BKEE: Pioneering Event Extraction in the Vietnamese Language

Thi-Nhung Nguyen[*♡], **Tien-Bang Tran**[*†], **Trong-Nghia Luu**[†],
Thien Huu Nguyen[‡], **Kiem-Hieu Nguyen**[†]
[♡]VinAI Research
v.nhungnt89@vinai.io

[‡]Department of Computer Science, University of Oregon
thien@cs.uoregon.edu

[†]School of Information and Communication Technology,
Hanoi University of Science and Technology
{bang.tt193988@sis, nghia.lt204888@sis, hieunk@soict}.hust.edu.vn

Abstract

Event Extraction (EE) is a fundamental task in information extraction, aimed at identifying events and their associated arguments within textual data. It holds significant importance in various applications and serves as a catalyst for the development of related tasks. Despite the availability of numerous datasets and methods for event extraction in various languages, there has been a notable absence of a dedicated dataset for the Vietnamese language. To address this limitation, we propose BKEE, a novel event extraction dataset for Vietnamese. BKEE encompasses over 33 distinct event types and 28 different event argument roles, providing a labeled dataset for entity mentions, event mentions, and event arguments on 1066 documents. Additionally, we establish robust baselines for potential downstream tasks on this dataset, facilitating the analysis of challenges and future development prospects in the field of Vietnamese event extraction.

Keywords: event extraction, less-resourced, information extraction, corpus

1. Introduction

Event Extraction stands as a pivotal and challenging endeavor within the realm of Information Extraction. In the context of EE, an event extraction pipeline consists of three main tasks: (1) Entity Mention Detection (EMD): to find words that refer to real-world entities and their types; (2) Event Detection (ED): to find the words (event trigger) that refer to the occurrence of the event and their types; and (3) Event Argument Extraction (EAE): to find entities that are involved in the event and their roles. To better understand these problems consider the example in Figure 1. EE has far-reaching applications in fields including information retrieval (Zhang et al., 2021; Kuhnle et al., 2021), recommendation systems (Gao et al., 2016; Liu et al., 2017) intelligent question answering (Boyd-Graber and Börschinger, 2020; Cao et al., 2020), knowledge graph construction (Wu et al., 2019; Bosselut et al., 2021), and numerous other areas (Liu et al., 2021; Ma et al., 2021).

Due to its important role, Event Extraction has received significant research attention over the past century (Ahn, 2006; Ji and Grishman, 2008; Nguyen et al., 2016, 2021; Veyseh et al., 2021; Veyseh and Nguyen, 2022; Liu et al., 2022). Most of these efforts have focused on resource-rich languages like English and Chinese, as illustrated by datasets used such as MAVEN (Wang et al., 2020), RAMS (Ebner et al., 2020), and WikiEvents (Li et al., 2021), which are only annotated for English. In addition, the growing need for multilingual event extraction systems has given rise to the development of multilingual datasets like ACE 2005 (Hsi et al., 2016), TAC KBP datasets (Mitamura and Liu, 2016, 2017), MINION (Veyseh et al., 2022b), and MEE (Veyseh et al., 2022a). Some initiatives have aimed at languages with fewer resources, such as French (Bittar et al., 2011) and Catalan (Saurı and Badia, 2012). However, it is important to acknowledge the absence of availability of Vietnamese language resources for event extraction. This scarcity significantly restricts research and application opportunities in this domain, further highlighting the gap between rich-resource and low-resource languages in the field of event extraction.

