Named Entity Recognition for Thai Historical Data

Nasith Laosen
Department of Digital Technology
Faculty of Science and Technology
Phuket Rajabhat University
Phuket, Thailand
nasith.l@pkru.ac.th

Kanjana Laosen
Andaman Intelligent Tourism and
Service Informatics Center
College of Computing
Prince of Songkla University
Phuket Campus
Phuket, Thailand
kanjana.l@phuket.psu.ac.th

Thummarat Paklao
Department of Mathematics
and Computer Science
Faculty of Science
Chulalongkorn University
Bangkok, Thailand
6373005223@student.chula.ac.th

***Abstract*—Named Entity Recognition (NER) is a fundamental task in Natural Language Processing (NLP), enabling various advanced NLP applications like information extraction, question answering, and text summarization. Thai NER presents unique challenges due to the absence of capitalization, explicit word boundaries, and sentence-ending punctuation. While NER on general domain Thai datasets has been explored, its application to historical text remains an underexplored area. Historical texts contain specialized terminology, necessitating domain-specific NER solutions for optimal performance. This study investigates Thai historical NER, utilizing historical textual data collected from the Wikipedia website. Our goal is to identify suitable word segmentation and NER methods. We evaluate the performance of the Attacut word segmentation algorithm against Deepcut and Newmm. Our findings demonstrate Attacut's superiority, achieving an F1-score of 0.9557. Furthermore, our proposed SBC model (Sentence Transformer + BiLSTM + CRF) outperforms pre-trained LLMs (BERT-th1, XLM-R, WangchanBERTa) in NER, achieving an average F1-score of 0.97. The overall performance of the Attacut algorithm and the SBC model highlights their suitability for developing advanced NLP applications within the Thai historical domain.**

***Index Terms*—named entity recognition, Thai language, historical data**

I. Introduction

Named Entity Recognition (NER) is a subtask within Natural Language Processing (NLP) that focuses on identifying and classifying named entities within unstructured text data. It involves pinpointing specific text data spans and assigning them corresponding categories such as person names, organizations, locations, data/time expressions, quantities, monetary values, and percentages [1]–[3]. NER plays a crucial role in enabling higher-level NLP applications such as information extraction, question answering, chatbots, knowledge graph construction, and text summarization [4]–[6]. Errors or inaccuracies in the NER process directly propagate to these high-level applications, resulting in a cascading effect on their overall performance. This dependence highlights the need for careful and meticulous execution of the NER process.

NER in the Thai language presents unique challenges stemming from its distinct linguistic characteristics. Firstly, the absence of capitalization, which is a key identifier of named entities in languages like English, significantly complicates the identification of proper nouns in Thai text. Secondly, Thai writing lacks explicit word boundaries, unlike languages that rely on spaces or punctuation to separate words. This ambiguity in word segmentation makes it difficult to accurately identify the start and end points of named entities within a sentence. Furthermore, the lack of sentence-ending punctuation such as periods adds another layer of complexity by hindering the ability to determine the exact scope of named entities within the context of a sentence. Finally, the shared character sets used for both proper nouns (e.g., names of people, organizations, locations) and common nouns create ambiguity for NER models.

While existing Thai NER techniques have demonstrated success on standard and general domain datasets [7]–[9], NER applied to Thai historical text presents a unique and under-explored research area. This domain is characterized by the presence of specialized terminology that deviates significantly from those encountered in modern Thai text, such as ancient cities, historical royal titles, and outdated personal titles. This work addresses this gap in research by investigating and identifying a combination of a word segmentation algorithm and an NER method that suitable for the Thai historical domain. The findings of this work can be used as a foundation for the development of advanced NLP applications within the domain of Thai historical text.

II. Literature Review

A. Thai Word Segmentation Algorithms

Several word segmentation algorithms have been developed specifically for processing Thai text. Some prominent Thai word segmentation algorithms are described below.

1) Newmm: Newmm [10] is a dictionary-based word segmentation algorithm for the Thai language. It primarily utilizes a maximal matching algorithm constrained by Thai Character Cluster (TCC) boundaries with refined TCC rules. Newmm is recognized for its simplicity and computational efficiency in standard Thai text segmentation tasks.

