ISSN 2393-9850
GEM — Georgian Educational Mine
Interdisciplinary International Peer-Reviewed Research Journal
St. George's College Aruvithura, Kerala, India

Volume 3, Issue 1, 2023
Pages 69–75
gem.sgcaruvithura.ac.in

ORIGINAL RESEARCH ARTICLE

Comparative Named Entity Recognition Analysis on Biblical Text Using Pre-Trained Models: A Case Study with SpaCy, CoreNLP, and ChatGPT

Anu Thomas[1,*]

[1]Department of Computer Applications, St. George's College Aruvithura, Kerala, India

[*]**Corresponding author:** Anu Thomas, anu_t@sgcaruvithura.ac.in

Received: 7 October 2023 · **Revised:** 20 November 2023 · **Accepted:** 16 December 2023 · **Published:** 18 December 2023

Keywords: Named Entity Recognition; ChatGPT;SpaCy;Stanford CoreNLP; Religious Text;Biblical Text;Pre-Trained Models

Abstract

In the landscape of Natural Language Processing (NLP), Named Entity Recognition (NER) stands as a fundamental technique for identifying and categorizing entities within text. Its outputs serve as foundational elements in the development of diverse applications, including question answering systems, information retrieval, entity linking, and the construction of comprehensive knowledge graphs etc. This paper undertakes a comparative analysis of the behavior exhibited by three prominent Named Entity Recognition tools—SpaCy, Stanford CoreNLP, and ChatGPT—when applied to Biblical text. The study focuses on unveiling nuanced distinctions in the tools' behavior patterns, shedding light on their adaptability to the unique linguistic and semantic features of religious texts. As no dedicated labelled dataset for Biblical entities exists, this research aims to provide valuable insights into the behavior of NER tools in this specialized context.

1. INTRODUCTION

Named Entity Recognition (NER) stands at the forefront of Natural Language Processing (NLP), serving as a foundational task that involves identifying and classifying entities within textual data. These entities can encompass a broad range, including persons, locations, organizations, dates, and more. NER plays a crucial role in numerous NLP applications, such as information retrieval[Mahalakshmi [2015]], aspect -based sentiment analysis [Lee and Kim [2023]], gene identification process, search queries, relation extraction [Thomas and Sivanesan [2022]], event extraction [Tsujimura et al. [2023]], machine translation [Marton and Zitouni [2014]], text summarization [Rohil and Magotra [2022]], and question answering systems [Nadapana and Kommanti [2022]] thereby enhancing the overall understanding and utility of textual data. Furthermore, NER models have emerged as pivotal tools with diverse applications across numerous domains, including healthcare [Tarcar et al. [2019]], judicial proceedings, aviation text, and beyond [Thomas and Sangeetha [2019], Thomas and Sivanesan [2022]]. In the realm of healthcare, NER models play a crucial role in extracting and categorizing entities like diseases, treatments, and patient information from vast volumes of medical texts. They aid in organizing and analyzing medical records, contributing to more efficient diagnoses and treatment plans. Similarly, in the judicial sector, NER models assist in identifying and categorizing entities such as legal statutes, entities involved in legal proceedings, and pertinent case details from legal documents [Thomas and Sangeetha [2021]].

Extracting named entities from Biblical text holds significant value across several domains. In theological studies, it aids in cataloging and analyzing individuals, locations, and events mentioned in religious scriptures, contributing to deeper textual understanding and scholarly research. Moreover, in linguistics and language studies, identifying and categorizing entities within Biblical texts supports linguistic analysis, semantic research, and the exploration of ancient language usage. Additionally, within the realm of literature and cultural studies, extracting named entities from these texts enables scholars to trace cultural references, historical contexts, and narrative themes, fostering insights into societal norms and beliefs prevalent during those times.

While Named Entity Recognition (NER) models have been extensively applied across various domains, their adaptation to specialized contexts, such as religious texts like the Bible, presents unique challenges. The Bible, as a foundational religious and literary text, contains linguistic nuances, archaic language, and cultural references that set it apart from more contemporary corpora. Applying NER to Biblical text requires models that not only understand general language constructs but also grasp the intricacies of ancient expressions and religious terminology. Traditional approaches to NER, often reliant on labelled datasets, are challenging in this context due to the scarcity of annotated Biblical data.

The methodology mentioned in the paper involves a comparative analysis of three Named Entity Recognition tools—SpaCy, Stanford CoreNLP, and ChatGPT—when applied to the Biblical text. This study aims to uncover the behavioral nuances and differences in how these tools identify and categorize entities in religious texts, particularly in the absence of a dedicated labeled dataset tailored to Biblical entities.