To address this limitation, we propose BKEE, the first event extraction dataset for the Vietnamese language. BKEE fully covers Event Mention Detection (EMD), Event Detection (ED), and Event Argument Extraction (EAE) tasks, with content spanning 11 different domains from news sources. Across the entire dataset, our dataset includes 12 entity types, 8 event types, 33 event sub-types, and 28 argument roles, totaling almost 9,000 event mentions, over 16,000 arguments, and entity mentions, making BKEE a valuable resource for Vietnamese event extraction and related applications. In addi-

[*] Equal contribution

LREC-COLING 2024, pages 2421–2427

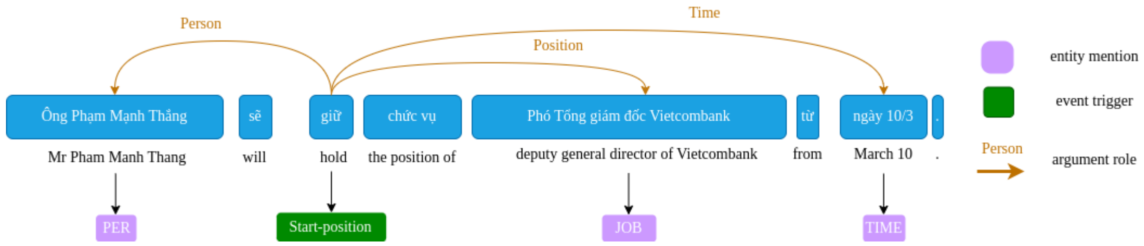

Figure 1: Entity Mention, Event Mention, and Argument Role example.

tion, we conduct extensive experiments on tasks derived from our dataset, establishing strong baselines and offering insights into specific Vietnamese EE challenges, useful for future works. Results highlight the struggles of current state-of-the-art EE models on BKEE, especially with overlapping contexts, complex event and entity mentions, and issues related to word-tokenization errors.

2. Data Construction

2.1. Data Preparation

We obtain the source data from news articles published on a major Vietnamese news media source *BaoMoi* during the period from 2018 to 2020. To ensure diversity in topics and focus on event-related data, we collect articles from 11 different domains, including entertainment, transportation, business, law, society, technology, demonstrations, elections, the military, startup organizations, and sports. After removing duplicate content, the articles are sentence tokenized using *VnCoreNLP* (Vu et al., 2018). As a result, we obtain a total of *1066* documents, *21318* sentences with a corresponding length of about *17* words/sentence. Subsequently, they are annotated across all three subtasks: Entity Mention Detection (EMD), Event Detection (ED), and Event Argument Extraction (EAE).

2.2. Data Annotation

We follow the entity, event, argument definitions, and labeling guidelines from the widely used ACE 2005 dataset (Walker et al., 2006) to leverage its well-structured documentation and maintain consistency with prior works in Event Extraction. However, due to limited resources, we only annotate entities that are directly related to events, that is, only entity mentions in sentences containing event mentions are labeled. Consequently, our dataset comprises all 8 event types, 33 event subtypes, 12 entity types, and 28 distinct argument roles. Tables 2 and 1 illustrate some examples of event types, event subtypes, and argument roles in BKEE.

To ensure label quality, we select experienced native speakers for tasks nearly similar tasks to annotation (e.g., Named-entity recognition). Initially, they are provided with annotation guidelines in Vietnamese, which are constructed based on ACE 2005 guidelines. Subsequently, they undergo an annotation verification process, which includes entity mentions, event mentions, and argument roles extraction. Annotators achieving accuracy rates above 95% in test cases then progress to the official labeling phase for our data. A total of 3 labelers participated throughout our EE project.

To minimize the complexity of labeling processes, our labeling process follows a sequential approach, where we annotate event mentions, entity mentions, and event arguments in that order. In addition, we divide documents into sentences, and then each sentence is annotated separately for EE tasks to reduce annotator overload in long documents. To evaluate the quality of our annotations, we employ a two-stage process. In the first stage, 10% of the documents are co-annotated by multiple annotators to assess agreement scores. Note that, in this co-annotated stage, annotators work independently to label the data. To quantify agreement scores, we utilize Krippendorff's alpha (Krippendorff, 2011)

https://baomoi.com/

Event Subtypes	Argument Roles
End-position	Time, Place, Person, Entity, Position
Transport	Time, Vehicle, Destination, Origin, Agentm, Artifact
Meet	Entity, Place, Time

Table 1: Example of argument roles corresponding to event subtypes in BKEE.