2) Deepcut: Deepcut [11] is a Thai word segmentation algorithm based on deep learning. Its architecture relies on character embeddings and character type embeddings as input

979-8-3503-8176-4/24/$31.00 ©2024 IEEE

features. A one-dimensional Convolutional Neural Network (CNN) with binary classification is employed to accurately determine the starting characters of words.

3) Attacut: Attacut [12] is another deep neural network-based word segmentation tool for Thai. It leverages CNNs to effectively segment text. Syllable embeddings, combined with character embeddings, serve as input features. Dilated CNN filters are employed to capture the complex and nuanced linguistic patterns of Thai.

B. Thai NER Methods

A range of methods have been explored for effectively performing NER on Thai text. Three key categories of methods are outlined below.

1) Rule-Based Method: This traditional NER method rely on pre-defined patterns and linguistic rules within Thai text to identify named entities [13] [14]. While these methods can be effective in constrained scenarios, they require extensive rule creation and a deep understanding of Thai sentence structures, making them cumbersome and less adaptable.

2) LLM-Based Method: This modern method leverage the power of machine learning techniques to address the complexities of the Thai language. It involves fine-tuning pre-trained large language models (LLMs) like BERT-th1 [15], XLM-RoBERTa (XLM-R) [16], WangchanBERTa [17], and HoogBERTa [18] for NER tasks [19]. This approach leverages the extensive linguistic knowledge already captured within these pre-trained models.

3) BiLSTM-CRF-Based Method: This method utilizes BiLSTM-CRF networks, which are deep learning networks, for Thai NER tasks [7]–[9] [20]. A BiLSTM-CRF network has three primary layers. The first layer is a embedding layer, which transforms input tokens (can be characters or words depending on the model configuration) into dense vector representations. This process captures the semantic and syntactic relationships between words, providing a richer representation for subsequent layers. The second layer is a Bidirectional LSTMs (BiLSTM) layer. This layer is responsible for analyzing and capturing long-range contextual dependencies within the text in both forward and backward directions. The last layer is a Conditional Random Fields (CRF) layer, which predicts sequence of labels based on conditional probability. It leverages the information from neighboring labels within the sequence, taking into account the previously predicted labels to inform the prediction of the current label. This characteristic is particularly beneficial for NER tasks, where the label of a word is often influenced by the labels of its surrounding words. For instance, if the model has already identified a word as the beginning of a location (e.g., a city name), the CRF layer is more likely to predict subsequent words as part of the same location entity.

III. Methodology

This work investigates the combined application of word segmentation and NER methods specifically tailored for Thai historical data. Fig. 1 provides a visual representation of our methodology. The initial step involves collecting historical textual data from the Thai Wikipedia website. The second step focuses on evaluating the efficiency of various word segmentation algorithms for the collected historical data. Three prominent algorithms are compared: Newmm, Deepcut, and Attacut. The most efficient algorithm will be selected (as a tokenizer) for subsequent processing. The third step involves constructing and comparing four NER models. Three of these are LLM-based models fine-tuned from three pre-trained LLMs: BERT-th1, XLM-R, and WangchanBERTa. The fourth model, proposed in this work, is a variation of the BiLSTM-CRF network architecture. It employs the simcse-model-roberta-base-thai sentence transformer model [21] for the embedding layer and utilizes the Dice loss function for the CRF layer. This unique model is denoted as the SBC (Sentence Transformer + BiLSTM + CRF) model. The performance of all four models will be evaluated using F1-score and a 5-fold cross-validation approach. Each step of our methodology will be further elaborated upon in subsequent sections of this work.

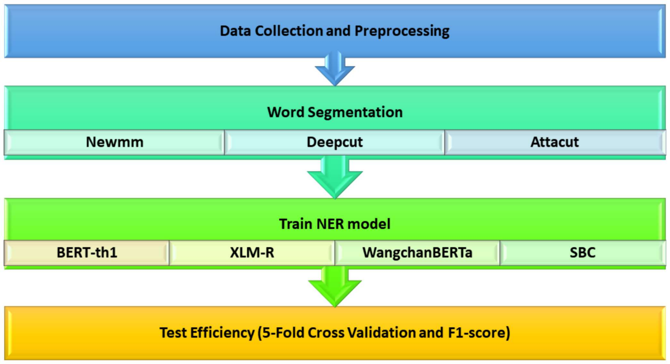
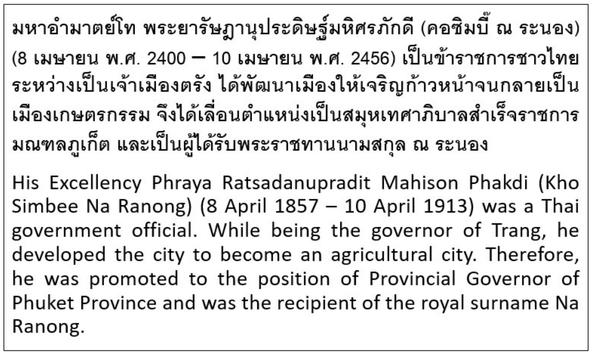