The key aspects of the methodology include:

- Selection of NER Tools: The paper focuses on SpaCy for its efficiency and accuracy, CoreNLP for its comprehensive linguistic analysis, and ChatGPT for its generative model approach.
- Analysis of Biblical Text: The study applies each of the NER tools to the Biblical text to evaluate their ability to recognize entities like persons, locations, organizations, and other relevant categories within the linguistic and semantic context of the scripture.
- Performance Comparison: The tools' performance is compared to identify how well they adapt to the unique challenges posed by the Biblical text, which includes archaic language, cultural references, and religious terminology.
- Evaluation of Adaptation Challenges: The paper also considers the challenges of applying traditional NER approaches to religious texts due to the scarcity of annotated Biblical data.

This comparative analysis provides insights into how NER tools function in a specialized con-

text and helps in understanding their strengths and areas for improvement when dealing with the complexities of ancient religious texts The paper is organized as follows: Section 2 outlines the existing state-of-the-art NER tools and ChatGPT. Section 3 presents the comparative analysis of SpaCy, CoreNLP, and ChatGPT, highlighting their adaptability to Biblical language and performance in entity recognition. Finally, Section 4 offers a conclusion, summarizing key findings and providing directions for future research in the evolving landscape of NER in religious texts.

2. NER TOOLS

The choice of these particular NER tools stems from their widespread usage, diverse capabilities, and varied approaches to NLP tasks. SpaCy is known for its efficiency and accuracy in entity recognition, CoreNLP provides comprehensive linguistic analysis, and ChatGPT, based on a generative language model, offers a unique perspective. The following section explains more on each of these NLP tools.

2.1 SpaCy

With its NER module, SpaCy [Honnibal and Montani [2017]] is a flexible solution for natural language processing tasks. Its powerful machine learning models enable SpaCy's NER to identify a wide range of entities, including people, organizations, places, date expressions, and more. The models are pre-trained on large corpora and can recognize entities in multiple languages.

2.2 Stanford CoreNLP

Stanford CoreNLP [Manning et al. [2014]] is a comprehensive natural language processing toolkit developed by Stanford University. It offers a wide range of linguistic analysis tools, making it a valuable resource for various natural language processing tasks. One of its key features is the NER, which identifies and classifies entities such as persons, organizations, and locations in text. CoreNLP uses state-of-the-art machine learning models to achieve high accuracy in entity recognition, contributing to its effectiveness in information extraction from diverse textual data.

CoreNLP is implemented in Java and can be used through a user-friendly interface or integrated into various applications and programming environments, making it a versatile and widely adopted tool in both research and industry for natural language processing tasks.

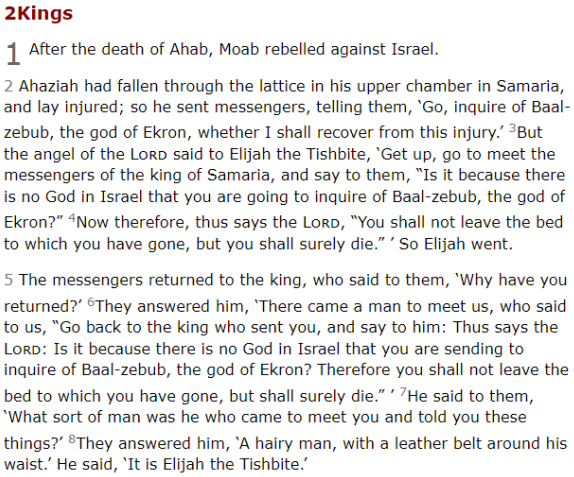
2.3 ChatGPT

ChatGPT [Brown et al. [2020]] is a language model designed for natural language understanding and generation, but it doesn't have specialized modules for tasks like NER. Even though ChatGPT doesn't have inherent NER capabilities, it can be leveraged for tasks related to NER in a conversational context. We can interact with ChatGPT by providing examples or queries related to NER, and it will attempt to understand and generate relevant responses.

3. EXPERIMENTALL CONDITIONS AND RESULTS

In this study, we specifically examine the working of SpaCy, Stanford CoreNLP and ChatGPT on biblical text. As there is no named entity-labeled dataset on Biblical text, we tested the above tools on the first chapter of 2 Kings in the English Bible. Figure 1 shows text extract from the dataset.

The test data is comprised of 42 sentences with 647 words. The test data is pre-processed to remove unnecessary characters, including numerals. Table 1 shows the count of persons and location entities present in the test data.