Event Type	Event Sub-Type
Business	Start-Org, Merge-Org, Declare-Bankruptcy, End-Org
Conflict	Attack, Demonstrate
Life	Be-Born, Marry, Divorce, Injure, Die

Table 2: Example of event types and corresponding event subtypes in BKEE.

Task	Count	IAA(%)	Challenges
EMD	8,717	83.0	JOB
ED	16,010	83.0	DEMONSTRATE
EAE	16,010	85.1	DESTINATION

Table 3: Annotation overview of BKEE. **Count**: Count of annotated events, entities, and argument roles. **IAA**: Inter-Annotator Agreement scores between annotators. **Challenges**: Identifying the most challenging type to annotate.

and the MASI distance metric (Passonneau, 2006). These enable us to calculate inter-annotator agreements (IAA) for each task based on the data annotated in the co-annotated stage. In the second stage, the remaining 90% of the documents are distributed among annotators for separate annotation. Here, annotators have the opportunity to collaborate by sharing their annotations and participating in discussions to resolve any discrepancies and reach a consensus on the final dataset.

2.3. Data Statistics

Table 3 reports a statistic of the total number of event mentions, entity mentions, and arguments labeled. Additionally, we report the agreement scores among annotators for each task, accompanied by the most challenging types to label. The challenging type is determined based on inter-annotator disagreements (total disagreements divided by the number of occurrences of that type), where higher disagreement indicates a higher level of challenge. As can be seen, with nearly *9k* events, over *16k* entities, and arguments, our dataset demonstrates significant potential for serving future deep learning applications. Furthermore, the diversity of types is evident through the specific distributions of event types, entity types, and argument types as illustrated in Figures 3, 2, and 4, respectively. This diversity underscores the dataset's versatility, making it suitable for a wide range of domains and applications.

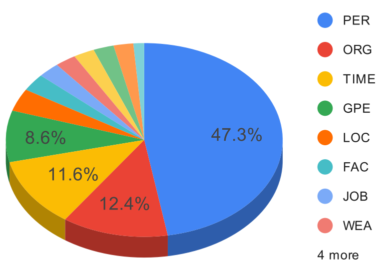

Figure 2: Distributions of entity types in BKEE.

3. Experiments

In this section, we conduct a performance evaluation of baselines and EE state-of-the-art models on BKEE to reveal its challenges inherent.

In our experiments, we adopt two distinct EE approaches:(1) Pipeline: We develop separate models for each of the three tasks (EMD, ED, and EAE). Each model is designed to handle a specific task independently. (2) Joint-learning: We utilize a model that simultaneously learns and infers all three tasks. This approach aims to mitigate error propagation and capitalize on the inherent interdependencies among these tasks. In both approaches, during the training, EMD and ED are modeled as BIO-labeled sequence tasks, while EAE is approached through the classification of the relationship between event mentions and entity mentions predicted by EMD and ED. Table 4 illustrates how these tasks are modeled, corresponding to the example in Figure1.

3.1. Baseline Models

For the pipeline, we use a pretrained transformer-based language model (XLM-Roberta/ PhoBERT) to encode input text for each task. For EMD and ED tasks, token representations are fed into a feed-forward network to compute label distributions, while event mention and entity mention word representations are concatenated and fed into a feed-forward network for argument role prediction. For the joint-learning models, we evaluate two SOTA joint-learning EE models OneIE (Lin et al., 2020) and FourIE (Nguyen et al., 2021) on BKEE.

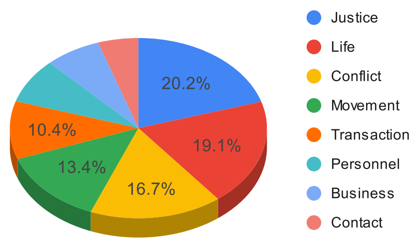

Figure 3: Distributions of event types in BKEE.