Data Collection and Preprocessing
Word Segmentation
Newmm
Deepcut
Attacut
Train NER model
BERT-th1
XLM-R
WangchanBERTa
SBC
Test Efficiency (5-Fold Cross Validation and F1-score)

Fig. 1. Our research methodology.

IV. Data Collection

We employed a web scraping methodology to collect historical textual data from publicly available sources. Specifically, we leveraged the Thai Wikipedia dataset (official distribution) in conjunction with the Beautiful Soup library [22] for HTML parsing. Our initial focus was on geographically extracting data related to provinces within southern Thailand.



มหาอำมาตย์โท พระยารัษฎานุประดิษฐ์มหิศรภักดี (คอซิมบี๊ ณ ระนอง) (8 เมษายน พ.ศ. 2400 – 10 เมษายน พ.ศ. 2456) เป็นข้าราชการชาวไทย ระหว่างเป็นเจ้าเมืองตรัง ได้พัฒนาเมืองให้เจริญก้าวหน้าจนกลายเป็นเมืองเกษตรกรรม จึงได้เลื่อนตำแหน่งเป็นสมุหเทศาภิบาลสำเร็จราชการมณฑลภูเก็ต และเป็นผู้ได้รับพระราชทานนามสกุล ณ ระนอง

His Excellency Phraya Ratsadanupradit Mahison Phakdi (Kho Simbee Na Ranong) (8 April 1857 – 10 April 1913) was a Thai government official. While being the governor of Trang, he developed the city to become an agricultural city. Therefore, he was promoted to the position of Provincial Governor of Phuket Province and was the recipient of the royal surname Na Ranong.

Fig. 2. An example of the collected textual data.

Subsequently, we expanded the dataset by recursively following relevant internal Wikipedia links, yielding approximately 400 distinct topics. These topics encompassed provincial and sub-provincial administrative divisions (districts, sub-districts), historical persons, landmarks, and key events. We also implemented a preprocessing pipeline that removed extraneous symbols and filtered entries based on minimal content length thresholds. This preprocessing improves quality of Thai textual data [23]. Fig. 2 presents a sample of the collected textual data, along with its corresponding English translation.

V. Word Segmentation

A. Data Labeling

Ten percent (10%) of the preprocessed historical data was allocated to evaluate the performance of the word segmentation algorithms. We adopted a BIO labeling scheme at the character level for word boundaries. This scheme assigns four labels to each character in a sentence:

- 'B' (Begin): The first character of a word.
- 'I' (Inside): An intermediate character within a word.
- 'E' (End): The end character of a word.
- 'O' (Other/Outside): A character that does not belong to any word (e.g., punctuation, space).

The labeling scheme resulted in approximately 107,000 labeled characters. An illustrative example of this labeling scheme is provided in Fig. 3.

B. Experiment Settings and Results

We assessed the performance of the word segmentation algorithms (Newmm, Deepcut, and Attacut) by having them segment text into words and comparing the results to the labeled text. Their performance is presented in Fig. 4 and Fig. 5. We employed two key metrics for evaluation: percentage accuracy and F1-score. The results indicate that Attacut emerged as the most effective word segmentation model for Thai historical data. Fig. 4 depicts the percentage accuracy (at the word level) achieved by each algorithm. Attacut achieved the highest accuracy with 84.58%, followed by Deepcut (83.52%) and Newmm (76.79%).

Fig. 5 delves deeper, presenting the precision, recall, and F1-score for each model. As can be seen form the figure, Attacut outperformed the others, attaining an F1-score of 0.9557. Deepcut and Newmm followed with F1-scores of 0.9495 and 0.9105, respectively. Attacut also demonstrated the highest precision and recall values.

