Mirror: A Universal Model for Various Information Extraction Tasks

Anonymous EMNLP submission

Abstract

The variety of information extraction tasks and data formats make it hard to share common knowledge among those tasks. This causes wastes to some extent and adds difficulties to build complex pipeline applications in real scenarios. Recent studies formulate IE tasks as a triplet extraction problem. However, such format does not support multi-span and n-ary extraction tasks, leading to weak versatility. To this end, we reorganize IE datasets into a unified format and propose a universal model for various IE tasks, namely Mirror. We regard IE tasks as a multi-span cyclic graph extraction problem, and devise a non-autoregressive graph decoding algorithm to extract all spans in a single step. Experiments on 22 datasets across 6 IE tasks show that our model has good compatibilities and achieves good performances under few-shot and zero-shot settings. The code, model weights and data will be publicly available at github.

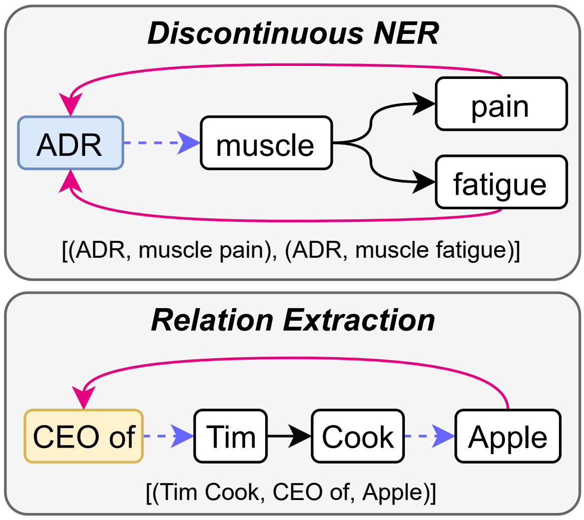

Figure 1: Multi-span cyclic graph for discontinuous NER and RE tasks (best viewed in color). The spans are connected by three types of edges, including ***consecutive connections***, dotted ***jump connections*** and ***tail-to-head connections***. *ADR* in discontinuous NER denotes the entity label of Adverse Drug Reaction.

1 Introduction

Information Extraction (IE) is a fundamental task in Natural Language Processing (NLP), which aims to extract structured information from unstructured text, such as Named Entity Recognition (NER), Relation Extraction (RE), Event Extraction (EE), etc. Each IE task is usually isolated, and it needs specific datasets and delicate models, which makes it difficult to build complex applications. It is a fascinating problem to combine different IE tasks into a whole, and make one model for universal IE tasks.

There are two ways to combine different IE tasks. The first is to utilize generative pretrained language models (PLMs) to generate the structured information directly. x, x and x structure the IE tasks as a sequence generation problem, and use T5 models to predict the structured information autoregressively. However, such methods cannot provide the exact positions of the structured information, which is important for some IE tasks, such as NER. Besides, the generation-based methods are usually slow, and it is difficult to train them on large-scale datasets. The second is to apply the extractive PLMs, which is way more faster to train and inference. UIE unifies the IE tasks into span spotting and associating problems. Based on this idea, x, x and x take the IE tasks into a triplet prediction problems, and propose different methods. However, such methods are not suitable for multi-span and n-ary IE tasks.

To extend the universal IE systems into more tasks, we propose a multi-task framework, namely Mirror, which can be applied on various IE tasks, including multi-span and n-ary extraction problems. We formulate IE tasks into a unified schema-guided multi-span graph, and regard labels as spans' leading tokens. We design three kind of connections and utilize the biaffine-style method to extract the structured information.

Model	TANL	UIE	DeepStruct	InstructUIE	USM	UniEX	RexUIE	Mirror
PLM #Params	T5-base 220M	T5-large 770M	GLM 10B	FlanT5 11B	RoBERTa large 372M	RoBERTa large 372M	DeBERTa-v3 large 434M	DeBERTa-v3 large 434M
Single-step	✓	✓	✓	✓	✓	✓	✗	✓
Indexing	✗	✗	✗	✗	✓	✓	✓	✓
NAR	✗	✗	✗	✗	✓	✓	✓	✓
Multi-span	✗	◦	◦	◦	✗	✗	✗	✓
N-ary	✗	✗	✗	✗	✗	✗	✓	✓

Table 1: Comparisons with other systems. Single-step represents that the model predicts results in a single step. Indexing means whether the model could provide exact information positions. NAR denotes the non-autoregressive decoding strategy. Multi-span means the model supports multi-span extraction, e.g. the discontinuous named entity recognition task. N-ary denotes the ability of n-ary tuple extraction. ◦ means the model supports the task theoretically, but the implementation is not available.

