Comparative Analysis of Named Entity Recognition in the Dungeons and Dragons Domain

Gayashan Weerasundara & Nisansa de Silva

Department of Computer Science & Engineering, University of Moratuwa, Sri Lanka.

Abstract

Some Natural Language Processing (NLP) tasks that are in the sufficiently solved state for general domain English still struggle to attain the same level of performance in specific domains. Named Entity Recognition (NER), which aims to find and categorize entities in text is such a task met with difficulties in adapting to domain specificity. This paper compares the performance of 10 NER models on 7 adventure books from the Dungeons and Dragons (D&D) domain which is a subdomain of fantasy literature. The aim is to identify challenges and opportunities for improving NER in fantasy literature. Fantasy literature, being rich and diverse in vocabulary, poses considerable challenges for conventional NER. In this study, we use open-source Large Language Models (LLM) to annotate the named entities and character names in each number of official D&D books and evaluate the precision and distribution of each model. The paper aims to identify the challenges and opportunities for improving NER in fantasy literature. Our results show that even in the off-the-shelf configuration, Flair, Trankit, and Spacy achieve better results for identifying named entities in the D&D domain compared to their peers.

Dungeons & Dragons

- Open-ended, pen-and-paper, table-top role playing game (RPG) in circulation since 1974 [1]
- *Limitless* game play possibilities. Has predefined rules
- Setting: lore, artifacts, rules, and species.
- Completely made up names, locations, items and even languages

Related Work

Pre-trained and NER-specific Models

- BERT [2], RoBERTa [3], ELECTRA [4], and XLM-RoBERTa [5] as foundational pre-trained models tailored for NER.
- BERT's introduction of the transformer architecture revolutionized NLP.
- ELECTRA's discriminative task offers efficiency in NER.
- WikiNEuRal [6], RoBERTaNER [3], and BERT-CRF [7] are specifically designed for NER tasks.
- BERT-CRF's combination of BERT with a CRF layer captures transition probabilities between entities.

Toolkits, Libraries, and Domain-specific Models

- Spacy [8] and Trankit [9] as comprehensive NLP tools with NER capabilities.
- Spacy's renowned speed and efficiency in NER tasks.
- Trankit's support for over 90 languages, making it ideal for multilingual NER.
- Flair's [10] use of contextual string embeddings offers state-of-the-art NER performance.
- StanfordAIMI's [11] potential focus on medical NER tasks, emphasizing domain-specialized models in NER research.

D&D Adventure Books

Book	Words	Topics
Lost mine of Phandelver	45947	29
Dragon Queen	74243	45
Rise of Tiamat	80065	48
Curse of Strahd	154519	62
Tomb of Annihilation	148605	35
Candlekeep Mysteries	141104	106
Wild Beyond the Witchlight	184135	60

Table 1. Result comparison between LLMs

Figure 1. Adventure sourcebooks.

Data Set Statistics

Chapter	Topic	Paragraph	Word Count
Introduction: Into the Feywild	Adventure Summary	The main antagonists of this story are three hags...	131
		One of the many novelties of this adventure is that...	43
		The characters are drawn into the adventure by one of two adventure hooks. You choose...	31
		Chapter 1 describes the Witchlight Carnival...	40
		...	...
	Running the Adventure	The Monster Manual contains stat blocks for most of the creatures encountered in this...	72
		Spells and equipment mentioned in the adventure are described in the Player's Handbook...	31

Table 2. Content hierarchy in a book

Model	PER	LOC	ORG	MSC	All
XLM-RoBERTa [5]	16	0	3	4	23
StanfordAIMI [11]	0	0	1	18	19
ELECTRA [4]	10	0	1	10	21
WikiNEuRal [6]	23	4	6	1	34
BERT [2]	9	1	1	0	11
RoBERTaNER [3]	1	0	0	17	18
BERT-CRF [7]	12	0	0	0	12
Flair [10]	28	14	6	4	54
Spacy [8]	21	11	7	18	57
Trankit [12]	25	15	2	2	44

Table 3. Statistics for the adventure book Candlekeep Mysteries. The NER tags are as follows, Person: PER, Location: LOC, Organization: ORG, and Miscellaneous: MSC