พ	ร	ะ	ย	า	ด	ํ	า	ร	ง	ส	ุ	จ	ร	ิ	ต	ม	ห	ิ	ศ	ร	ภ	ั	ก	ด	ี
B	I	I	I	E	B	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	E

ม	ี	น	า	ม	เ	ด	ิ	ม	ว	่	า	ค	อ	ซ	ู	้	เ	จ	ี	ย	ง	
B	E	B	I	E	B	I	I	E	B	I	E	B	I	I	I	I	I	I	I	I	E	O

Fig. 3. An example of labeled data for word segmentation.

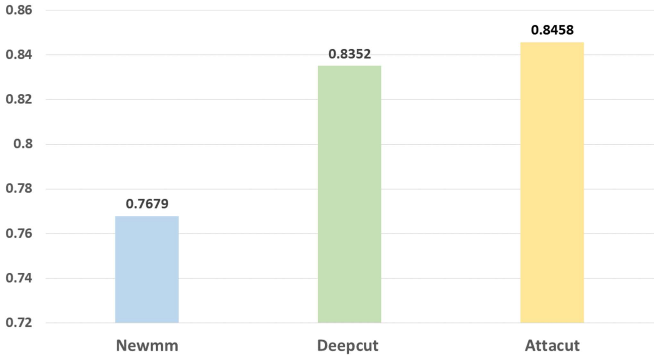


Fig. 4. Accuracy scores (at the word level) of word segmentation algorithms.

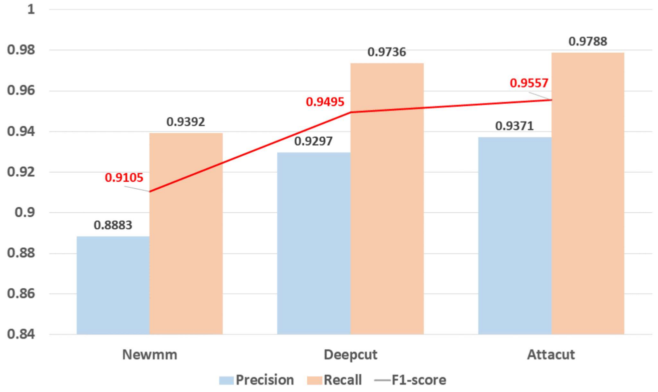


Fig. 5. Precision, recall, and F1 scores of word segmentation algorithms.

As Attacut achieved the most compelling word segmentation results for historical data, we delve into a finer-grained analysis using a confusion matrix focusing on three classes: 'B' (Begin character), 'I' (Inside character), and 'O' (Other/Outside). We disregard class 'E' (End character) since its prediction is inherently dependent on the subsequent 'B' class. Fig. 6 reveals that Attacut excels at correctly predicting the 'O' class, likely due to the clear distinction of spaces within the Thai language. However, the model exhibits some errors in classifying 'B' and 'I' classes. These classes pose inherent complexity due to their dependence on factors like consonant/vowel characters, neighboring words, and overall sentence structure.

VI. Named Entity Recognition

A. Data Labeling

We mainly focused on four primary entity types: Person, Location, Date/Time, and Organization. To prepare the data for NER, we first employed the Attacut algorithm for tokenization. Each token was then assigned a corresponding label using a BIO tagging scheme:

- B-PER (Begin-Person): Identifies the beginning of a person's name.
- PER (Person): Represents person's name.
- B-ORG (Begin-Organization): Identifies the beginning of an organization's name.
- ORG (Organization): Represents organization's name.

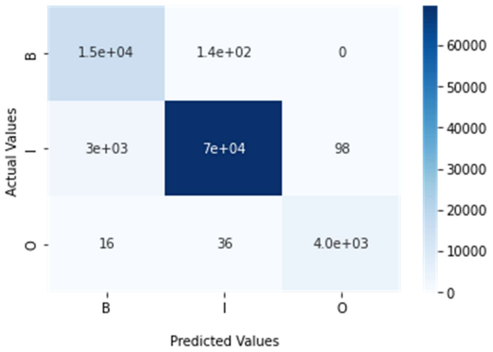


Fig. 6. The confusion matrix of the Attacut algorithm.