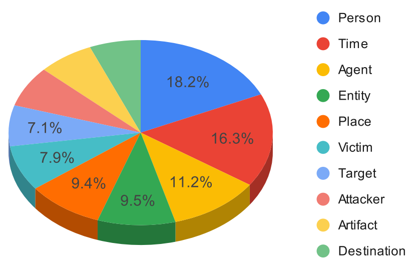

Figure 4: Distributions of key arguments in BKEE.

Input	Mr	Pham	Manh	Thang	will	hold	the	position	of	deputy	general	director	of	VietcomBank	from	March	10/3	.
ED	O	O	O	O	O	B-start-position	O	O	O	O	O	O	O	O	O	O	O	O
EMD	B-PER	I-PER	I-PER	I-PER	O	O	O	O	O	B-JOB	I-JOB	I-JOB	I-JOB	I-JOB	O	B-TIME	I-TIME	O
EAE	(hold-Mr Pham Manh Thang):Person; (hold-deputy general director of Vietcombank):JOB; (hold-March 10/3):TIME																	

Table 4: Illustration of ED, EMD, and EAE tasks in the training process.

3.2. Hyper-parameters

For pre-trained language models (PLMs), we conducted experiments using models based on PhoBERT (Nguyen and Tuan Nguyen, 2020) and XLM-Roberta (Conneau et al., 2020) (large version). During training, EAE is provided with golden EMD and ED labels, whereas during the evaluation, it only accesses event mentions and entity mentions previously predicted by EMD and ED. In the model pipeline, we fine-tuned PLMs, where the feed-forward network consists of 2 linear layers with a dimension of 250, a learning rate of 1e-5, a batch size of 16, and used the Adam optimizer. For OneIE and FourIE, we leveraged the parameters recommended by the original works (Lin et al., 2020; Nguyen et al., 2021). For data, we randomly divide the dataset into training, development, and testing sets with a ratio of 3:1:1 by documents. Following prior EE works (Lin et al., 2020; Nguyen et al., 2021, 2022), we report F1 scores of EE models over three tasks EMD, ED, and EAE for performance measure. We report the average performances of five different runs with random seeds.

Unlike several languages, e.g. English, where words are separated by spaces, Vietnamese presents a more complex word tokenization task due to the presence of multi-syllabic words, while PhoBERT requires the input that is word-tokenized. To address this issue, we employ *VnCoreNLP* (Vu et al., 2018), along with human quality checks to enhance word tokenization accuracy.

3.3. Results

Tables 5 and 6 illustrate the performance of baselines on BKEE when utilizing pretrained Phobert and Roberta models. Despite not being specifically designed for the Vietnamese language, XLM-Roberta still delivers competitive results on the pipeline, but in general EE PhoBERT-based models perform better than XLM-Roberta-based models. Furthermore, the joint learning models (OneIE, FourIE) outperform the pipeline models, especially in argument role classification, demonstrating that joint learning can help mitigate error propagation in end-to-end models. Therefore, future research efforts may benefit from a more focused approach in this direction.