We conduct extensive experiments on 6 IE tasks, including NER, RE, EE, Aspect-based Sentiment Analysis (ABSA), multi-span discontinuous NER and n-ary hyper RE. Our Mirror shows good compatibility across different tasks and datasets, and achieves competitive results on few-shot settings. Besides, 434M Mirror outperforms FlanT5-11B InstructUIE in zero-shot NER tasks.

Our contributions are summarized as follows:

- We propose a multi-task framework, namely Mirror, which can be trained on various IE tasks, including multi-span and n-ary extraction problems.
- We conduct extensive experiments on x IE tasks, and the results show that our model achieves state-of-the-art performance on most of the tasks.

2 Related Work

2.1 Multi-task Information Extraction

Multi-task IE is a popular research topic in recent years. The main idea is to use a single model to perform multiple IE tasks. IE tasks could be formulated as different graph structures. Li et al. (2022) formulate flat, nested, and discontinuous NER tasks as a graph with next-neighboring and tail-to-head connections. Maximal cliques also have been used to flat & discontinuous NER tasks (Wang et al., 2021) and trigger-available & trigger-free event extractions (Zhu et al., 2022). DyGIE++ takes NER, RE and EE tasks as span graphs, and apply iterative propagation to enhance spans' contextual representations (Wadden et al., 2019). OneIE uses the similar graph structures with global constraint features (Lin et al., 2020).

In addition to explicit graph-based multi-task IE systems, generative language models are also been widely used. Yan et al. (2021b) and Yan et al. (2021a) add special index tokens into BART (Lewis et al., 2020) vocabulary to help perform various NER and ABSA tasks and obtain explicit span positions. TANL (Paolini et al., 2021) apply T5 (Raffel et al., 2020) to generate texts with special enclosures as the predicted information. GenIE (Josifoski et al., 2022) and DeepStruct (Wang et al., 2022) share a similar idea to generate subject-relation-object triplets, and DeepStruct extends the model size to 10B with GLM as the backbone (Du et al., 2022).

2.2 Schema-guided Information Extraction

In schema-guided IE systems, schemas are input as an guidance signal to help the model extracting target information. UIE (Lu et al., 2022) categorize IE tasks into span spotting and asscociating elementary tasks and devise a linearized query language. Fei et al. (2022) introduces the hyper relation extraction task to represent complex IE tasks like EE, and utilize external parsing tools to enhance the text representations. InstructUIE (Wang et al., 2023) formulates schemas into instructions and uses FlanT5-11B (Chung et al., 2022) to performing multi-task instruction tuning.

While the above methods use flexible generative language models, they cannot predict exact positions, which brings ambiguity when evaluating. Besides, large generative language models are usually slow to train and inference, and requires tons of computing resources. USM (Lou et al., 2023) and UniEX (Lu et al., 2023) utilize BERT-family models to extract triplets non-autoregressively. USM regards IE tasks into a unified schema matching