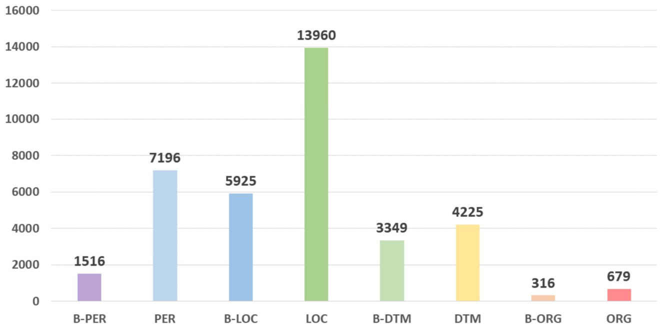
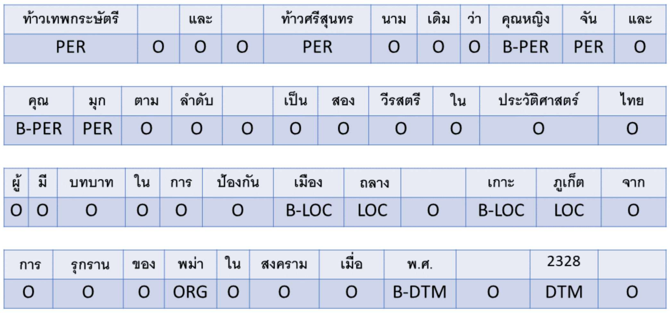


Fig. 7. The distribution of the data labeled for NER.

- B-LOC (Begin-Location): Identifies the beginning of a location name.
- LOC (Location): Represents location's name.
- B-DTM (Begin-Date/Time): Identifies the beginning of a date or time expression.
- DTM (Date/Time): Represents a date or time expression.
- O (Other/Outside): Denotes words that are not part of any named entity.

Fig. 7 illustrates the distribution of labeled tokens across different entity types (excluding the type 'O', which has approximately 218,000 tokens). Fig. 8 shows an example of the data labeled for NER.



ท้าวเทพกระษัตรี		และ		ท้าวศรีสุนทร	นาม	เดิม	ว่า	คุณหญิง	จัน	และ
PER	O	O	O	PER	O	O	O	B-PER	PER	O

คุณ	มุก	ตาม	ลำดับ		เป็น	สอง	วีรสตรี	ใน	ประวัติศาสตร์	ไทย
B-PER	PER	O	O	O	O	O	O	O	O	O

ผู้	มี	บทบาท	ใน	การ	ป้องกัน	เมือง	ถลาง		เกาะ	ภูเก็ต	จาก
O	O	O	O	O	O	B-LOC	LOC	O	B-LOC	LOC	O

การ	รุกราน	ของ	พม่า	ใน	สงคราม	เมื่อ	พ.ศ.		2328	
O	O	O	ORG	O	O	O	B-DTM	O	DTM	O

Fig. 8. An example of the labeled data for NER.

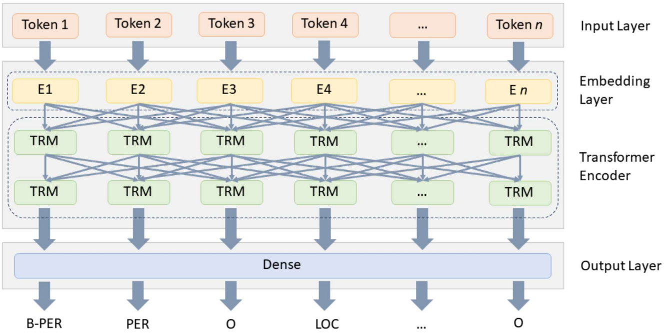


Fig. 9. A general architecture of LLM-based NER models.

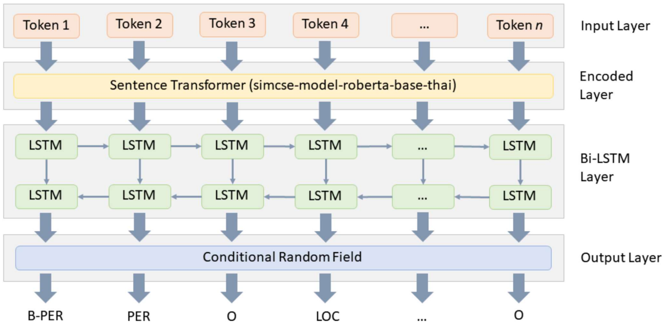


Fig. 10. The architecture of the SBC model.