3.4. Error Analysis

To better grasp the dataset's challenges for extraction systems, we analyze 100 random errors of FourIE using PhoBERT on the test set and describe the main error categories below:
Overlapping context (35%): Usually occurs in the EAE task. Sentences containing events are often long and have many overlapping contextual elements, leading to errors in prediction. For example, in the sentence *After the crime, Hung visited Tran Van Chien's house to discuss it, and Chien purchased a SIM card to stay in touch with Hung during his escape.*, the entity "Chien" was assigned the wrong role "Agent" to activate event "escape", while "Chien" is not the subject of this escape.
Span errors (28%): Usually occurs in the EMD and ED task. These errors occur when the model captures part of a mention but does not overlap completely with the gold one. For example, the entity "Ngan_hang Nong_nghiep và Phat_trien nong_thon Viet_Nam Agribank" (*Vietnam Bank for Agriculture and Rural Development Agribank*) is only partially detected as "Agribank". Through our in-house experiments, these errors primarily stem from (1) the complex structure of entity mentions and event mentions in Vietnamese. For instance, organizational names in Vietnam are often lengthy (40% of entity mentions contain more than 2 words, the longest case even up to 82 words), and (2) the error of word tokenization leads to mentions ending too early, starting too late, or missing a syllable in the middle of a span. Our in-house experiments have revealed that relying solely on whitespace for word segmentation, a common practice in English, significantly diminishes the baseline performance of FourIE with pre-trained Pho-bert. This resulted in a noteworthy 1.7% drop in F1-score for entity mention detection, 0.6% for event detection, and 2.3% for event argument extraction. These findings underscore the tangible impact of word segmentation on model performance.
Potentially relevant (12%): Entities, triggers, and arguments are identified that can be considered

Task	Pipeline	OneIE	FourIE
Entity	54.4	55.8	**57.6**
Event	61.8	**62.8**	61.9
Argument	44.4	53.0	**53.4**

Table 5: The performance (F1-score) of baselines using PhoBERT on BKEE.

Task	Pipeline	OneIE	FourIE
Entity	55.0	56.3	**56.4**
The Event	60.3	60.0	**61.5**
Argument	44.9	**51.7**	51.6

Table 6: The performance (F1-score) of baselines using XLM-RoBERTa on BKEE.

valid based on manual review. For example, while the golden entity is "Iraq and Syria", the model identifies two entities as "Iraq" and "Syria".

Abbreviations (6%): Abbreviations in the text are sometimes misunderstood. For example, the acronym "CEO" is mislabeled as "JOB" instead of "PER", or vice versa.

Proportions do not add up to 100% because we exclude less common errors or undefined classes.

4. Conclusion

In this work, we propose BKEE, the first Vietnamese-language EE dataset that achieves three main goals: (1) reducing the gap between rich-resource and low-resource languages in the field of EE, (2) pioneering EE development for the Vietnamese language, and (3) establishing strong baselines to support future works and analyzing the challenges faced by Vietnamese EE. Experimental results indicate that Vietnamese EE encounters cases of overlapping context, complex event and entity mentions, and the critical preprocessing task of word-tokenization to enhance performance.

5. Limitations

As the first EE dataset for the Vietnamese language, BKEE reduces the gap between rich-resource and low-resource languages in the field of EE. However, some limitations can be improved in the future. First, although BKEE offers a significant amount of events, entities, and arguments, it is currently labeled intra-sentence due to limitations in our human resources. Expanding to the document level would result in an exponential increase in the number of labeled samples required. Therefore, future works may consider extending BKEE beyond the sentence level to enhance the overall understanding of global semantics and complex information processing. However, scaling annotations at the document level can lead to an exponential increase in labeling effort, leading to overwhelming annotators and affecting data quality. To overcome this, future efforts could adopt strategies used in several EE volumes, such as RAMS (Ebner et al., 2020), in which articles are divided into a number of consecutive sentences (called a segment). This approach allows individual annotations for each segment, allowing annotators to better grasp context and provide more accurate event and entity annotations. Second, our experiments indicate that the quality of word-tokenization might affect the EE performance of PhoBERT-based models. Future works can improve word boundary detection to minimize span errors or increase the ability to understand document structure to minimize context complexity. Nevertheless, there is still room to investigate the performance of syllable-based EE models. Finally, compared to resource-rich languages, SOTA EE models on BKEE exhibit significantly lower performance, as can be seen in (Veyseh et al., 2022a). Future works may delve deeper into addressing this gap to achieve better performance for Vietnamese EE.