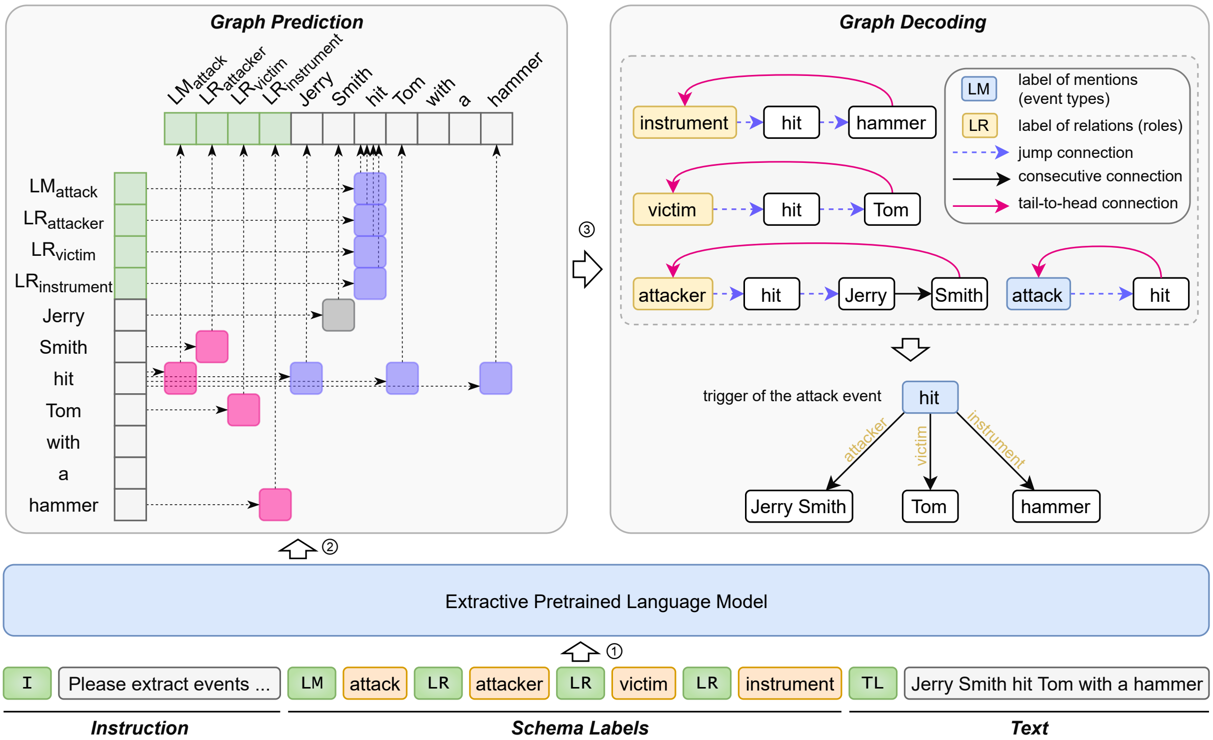

Figure 2: Model framework (best viewed in color).

task and use a label-text matching model to extract triplets. UniEX inherits the idea and uses a triaffine module to predict hyper-dimensional links. Beyond triplets extraction, RexUIE (Liu et al., 2023) perform to extract n-ary tuples step-by-step. Although RexUIE supports n-ary extraction, it sacrifices the efficiency and requires more time when predicting results (one slot per step).

3 Mirror

3.1 Unified Data Interface

To make the model able to handle different IE tasks, we propose a unified data interface for the input. As shown in Figure 2, there are three parts in the model input: the *instruction*, the *schema labels*, and the *text*. The instruction is composed of a leading token `[I]` and a natural language sentence. The leading token indicates the instruction part while the sentence tells the model what it should do. For example, the instruction of NER could be *Please identify any possible entities in the given text and label them with the following types*. The instruction is the question in Machine Reading Comprehension (MRC) and Question Answering (QA) datasets. For each task, we manually design a set of instructions, and randomly pick one of them for each training sample. The number of instructions for each IE task is listed in Table 2.

Task	#Dataset	#Instruction	#Instance
NER		42	
RE		53	
EE		40	

Table 2: Pretraining dataset statistics.

The schema labels are task ontologies that used for schema-guided extraction. This part is consists of special token labels (`[LM]`, `[LR]` and `LC`) and corresponding label texts. Among the special tokens, `[LM]` denotes the label of mentions (or event types), `[LR]` denotes the label of relations (or argument roles), and `[LC]` denotes the label of classes. `[LC]` token is designed for classification tasks when pretraining.

The text part is the input text that the model should extract information from. It is composed of a leading token (`[TL]` or `[TP]`) and a natural language sentence. If the leading token is `[TL]`, the model should link labels from schema labels to spans in the text. While the `[TP]` token indicates the target spans are only in the text, and the model should extract information from the text without schema labels. The `[TP]` label is used in the pretraining stage to make the model able to extract information in MRC tasks without schema. In classification tasks when pretraining, the model should not extract anything from the text part. So we add a special background area with a leading token `[B]` to distinguish from extractive texts.