B. NER Models

As detailed in Sect. III, this work constructed and compared four distinct NER models. The first three models obtained from fine-tuning three pre-trained LLMs, i.e., BERT-th1, XLM-R, and WangchanBERTa, for the NER downstream task. The general architecture of these LLM-based models is depicted in Fig. 9. The fourth model is the SBC model, which is our proposed contribution. Its architecture is presented in Fig. 10.

C. Experiment Settings and Results

The dataset was partitioned into two primary sections: training data (80%) and testing data (20%). To mitigate overfitting during the training process, a 10% validation split was further established from the training data. Subsequently, a 5-fold cross-validation technique was employed to train the model in an iterative fashion across five folds. During the training phase, a learning rate of 1e-4 and a maximum sequence length of 128 were utilized. For all pre-trained models, a batch size of 64 and a maximum of 30 epochs were employed for training.

Since the data is imbalance, we employed the micro-averaged F1-score as a primary metric to assess the efficiency of our NER models. The results, presented in Fig. 11, indicate that the SBC model achieved the highest performance with an average F1-score of 0.97. Additionally, the average F1-scores for XLM-R, BERT-th1, and WangchanBERTa were 0.92, 0.88, and 0.70, respectively.

Delving deeper into the SBC model's performance, we observed consistently high F1-scores exceeding 0.96 in both

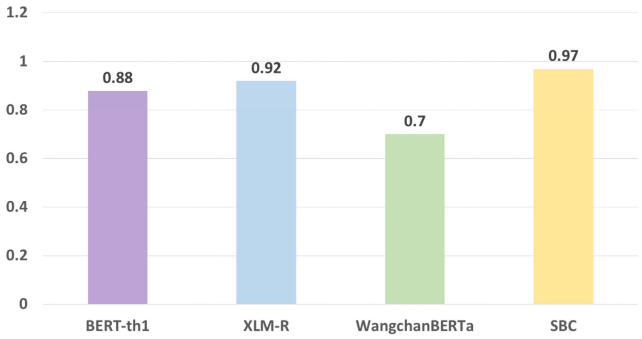


Fig. 11. F1 scores of NER models.

TABLE I
F1 SCORES OF THE SBC MODEL.

Folds	F1 score	
	Validation	Test
1^{st}	0.9695	0.9689
2^{nd}	0.9690	0.9681
3^{rd}	0.9733	0.9680
4^{th}	0.9687	0.9678
5^{th}	0.9870	0.9646
Mean	0.9735	0.9675
std.	0.0078	0.0017

the validation and test sets, as presented in Table I. The average F1-score across the five folds hovers around 0.97, with a standard deviation of approximately 0.0017. These results demonstrate the SBC model's capability of achieving high accuracy and maintaining stability throughout the evaluation process.

In addition to the overall F1-scores, we calculated entity-type-wise F1-scores for each NER model. An entity-type F1-score is derived by computing the F1-score of the constituent classes belonging to that particular entity type. For instance, the F1-score for the Person entity type is obtained by calculating the F1-score of the B-PER and PER classes. The resulting entity-type-wise F1-scores of NER models are shown in Fig. 12. As can be seen from the figure, the SBC model demonstrates superior performance across Person, Location, and Date/Time types. While exhibiting slightly reduced performance for the Organization type, the SBC model still outperforms BERT-th1 and WangchanBERTa. The lower accuracy in the Organization type may be attributed to limited and less diverse Organizational entity examples within the training data. WangchanBERTa, despite exhibiting a moderate overall F1-score, shows notably lower performance on named entity recognition, suggesting its limitations for use with Thai historical data.

VII. Discussion and Conclusion

In this study, we evaluated the performance of word segmentation algorithms and named entity recognition (NER) models on Thai historical textual data collected from the Wikipedia website. Our results showcase the efficacy of the Attacut algorithm for word segmentation and the superiority of the SBC model for the NER task. Attacut demonstrated outstanding performance in word segmentation, achieving the highest percentage accuracy of 84.58% among the evaluated algorithms. This superiority was further affirmed by its F1-score of 0.9557, outperforming the Deepcut and Newmm algorithms.