2Kings

1 After the death of Ahab, Moab rebelled against Israel.

2 Ahaziah had fallen through the lattice in his upper chamber in Samaria, and lay injured; so he sent messengers, telling them, 'Go, inquire of Baal-zebub, the god of Ekron, whether I shall recover from this injury.' 3But the angel of the LORD said to Elijah the Tishbite, 'Get up, go to meet the messengers of the king of Samaria, and say to them, "Is it because there is no God in Israel that you are going to inquire of Baal-zebub, the god of Ekron?" 4Now therefore, thus says the LORD, "You shall not leave the bed to which you have gone, but you shall surely die." ' So Elijah went.

5 The messengers returned to the king, who said to them, 'Why have you returned?' 6They answered him, 'There came a man to meet us, who said to us, "Go back to the king who sent you, and say to him: Thus says the LORD: Is it because there is no God in Israel that you are sending to inquire of Baal-zebub, the god of Ekron? Therefore you shall not leave the bed to which you have gone, but shall surely die." ' 7He said to them, 'What sort of man was he who came to meet you and told you these things?' 8They answered him, 'A hairy man, with a leather belt around his waist.' He said, 'It is Elijah the Tishbite.'

Figure 1. Text extract from the dataset

Table 1. Dataset Statistics

	Person	Location	Misc
Total	29	13	29
Unique	6	5	4

4. EVALUATION

Assessing the performance and accuracy of Named Entity Recognition systems is crucial, and various measures are employed for this purpose. Key metrics in NER evaluation include precision, recall, and FScore. Precision quantifies the ratio of accurately identified entities to the total number predicted by the system, while recall measures the proportion of correctly identified entities among the total entities in the reference data. Fscore is the harmonic mean between Precision and Recall. Table 2 shows the performance of the model in identifying each of the entities in terms of precision, and recall. Figure 2 shows the FScore values given by each NER tool for Person and Location entities.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \tag{1}$$

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{2}$$

$$FScore = \frac{2 * Precision * Recall}{Precision + Recall} \tag{3}$$

4.1 Result Analysis

Based on the results, different NER systems exhibited varying levels of performance across these measures when applied to biblical text: ChatGPT performed exceptionally well in recognizing person

Table 2. Experimental Results

	SpaCy		**CoreNLP**		**ChatGPT**	
	Precision	**Recall**	**Precision**	**Recall**	**Precision**	**Recall**
Person	40	67	44	67	67	100
Location	50	80	80	80	100	60

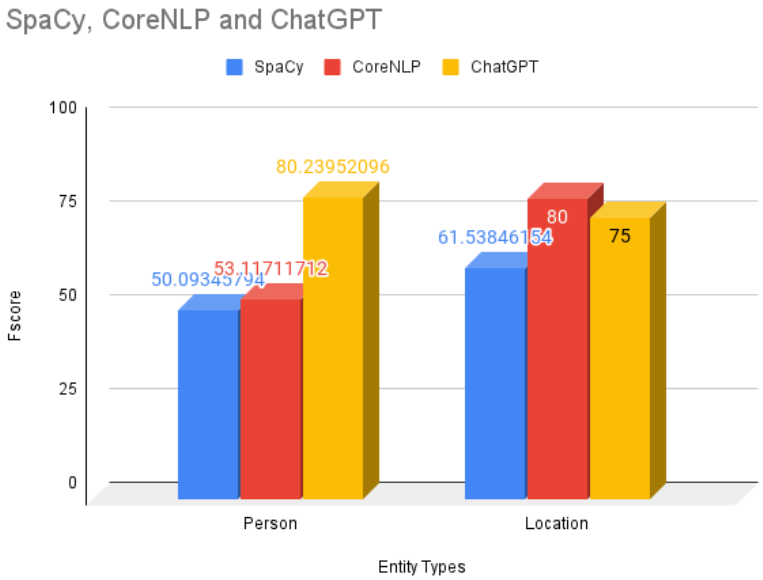


Figure 2. F Score values of each entity type

entities with perfect recall (100%), indicating it was able to identify all person entities in the sample. However, its recall for location entities was lower (60%), which might imply it missed several location entities the annotators expected to find. SpaCy and CoreNLP showed different levels of precision and recall for person and location entities, indicating varying abilities to accurately identify and classify these entities.

Impact of Precision on Effectiveness: High Precision means that when the NER system identifies an entity, there is a high chance that it is correct. In biblical texts, this ensures that entities such as historical figures or places are accurately recognized, which is crucial for scholarly accuracy. However, high precision can come at the expense of recall; the system may miss some entities, which can be problematic if complete information extraction is required. Low Precision indicates the system identifies many incorrect entities. This could lead to skewed analysis, as the extracted information might include irrelevant or wrong data, which can be particularly misleading when studying texts where context and accuracy are paramount.