With the above three parts, we can formulate classification, extractive MRC, multi-choice MRC, and IE tasks into a unified data interface, and the model can be trained in a unified way even the model is not based on generative language models.

3.2 Multi-span Cyclic Graph (MCG)

We formulate IE tasks as a unified multi-span extraction problem. For the flat NER task, the model is expected to extract a tuple like: "((entity label position,), (span start position, span end position))". As shown in Figure 1 and the top right of Figure 2, we formulate IE tasks into a unified multi-span cyclic graph, and regard labels as the leading tokens in schema labels. There are three types of connections in the graph: the ***consecutive*** connection, the ***jump*** connection, and the ***tail-to-head*** connection.

The consecutive connection is used to **spans in the same entity**. For an entity that has multiple tokens, the consecutive connection connects from the first token to the last token. As shown in Figure 2, "Jerry" connects to "Smith". If there is only one token in an entity, the consecutive connection is not used. For example, entities in "muscle pain and fatigue" contains two entities "muscle pain" and "muscle fatigue". The consecutive connection is used to connect from "muscle" to "pain", and "muscle" to "fatigue". The jump connection connects **different slots** in a tuple. Schema labels and spans from texts are in different slots, so they are connected in jump connections. In addition, the head entity and the tail entity of a relation triplet are in different slots, so they are also connected in jump connections. The tail-to-head connection helps **locate the start & end boundaries**, and forms a cycle in the graph. It connects from the last token of the last slot to the first token of the first slot in a tuple.

3.3 Model Structure

Text text text

If we

4 Experiments

4.1 Experiment Setup

4.2 Datasets

4.3 Main Results

4.4 Few-shot Results

4.5 Zero-shot Results

4.6 Analysis on Training Strategies

4.7 Ablation Study

Limitations

Content input length and model compatibility. Multi-turn result modification. Laborious data cleaning and format unification.

Ethics Statement

All datasets are publicly available without further annotation, and the NLU tasks are traditional tasks in natural language processing communities. So there are no ethical issues.

References

Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Eric Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, Albert Webson, Shixiang Shane Gu, Zhuyun Dai, Mirac Suzgun, Xinyun Chen, Aakanksha Chowdhery, Sharan Narang, Gaurav Mishra, Adams Yu, Vincent Y. Zhao, Yanping Huang, Andrew M. Dai, Hongkun Yu, Slav Petrov, Ed H. Chi, Jeff Dean, Jacob Devlin, Adam Roberts, Denny Zhou, Quoc V. Le, and Jason Wei. 2022. Scaling instruction-finetuned language models. *CoRR*, abs/2210.11416.

Zhengxiao Du, Yujie Qian, Xiao Liu, Ming Ding, Jiezhong Qiu, Zhilin Yang, and Jie Tang. 2022. GLM: General language model pretraining with autoregressive blank infilling. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 320–335, Dublin, Ireland. Association for Computational Linguistics.

Hao Fei, Shengqiong Wu, Jingye Li, Bobo Li, Fei Li, Libo Qin, Meishan Zhang, Min Zhang, and Tat-Seng Chua. 2022. Lasuie: Unifying information extraction with latent adaptive structure-aware generative language model. In *NeurIPS*.

Martin Josifoski, Nicola De Cao, Maxime Peyrard, Fabio Petroni, and Robert West. 2022. GenIE: Generative information extraction. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4626–4643, Seattle, United States. Association for Computational Linguistics.