The SBC model emerged as the top performer in NER, achieving an average F1-score of 0.97, surpassing other evaluated models: BERT-th1, XLM-R, and WangchanBERTa. Consistently high F1-scores across validation and test sets demonstrate the SBC model's accuracy and stability. Additionally, entity-type-wise F1-scores revealed the SBC model's superior performance across the Person, Location, Date/Time, and Other entity types. However, the SBC model exhibited slightly lower accuracy in recognizing Organization entities compared to other entity types, likely due to a limited number of diverse training examples. Expanding the training data for the Organization entity type could improve performance.

Our findings underscore the importance of selecting appropriate models for specific tasks and data domains. The robust performance of the Attacut algorithm and the SBC model signifies their potential in enhancing higher-level NLP applications within the domain of Thai historical text. Our future work will focus on constructing a Thai historical knowledge graph and developing Thai historical content similarity calculations, leveraging the capabilities of the Attacut algorithm and the SBC model.

Acknowledgment

This research is supported by National Research Council of Thailand (NRCT) under the "Artificial Intelligent Platform Development for Metaverse Historical Tourism: AI-MHT" and Super AI Engineer season II program.

References

[1] B. Mohit, *Named Entity Recognition*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, ch. 7, pp. 221–245. [Online]. Available: https://doi.org/10.1007/978-3-642-45358-8-7

[2] A. Roy, "Recent trends in named entity recognition (NER)," 2021. [Online]. Available: https://arxiv.org/abs/2101.11420

[3] X. Liu, H. Chen, and W. Xia, "Overview of named entity recognition," *Journal of Contemporary Educational Research*, vol. 6, no. 5, pp. 65–68, 2022.

[4] T. Noraset, L. Lowphansirikul, and S. Tuarob, "WabiQA: A wikipedia-based Thai question-answering system," *Information Processing Man-agement*, vol. 58, no. 1, p. 102431, 2021.

[5] P. Wongpraomas, C. Soomlek, W. Sirisangtragul, and P. Seresangtakul, "Thai question-answering system using pattern-matching approach," in *Proceedings of the 1st International Conference on Technology Innovation and Its Applications (ICTIIA)*, 2022, pp. 1–5.

[6] S. Chotirat and P. Meesad, "Part-of-speech tagging enhancement to natural language processing for Thai wh-question classification with deep learning," *Heliyon*, vol. 7, no. 10, 2021.

[7] S. Thattinaphanich and S. Prom-on, "Thai named entity recognition using Bi-LSTM-CRF with word and character representation," in *Proceedings of the 4th International Conference on Information Technology (InCIT)*, 2019, pp. 149–154.

[8] V. Sornlertlamvanich and S. Yuenyong, "Thai named entity recognition using BiLSTM-CNN-CRF enhanced by TCC," *IEEE Access*, vol. 10, pp. 53 043–53 052, 2022.

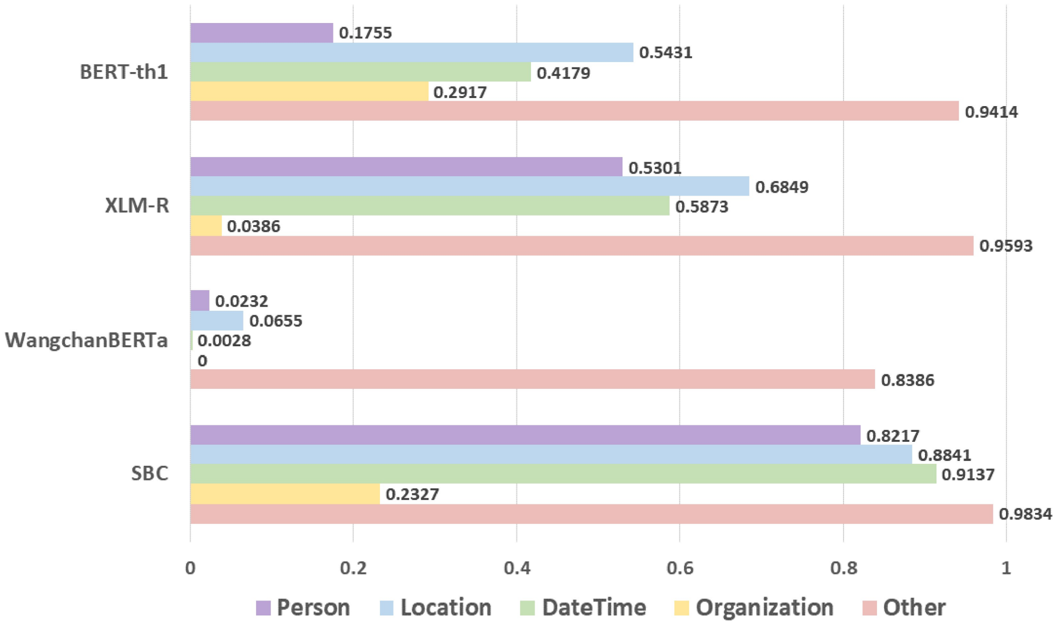


Fig. 12. Entity-type-wise F1-scores of NER models.