Book	Bloom		Dolly		OpenLLaMA		Total Unique Entities
	Count	Recall	Count	Recall	Count	Recall	
Lost Mine of Phandelver	21	0.47	32	0.73	40	0.91	44
Hoard of the Dragon Queen	58	0.89	62	0.95	60	0.92	65
Rise of Tiamat	54	0.88	57	0.93	53	0.87	61
Curse Of Strahd	92	0.90	96	0.94	101	0.99	102
Tomb of Annihilation	101	0.80	99	0.79	112	0.89	126
Candle keep Mysteries	60	0.87	61	0.88	64	0.93	69
The Wild Beyond Witch Light	66	0.84	67	0.85	71	0.89	79

Table 4. Result comparison between LLMs

Dataset Derivation Process

```
Input: Books;
Output: Named entities;
foreach book do
    segments ← divideIntoSegments(book);
    foreach segment in segments do
        paragraphs ←
          divideIntoParagraphs(segment);
        foreach paragraph in paragraphs
          do
            foreach LLM in LLMs do
                prompt ←
                  createPrompt(paragraph);
                namedEntities ←
                  LLM(prompt);
                processNamedEntities(namedEntities);
            end
        end
    end
end
removeDuplicates(namedEntities);
```

Figure 2. Named Entity Recognition using Multiple LLMs.

Please identify and list all named entities in the following text using the BIO (beginning-inside-outside) scheme:		
"The traveling extravaganza known as the Witchlight Carnival visits your world once every eight years. You have a dim memory of sneaking into the carnival as a child without paying... ...pair of elves named Mister Witch and Mister Lightwere decidedly unhelpful."		
B-Organization:	Witchlight	Carnival
I-Person:	Mister	Witch
I-Person:	Mister	Light

Figure 3. Process of Annotation.

JSON object


```
{
    "book": "Candlekeep␣Mysteries",

    "chapter": 1,
    "text": "The␣Book␣of␣Inner␣
   Alchemy␣is␣one␣of␣Candlekeeps
   ...",
    "entities": [
        {
            ...
        },
        {
            "entity": "B-Location",

            "score": 0.9659823,
            "index": 8,
            "word": "Candlekeep",
            "start": 42,
            "end": 51
        },
    ]
}
```


Figure 4. Annotated output.

Results

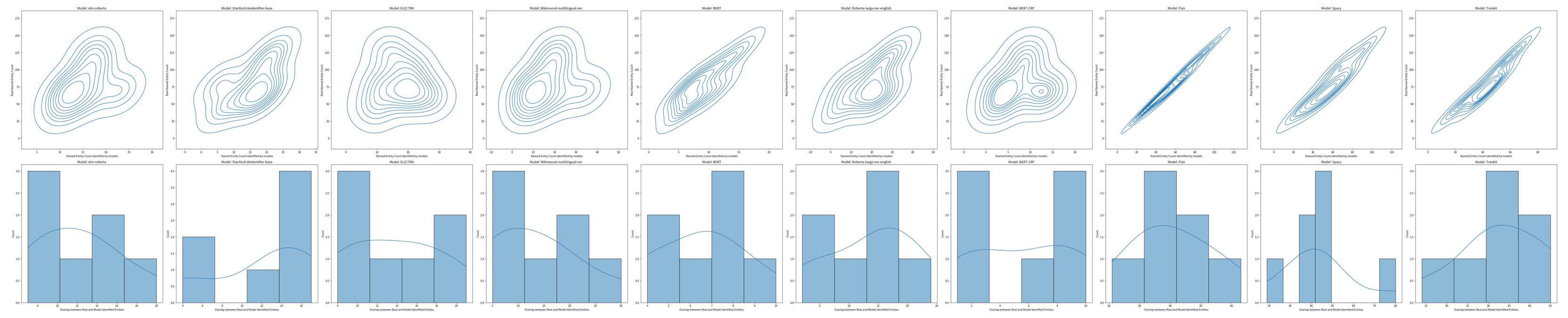

Figure 5. Distribution plot for each model

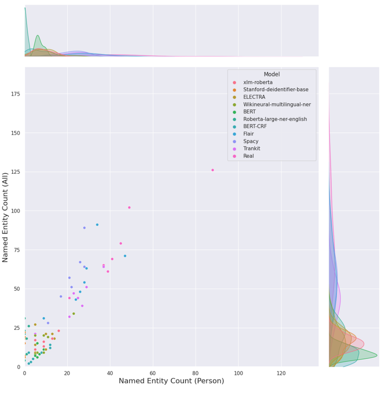

Figure 6. Models

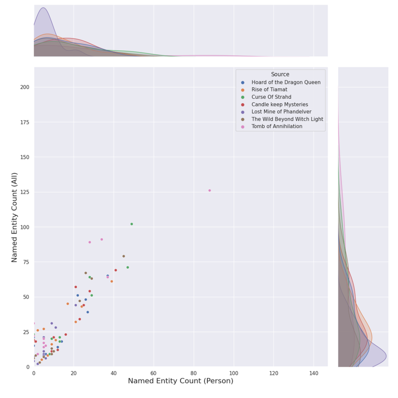

Figure 7. Adventure sourcebooks.