Impact of Recall on Effectiveness: High Recall means the system identifies most true entities in the text. For biblical texts, this could be essential for thematic studies or constructing genealogies, where missing out on a single entity could lead to incomplete or incorrect conclusions. Low Recall results in many true entities being overlooked. Researchers might miss out on important connections or information, leading to an incomplete picture of the text.

Impact of FScore on Effectiveness: The FScore represents a balance between precision and recall, providing a single measure that weighs how well the NER system is performing overall. A high FScore means that the system has a suitable trade-off between not missing entities (recall) and not incorrectly tagging non-entities (precision). For specialists in biblical texts, a high FScore would be indicative of a robust NER system that can be reliably used for extracting entities without needing to make significant trade-offs between recall and precision—which is often the goal for complex analytical tasks. Different applications have different requirements depending on the intended use:

- For applications such as creating a searchable index or database, high recall might be preferred to ensure no entity is left out.
- For tasks like semantic analysis or literature studies, high precision might be crucial to ensure the analysis is based on accurately identified entities.
- For balanced needs, such as developing a recommendation system or supporting academic research, a high FScore would signify that the NER system is quite effective overall.

Ultimately, the performance measures reflect the strengths and weaknesses of NER systems in handling the complexities and nuances of ancient texts like the Bible. Depending on the specific application requirements and the desired outcome (whether that be precision-oriented, recalloriented, or a balance of both), researchers would need to choose the entity recognition system that aligns best with their objectives. In scholarly endeavors and applications where any misidentification could lead to significant errors in interpretation or analysis, such as in precise biblical exegesis or the study of historical interrelations within the text, researchers might place a greater emphasis on precision. However, for qualitative studies seeking to understand the broader narratives or thematic elements in religious texts, systems with a higher recall may be more beneficial, as they ensure fewer entities are missed, albeit with some risk of including incorrect identifications.

5. CONCLUSION

Our study embarked on a journey to evaluate the capabilities of Named Entity Recognition technologies in interpreting religious manuscripts, with a focus on the Bible. We found the Bible's archaic lexicon and cultural idioms pose substantial hurdles for contemporary NER systems. This research provided a detailed examination of how SpaCy, Stanford CoreNLP, and ChatGPT perform in identifying entities within such a complex text.

The findings presented in this analysis offer valuable contributions to the domain of NER, particularly when dealing with specialized literature. They not only enhance our understanding of the strengths and limitations of current NER models in processing theological and ancient texts but also lay the groundwork for the evolution of Natural Language Processing methodologies. Advances from this study could pave the way for more sophisticated interpretive algorithms capable of navigating the intricacies of historical and religious narratives.

REFERENCES

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

Matthew Honnibal and Ines Montani. spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing. *To appear*, 7(1):411–420, 2017.

Sun-Kyung Lee and Jong-Hwan Kim. Sener: Sentiment element named entity recognition for aspect-based sentiment analysis. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023.

G Suryanarayanan Mahalakshmi. Content-based information retrieval by named entity recognition and verb semantic role labelling. *Journal of universal computer science*, 21(13):1830, 2015.

Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60, 2014.

Yuval Marton and Imed Zitouni. Transliteration normalization for information extraction and machine translation. *Journal of King Saud University-Computer and Information Sciences*, 26(4):379–387, 2014.

Vasuki Nadapana and Hima Bindu Kommanti. Investigating the role of named entity recognition in question answering models. In *2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT)*, pages 1–7. IEEE, 2022.

Mukesh Kumar Rohil and Varun Magotra. An exploratory study of automatic text summarization in biomedical and healthcare domain. *Healthcare Analytics*, 2:100058, 2022.

Amogh Kamat Tarcar, Aashis Tiwari, Vineet Naique Dhaimodker, Penjo Rebelo, Rahul Desai, and Dattaraj Rao. Healthcare ner models using language model pretraining. *arXiv preprint arXiv:1910.11241*, 2019.

Anu Thomas and S Sangeetha. An innovative hybrid approach for extracting named entities from unstructured text data. *Computational Intelligence*, 35(4):799–826, 2019.

Anu Thomas and Sivanesan Sangeetha. Semi-supervised, knowledge-integrated pattern learning approach for fact extraction from judicial text. *Expert Systems*, 38(3):e12656, 2021.

Anu Thomas and Sangeetha Sivanesan. An adaptable, high-performance relation extraction system for complex sentences. *Knowledge-Based Systems*, 251:108956, 2022.

Tomoki Tsujimura, Koshi Yamada, Ryuki Ida, Makoto Miwa, and Yutaka Sasaki. Contextualized medication event extraction with striding ner and multi-turn qa. *Journal of Biomedical Informatics*, 144:104416, 2023.