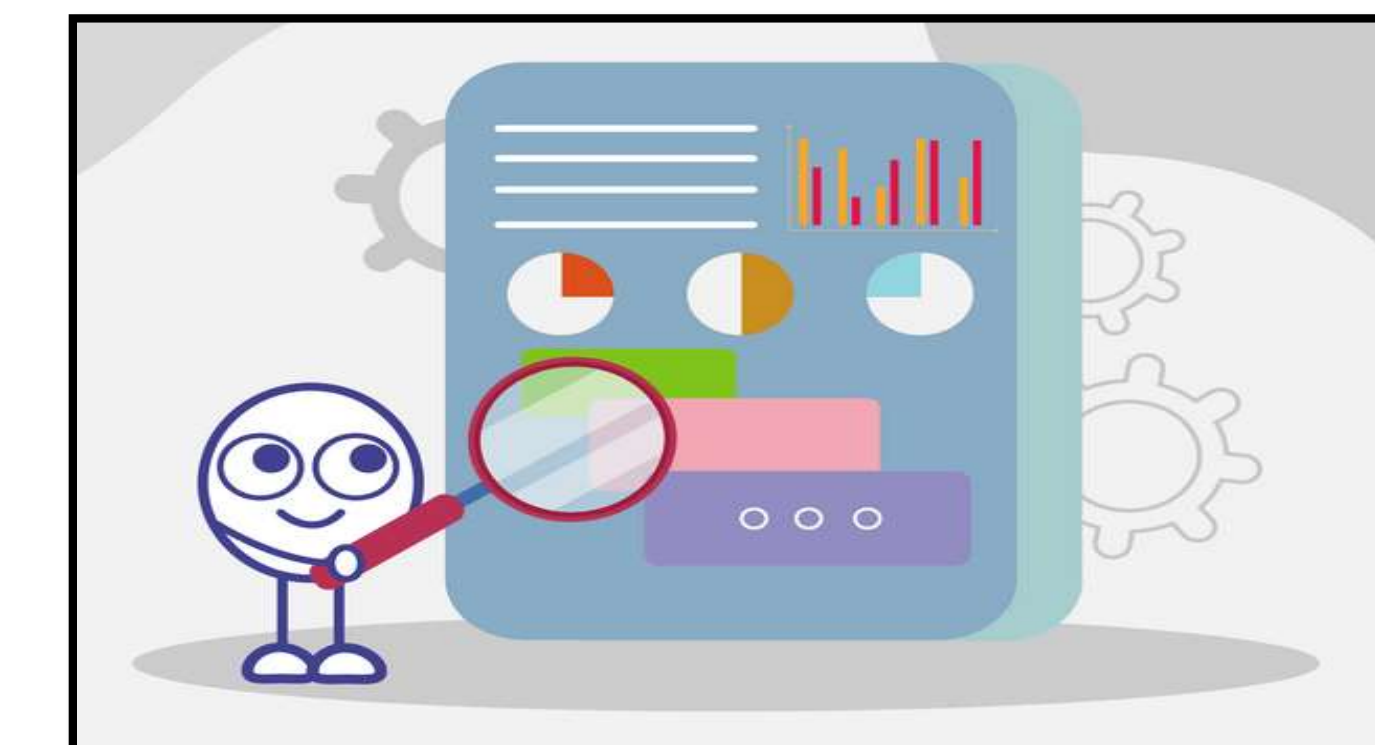
Fig.2 Results of Named Entity Recognition using Context-Free Grammar.



Scan this QR code for Project review. A complete website of the project which helps you to understand it clearly.

Conclusions

This project has successfully addressed the limitations of Spacy in recognizing specific person names within the context of Named Entity Recognition (NER). The exploration of Spacy's performance highlighted its inherent challenges, particularly in accurately identifying certain names. Through the implementation of a user-friendly interface, users engaged in direct interactions that effectively showcased Spacy's drawbacks in real-world scenarios. The introduction of Context-Free Grammar (CFG) as an alternative tool demonstrated promising results, indicating improved accuracy in person name recognition.



In conclusion, the project successfully achieves its objectives by not only identifying the limitations of Spacy but also introducing a viable solution through CFG. The findings encourage further exploration and discussion on enhancing NER systems to address the complexities of person name recognition in natural language text.

Acknowledgments

On behalf of our research team, we extend our sincere gratitude to Dr. Dan Lo, our esteemed professor at Kennesaw State University, for his invaluable guidance and support throughout the collaborative exploration of Named Entity Recognition using Context-Free Grammar. Dr. Lo's expertise in the theory of computation has significantly influenced the trajectory and depth of our research.

We would also like to express our appreciation to Kennesaw State University for providing the necessary academic infrastructure and resources that have been instrumental in the successful completion of our group project. This work is a testament to the collaborative efforts of the academic community, and we are grateful for the support and encouragement received from both Dr. Lo and Kennesaw State University in our collective pursuit of knowledge and research endeavors.

Contact Information

- Sai Chandana : skogant3@students.kennesaw.edu
- Varshini : vyaganti@students.kennesaw.edu
- Vinesh: vyagant1@students.kennesaw.edu
- Pavani : pthullur@students.kennesaw.edu
- Surya Kiran : skatrag1@students.kennesaw.edu

References

- [1] Named Entity Recognition (NER) for Hindi Prince Rana, Sunil Kumar Gupta, Kamlesh Dutta International Journal of Computer Sciences and Engineering Vol.-6, ssue-7, E-ISSN: 2347-2693 July 2018
- [2] D. Nadeau and S. Sekine, "A survey of named entity recognition and classification," *Lingvist. Investig.*, vol. 30, no. 1, pp. 3–26, 2007
- [3] Thomas C. Rindfleisch, Lorraine Tanabe, John N. Weinstein, and Lawrence Hunter. 2000. Edgar: Extraction of drugs, genes and relations from the biomedical literature.