Task	Datasets	TANL	UIE	DeepStruct	InstructUIE	USM	UniEX	RexUIE	Mirror
NER	ACE04	-	86.89	-	-	87.62	87.12	87.25	87.63
	ACE05	84.90	85.78	86.9	86.66	87.14	87.02	87.23	86.81
	CoNLL03	91.70	92.99	93	92.94	93.16	92.65	93.67	92.83
RE	ACE05	63.70	66.06	66.8	-	67.88	66.06	64.87	66.70
	CoNLL04	71.40	75.00	78.3	78.48	78.84	73.40	78.39	71.48
	NYT	-	93.54	93.3	90.47	94.07	-	94.55	91.66
	SciERC	-	36.53	-	45.15	37.36	38.00	38.37	36.08
Event	ACE05-Tgg	68.40	73.36	69.8	77.13	72.41	74.08	75.17	69.95
	ACE05-Arg	47.60	54.79	56.2	72.94	55.83	53.92	59.15	48.12
	CASIE-Tgg	-	69.33	-	67.80	71.73	71.46	73.01	69.19
	CASIE-Arg	-	61.30	-	63.53	63.26	62.91	63.87	55.09
ABSA	14-res	-	74.52	-	-	77.26	74.77	77.46	77.20
	14-lap	-	63.88	-	-	65.51	65.23	66.41	62.84
	15-res	-	67.15	-	-	69.86	68.58	70.84	68.82
	16-res	-	75.07	-	-	78.25	76.02	77.20	75.53

Table 3: Main results.

	P	R	F1
Discontinuous NER: CADEC			
BART-NER	70.08	71.21	70.64
W2NER	74.09	**72.35**	**73.21**
Mirror	**74.83**	67.88	71.19
N-ary Tuples: HyperRED			
CubeRE	66.39	67.12	66.75
RexUIE	-	-	**75.20**
Mirror	70.88	64.05	67.29

Table 4: Results on multi-span and n-ary inforamtion extraction tasks. The best results are in **bold**, and the second best results are underlined.

	Model	1-shot	5-shot	10-shot	Avg.
NER CoNLL03	UIE	57.53	75.32	79.12	70.66
	USM	71.11	83.25	84.58	79.65
	RexUIE	**86.57**	**89.63**	**90.82**	**89.07**
	Mirror	77.50	82.73	84.48	81.57
RE CoNLL04	UIE	34.88	51.64	58.98	48.50
	USM	36.17	53.20	60.99	50.12
	RexUIE	**43.80**	**54.90**	**61.68**	**53.46**
	Mirror	34.66	52.23	58.68	48.52
Event Trigger ACE05	UIE	42.37	53.07	54.35	49.93
	USM	40.86	55.61	58.79	51.75
	RexUIE	**56.95**	64.12	**65.41**	**62.16**
	Mirror	49.50	**65.61**	60.68	58.60
Event Arg ACE05	UIE	14.56	31.20	35.19	26.98
	USM	19.01	36.69	42.48	32.73
	RexUIE	**30.43**	41.04	**45.14**	**38.87**
	Mirror	23.46	**48.32**	41.90	37.89
ABSA 16res	UIE	23.04	42.67	53.28	39.66
	USM	30.81	52.06	58.29	47.05
	RexUIE	37.70	49.84	60.56	49.37
	Mirror	**67.06**	**73.51**	**68.70**	**69.76**

Table 5: Few-shot results. The best results are in **bold**, and the second best results are underlined.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.

Jingye Li, Hao Fei, Jiang Liu, Shengqiong Wu, Meishan Zhang, Chong Teng, Donghong Ji, and Fei Li. 2022. Unified named entity recognition as word-word relation classification. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, pages 10965–10973. AAAI Press.

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, Online. Association for Computational Linguistics.

Chengyuan Liu, Fubang Zhao, Yangyang Kang, Jingyuan Zhang, Xiang Zhou, Changlong Sun, Fei Wu, and Kun Kuang. 2023. RexUIE: A Recursive Method with Explicit Schema Instructor for Univer-

Model	Movie	Restaurant	AI	Literature	Music	Politics	Science	Avg.
Davinci	0.84	2.94	2.97	9.87	13.83	18.42	10.04	8.42
ChatGPT	41.00	**37.76**	**54.40**	54.07	**61.24**	59.12	**63.00**	**52.94**
USM	37.73	14.73	28.18	**56.00**	44.93	36.10	44.09	37.39
InstructUIE	**63.00**	20.99	49.00	47.21	53.61	48.15	49.30	47.32
Mirror	40.96	20.02	51.13	44.80	60.63	**61.19**	53.65	47.48
Upper Bound	85.94	83.30	65.72	67.93	78.25	75.92	70.96	75.43

Table 6: Zero-shot NER results. The best results are in **bold**, and the second best results are underlined. The upper bound is the Mirror performance where these zero-shot NER training sets are included in the pretraining phase.