Acknowledgements

This research has been supported by the Army Research Office (ARO) grant W911NF-21-1-0112, the NSF grant CNS-1747798 to the IUCRC Center for Big Learning, and the NSF grant # 2239570. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI or the U.S. Government.

Bibliographical References

David Ahn. 2006. The stages of event extraction. In *Proceedings of the Workshop on Annotating and Reasoning about Time and Events*, pages 1–8, Sydney, Australia. Association for Computational Linguistics.

André Bittar, Pascal Amsili, Pascal Denis, and Laurence Danlos. 2011. French timebank: an iso-timeml annotated reference corpus. In *Proceedings of the 49th annual meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 130–134.

Antoine Bosselut, Ronan Le Bras, and Yejin Choi. 2021. Dynamic neuro-symbolic knowledge graph construction for zero-shot commonsense question answering. In *Proceedings of the AAAI conference on Artificial Intelligence*, volume 35, pages 4923–4931.

Jordan Boyd-Graber and Benjamin Börschinger. 2020. What question answering can learn from trivia nerds. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7422–7435.

Qingqing Cao, Harsh Trivedi, Aruna Balasubramanian, and Niranjan Balasubramanian. 2020. Deformer: Decomposing pre-trained transformers for faster question answering. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4487–4497.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

George R Doddington, Alexis Mitchell, Mark A Przybocki, Lance A Ramshaw, Stephanie M Strassel, and Ralph M Weischedel. 2004. The automatic content extraction (ace) program-tasks, data, and evaluation. In *Lrec*, volume 2, pages 837–840. Lisbon.

Seth Ebner, Patrick Xia, Ryan Culkin, Kyle Rawlins, and Benjamin Van Durme. 2020. Multi-sentence argument linking. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8057–8077, Online. Association for Computational Linguistics.

Li Gao, Jia Wu, Zhi Qiao, Chuan Zhou, Hong Yang, and Yue Hu. 2016. Collaborative social group influence for event recommendation. In *Proceedings of the 25th ACM international on conference on information and knowledge management*, pages 1941–1944.

Andrew Hsi, Yiming Yang, Jaime G Carbonell, and Ruochen Xu. 2016. Leveraging multilingual training for limited resource event extraction. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1201–1210.

Heng Ji and Ralph Grishman. 2008. Refining event extraction through cross-document inference. In *Proceedings of ACL-08: Hlt*, pages 254–262.

Klaus Krippendorff. 2011. Computing krippendorff's alpha-reliability.

Alexander Kuhnle, Miguel Aroca-Ouellette, Anindya Basu, Murat Sensoy, John Reid, and Dell Zhang. 2021. Reinforcement learning for information retrieval. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2669–2672.

Sha Li, Heng Ji, and Jiawei Han. 2021. Document-level event argument extraction by conditional generation. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 894–908, Online. Association for Computational Linguistics.

Ying Lin, Heng Ji, Fei Huang, and Lingfei Wu. 2020. A joint neural model for information extraction with global features. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 7999–8009.

Chun-Yi Liu, Chuan Zhou, Jia Wu, Hongtao Xie, Yue Hu, and Li Guo. 2017. Cpmf: A collective pairwise matrix factorization model for upcoming event recommendation. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 1532–1539. IEEE.

Fanzhen Liu, Shan Xue, Jia Wu, Chuan Zhou, Wenbin Hu, Cecile Paris, Surya Nepal, Jian Yang, and Philip S Yu. 2021. Deep learning for community detection: progress, challenges and opportunities. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 4981–4987.

Xiao Liu, Heyan Huang, Ge Shi, and Bo Wang. 2022. Dynamic prefix-tuning for generative template-based event extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5216–5228, Dublin, Ireland. Association for Computational Linguistics.

Weiyi Lu and Thien Huu Nguyen. 2018. Similar but not the same: Word sense disambiguation improves event detection via neural representation matching. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4822–4828, Brussels, Belgium. Association for Computational Linguistics.