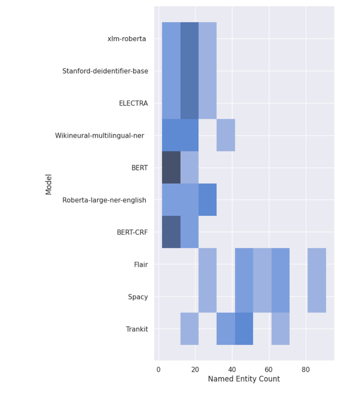

Figure 8. Density plot for each model

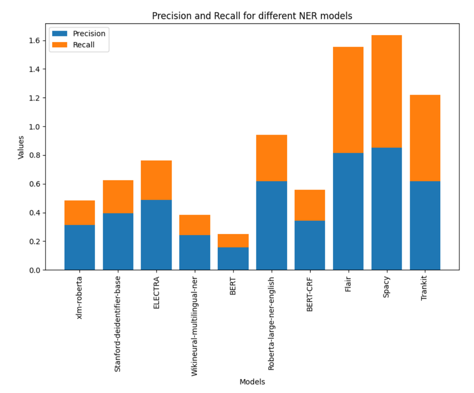

Figure 9. Precision graph for different NER models

Conclusion

We explored NER capabilities in the D&D universe using seven adventure books and ten NER models.

- Models like Flair, Trankit, and Spacy showed strong baseline performance in this domain.
- Our study highlights the potential of general models in specialized domains.
- Our guidelines and dataset set the stage for further domain-specific NER evaluations.
- Future work can use our resources for tasks like text generation or domain-specific summarization.

References

[1] G. Gygax and D. Arneson, *dungeons & dragons*. Tactical Studies Rules Lake Geneva, WI, 1974, vol. 19.

[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[3] J. Baptiste, "roberta-large-ner-english: model fine-tuned from roberta-large for ner task," https://huggingface.co/Jean-Baptiste/roberta-large-ner-english, Jan 2022, (Accessed on 05/10/2023).

[4] K. Clark, M.-T. Luong, Q. V. Le, and C. D. Manning, "Electra: Pre-training text encoders as discriminators rather than generators," in *International Conference on Learning Representations*, 2020.

[5] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised cross-lingual representation learning at scale," *arXiv preprint arXiv:1911.02116*, 2019.

[6] S. Tedeschi, V. Maiorca, N. Campolungo, F. Cecconi, and R. Navigli, "WikiNEuRal: Combined neural and knowledge-based silver data creation for multilingual NER," in *Findings of the Association for Computational Linguistics: EMNLP 2021*, 2021, pp. 2521–2533.

[7] F. Souza, R. Nogueira, and R. Lotufo, "Portuguese named entity recognition using bert-crf," *arXiv preprint arXiv:1909.10649*, 2019.

[8] M. Honnibal and I. Montani, "spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing," *To appear*, 2017.

[9] M. Nguyen and Z. Yu, "Improving named entity recognition in spoken dialog systems by context and speech pattern modeling," in *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Singapore and Online: Association for Computational Linguistics, Jul. 2021, pp. 45–55. [Online]. Available: https://aclanthology.org/2021.sigdial-1.6

[10] A. Akbik, D. Blythe, and R. Vollgraf, "Contextual string embeddings for sequence labeling," in *COLING 2018, 27th International Conference on Computational Linguistics*, 2018, pp. 1638–1649.

[11] P. J. Chambon, C. Wu, J. M. Steinkamp, J. Adleberg, T. S. Cook, and C. P. Langlotz, "Automated deidentification of radiology reports combining transformer and "hide in plain sight" rule-based methods," *Journal of the American Medical Informatics Association*, vol. 30, no. 2, pp. 318–328, 2023.

[12] T. V. Nguyen, T. Nguyen, V. S. Ta, P. Q. N. Minh, V. X. Son, N. M. Hieu, N. P. Thai, V. T. Hung, P. M. Quang, N. T. M. Huyen, and H. Li, "Trankit: A light-weight transformer-based toolkit for multilingual natural language processing," in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*, 2021.