Dataset	Training	NER CoNLL03	RE CoNLL04	Event Trigger ACE05	Event Arg ACE05	ABSA 16res	Avg.
Included in PT	Multi-task FT						
	Single-task FT	92.83	71.48	69.95	48.12	75.53	71.58
Excluded in PT	Multi-task FT	91.84	72.39	68.24	47.73	75.31	71.10
	Single-task FT	92.45	73.70	71.01	52.57	75.55	73.06

Table 7: Analysis on training strategy.

sal Information Extraction. ArXiv:2304.14770 [cs] version: 1.

Jie Lou, Yaojie Lu, Dai Dai, Wei Jia, Hongyu Lin, Xianpei Han, Le Sun, and Hua Wu. 2023. Universal information extraction as unified semantic matching. *CoRR*, abs/2301.03282.

Junyu Lu, Ping Yang, Ruyi Gan, Junjie Wang, Yuxiang Zhang, Jiaxing Zhang, and Pingjian Zhang. 2023. UniEX: An Effective and Efficient Framework for Unified Information Extraction via a Span-extractive Perspective. ArXiv:2305.10306 [cs].

Yaojie Lu, Qing Liu, Dai Dai, Xinyan Xiao, Hongyu Lin, Xianpei Han, Le Sun, and Hua Wu. 2022. Unified structure generation for universal information extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5755–5772, Dublin, Ireland. Association for Computational Linguistics.

Giovanni Paolini, Ben Athiwaratkun, Jason Krone, Jie Ma, Alessandro Achille, Rishita Anubhai, Cícero Nogueira dos Santos, Bing Xiang, and Stefano Soatto. 2021. Structured prediction as translation between augmented natural languages. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21:140:1–140:67.

David Wadden, Ulme Wennberg, Yi Luan, and Hannaneh Hajishirzi. 2019. Entity, relation, and event extraction with contextualized span representations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5784–5789, Hong Kong, China. Association for Computational Linguistics.

Chenguang Wang, Xiao Liu, Zui Chen, Haoyun Hong, Jie Tang, and Dawn Song. 2022. DeepStruct: Pretraining of language models for structure prediction. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 803–823, Dublin, Ireland. Association for Computational Linguistics.

Xiao Wang, Weikang Zhou, Can Zu, Han Xia, Tianze Chen, Yuansen Zhang, Rui Zheng, Junjie Ye, Qi Zhang, Tao Gui, Jihua Kang, Jingsheng Yang, Siyuan Li, and Chunsai Du. 2023. InstructUIE: Multi-task Instruction Tuning for Unified Information Extraction. ArXiv:2304.08085 [cs].

Yucheng Wang, Bowen Yu, Hongsong Zhu, Tingwen Liu, Nan Yu, and Limin Sun. 2021. Discontinuous named entity recognition as maximal clique discovery. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 764–774, Online. Association for Computational Linguistics.

Hang Yan, Junqi Dai, Tuo Ji, Xipeng Qiu, and Zheng Zhang. 2021a. A unified generative framework for aspect-based sentiment analysis. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*

	NER CoNLL03	RE CoNLL04	Event Trigger ACE05	Event Arg ACE05	ABSA 16res	Avg.
Mirror						
- Pretrain						
- Pretrain & Instruction						

Table 8: Ablation study.

(Volume 1: Long Papers), pages 2416–2429, Online. Association for Computational Linguistics.

Hang Yan, Tao Gui, Junqi Dai, Qipeng Guo, Zheng Zhang, and Xipeng Qiu. 2021b. A unified generative framework for various NER subtasks. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 5808–5822, Online. Association for Computational Linguistics.

Tong Zhu, Xiaoye Qu, Wenliang Chen, Zhefeng Wang, Baoxing Huai, Nicholas Yuan, and Min Zhang. 2022. Efficient document-level event extraction via pseudo-trigger-aware pruned complete graph. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 4552–4558. International Joint Conferences on Artificial Intelligence Organization. Main Track.

A Dataset Statistics

test

This is a section in the appendix.