Xiaoxiao Ma, Jia Wu, Shan Xue, Jian Yang, Chuan Zhou, Quan Z Sheng, Hui Xiong, and Leman Akoglu. 2021. A comprehensive survey on graph anomaly detection with deep learning. *IEEE Transactions on Knowledge and Data Engineering*.

Teruko Mitamura and Zhengzhong Liu. 2016. Overview of tac kbp 2015 event nugget track.

Teruko Mitamura and Zhengzhong Liu. 2017. Events detection, coreference and sequencing: What's next? overview of the tac kbp 2017 event track.

Dat Quoc Nguyen and Anh Tuan Nguyen. 2020. PhoBERT: Pre-trained language models for Vietnamese. In *Findings of the Association for*

Computational Linguistics: EMNLP 2020, pages 1037–1042, Online. Association for Computational Linguistics.

Minh Van Nguyen, Viet Dac Lai, and Thien Huu Nguyen. 2021. Cross-task instance representation interactions and label dependencies for joint information extraction with graph convolutional networks. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 27–38.

Minh Van Nguyen, Bonan Min, Franck Dernoncourt, and Thien Nguyen. 2022. Learning cross-task dependencies for joint extraction of entities, events, event arguments, and relations. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9349–9360, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Thien Huu Nguyen, Kyunghyun Cho, and Ralph Grishman. 2016. Joint event extraction via recurrent neural networks. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 300–309.

Rebecca J Passonneau. 2006. Measuring agreement on set-valued items (masi) for semantic and pragmatic annotation. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*.

Roser Saurı and Toni Badia. 2012. Catalan timebank 1.0 corpus documentation. Technical report, Technical report.

Amir Pouran Ben Veyseh, Javid Ebrahimi, Franck Dernoncourt, and Thien Nguyen. 2022a. Mee: A novel multilingual event extraction dataset. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9603–9613.

Amir Pouran Ben Veyseh, Minh Van Nguyen, Franck Dernoncourt, and Thien Nguyen. 2022b. MINION: a large-scale and diverse dataset for multilingual event detection. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2286–2299, Seattle, United States. Association for Computational Linguistics.

Amir Pouran Ben Veyseh, Minh Van Nguyen, Nghia Ngo Trung, Bonan Min, and Thien Huu Nguyen. 2021. Modeling document-level context for event detection via important context selection. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 5403–5413, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Amir Pouran Ben Veyseh and Thien Nguyen. 2022. Word-label alignment for event detection: A new perspective via optimal transport. In *Proceedings of the 11th Joint Conference on Lexical and Computational Semantics*, pages 132–138, Seattle, Washington. Association for Computational Linguistics.

Thanh Vu, Dat Quoc Nguyen, Mark Dras, Mark Johnson, et al. 2018. Vncorenlp: A vietnamese natural language processing toolkit. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pages 56–60.

Christopher Walker, Stephanie Strassel, Julie Medero, and Kazuaki Maeda. 2006. Ace 2005 multilingual training corpus. In *Technical report, Linguistic Data Consortium*.

Xiaozhi Wang, Ziqi Wang, Xu Han, Wangyi Jiang, Rong Han, Zhiyuan Liu, Juanzi Li, Peng Li, Yankai Lin, and Jie Zhou. 2020. Maven: A massive general domain event detection dataset. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1652–1671.

Xindong Wu, Jia Wu, Xiaoyi Fu, Jiachen Li, Peng Zhou, and Xu Jiang. 2019. Automatic knowledge graph construction: A report on the 2019 icdm/icbk contest. In *2019 IEEE International Conference on Data Mining (ICDM)*, pages 1540–1545. IEEE.

Weinan Zhang, Xiangyu Zhao, Li Zhao, Dawei Yin, and Grace Hui Yang. 2021. Drl4ir: 2nd workshop on deep reinforcement learning for information retrieval. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2681–2684.