[9] A. Pimpisal, T. Simud, N. Sanglerdsinlapachai, N. Surasvadi, and A. Plangprasopchok, "Named entity recognition of Thai documents using CRF with a simple data masking technique," in *Proceedings of the 16th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, 2021, pp. 1–6.

[10] W. Phatthiyaphaibun, K. Chaovavanich, C. Polpanumas, A. Suriyawongkul, L. Lowphansirikul, and P. Chormai, "PyThaiNLP: Thai natural language processing in Python," in *Proceedings of the 3rd Workshop for Natural Language Processing Open Source Software (NLP-OSS 2023)*, 2023, pp. 25–36.

[11] R. Kittinaradorn, T. Achakulvisut, K. Chaovavanich, K. Srithaworn, P. Chormai, C. Kaewkasi, T. Ruangrong, and K. Oparad, "Deepcut: A Thai word tokenization library using deep neural network," Sep. 2019. [Online]. Available: http://doi.org/10.5281/zenodo.3457707

[12] P. Chormai, P. Prasertsom, and A. Rutherford, "Attacut: A fast and accurate neural Thai word segmenter," 2019. [Online]. Available: https://arxiv.org/abs/1911.07056

[13] P. Sutheebanjard and W. Premchaiswadi, "Thai personal named entity extraction without using word segmentation or POS tagging," in *Proceedings of the 8th International Symposium on Natural Language Processing*, 2009, pp. 221–226.

[14] N. Saetiew, T. Achalakul, and S. Prom-on, "Thai person name recognition (PNR) using likelihood probability of tokenized words," in *Proceedings of the 2017 International Electrical Engineering Congress (iEECON)*, 2017, pp. 1–4.

[15] Monsoon-NLP, "BERT-th," 2020. [Online]. Available: https: //huggingface.co/monsoon-nlp/bert-base-thai

[16] 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," 2019. [Online]. Available: https://arxiv.org/abs/1911.02116

[17] L. Lowphansirikul, C. Polpanumas, N. Jantrakulchai, and S. Nutanong, "WangchanBERTa: Pretraining transformer-based Thai language models," 2021. [Online]. Available: https://arxiv.org/ abs/2101.09635

[18] P. Porkaew, P. Boonkwan, and T. Supnithi, "HoogBERTa: Multi-task sequence labeling using Thai pretrained language representation," in *Proceedings of the 16th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, 2021, pp. 1–6.

[19] N. Praechanya and O. Sornil, "Improving thai named entity recognition performance using BERT transformer on deep networks," in *Proceedings of the 6th International Conference on Machine Learning Technologies*, 2021, p. 177–183. [Online]. Available: https://doi.org/10.1145/3468891.3468918

[20] K. Suriyachay, T. Charoenporn, V. Sornlertlamvanich, and N. Kaothanthong, "Enhancement of character-level representation in Bi-LSTM model for Thai NER," *Science and Technology Asia*, vol. 26, no. 2, p. 61–78, Jun. 2021. [Online]. Available: https://ph02.tci-thaijo.org/index.php/SciTechAsia/article/view/230527

[21] T. Gao, X. Yao, and D. Chen, "SimCSE: Simple contrastive learning of sentence embeddings," 2021. [Online]. Available: https: //arxiv.org/abs/2104.08821

[22] L. Richardson, "Beautiful Soup documentation," 2016. [Online]. Available: https://beautiful-soup-4.readthedocs.io/en/latest/

[23] T. Tanantong and M. Parnkow, "A survey of automatic text classification based on Thai social media data," *International Journal of Knowledge and Systems Science (IJKSS)*, vol. 13, no. 1, p. 1–25, 2022. [Online]. Available: https://www.igi-global.com/gateway/ article/312578