Universal Information Extraction as Unified Semantic Matching

Jie Lou[1*], Yaojie Lu[2*], Dai Dai[1†], Wei Jia[1], Hongyu Lin[2], Xianpei Han[2,3†], Le Sun[2,3], Hua Wu[1]

[1]Baidu Inc., Beijing, China
[2]Chinese Information Processing Laboratory, Institute of Software, Chinese Academy of Sciences, Beijing, China
[3]State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing, China
{loujie, daidai, jiawei07, wu_hua}@baidu.com
{luyaojie, hongyu, xianpei, sunle}@iscas.ac.cn

Abstract

The challenge of information extraction (IE) lies in the diversity of label schemas and the heterogeneity of structures. Traditional methods require task-specific model design and rely heavily on expensive supervision, making them difficult to generalize to new schemas. In this paper, we decouple IE into two basic abilities, structuring and conceptualizing, which are shared by different tasks and schemas. Based on this paradigm, we propose to universally model various IE tasks with **U**nified **S**emantic **M**atching (USM) framework, which introduces three unified token linking operations to model the abilities of structuring and conceptualizing. In this way, USM can jointly encode schema and input text, uniformly extract substructures in parallel, and controllably decode target structures on demand. Empirical evaluation on 4 IE tasks shows that the proposed method achieves state-of-the-art performance under the supervised experiments and shows strong generalization ability in zero/few-shot transfer settings.

Introduction

Information extraction aims to extract various information structures from texts (Andersen et al. 1992; Grishman 2019). For example, given the sentence "Monet was born in Paris, the capital of France", an IE system needs to extract various task structures such as entities, relations, events, or sentiments in the sentence. It is challenging because the target structures have diversified label schemas (person, work for, positive sentiment, etc.) and heterogeneous structures (span, triplet, etc.).

Traditional IE model leverages task- and schema-specialized architecture, which is commonly specific to different target structures and label schemas. The expensive annotation leads to limited predefined categories and small data size in general domains for information extraction tasks. From another perspective, task-specific model design makes it challenging to migrate learned knowledge between different tasks and extraction frameworks. The above problems lead to the poor performance of IE models in low-resource settings or facing new label schema, which greatly restricts the application of IE in real scenarios.

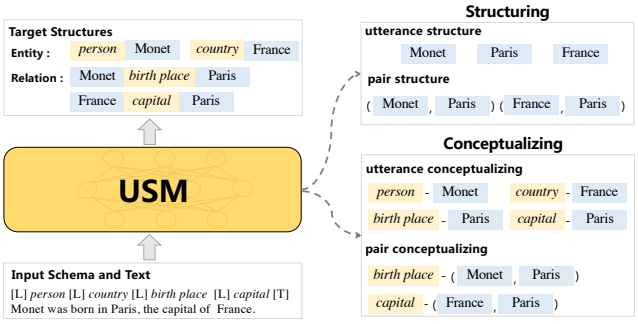

Figure 1: The USM framework for UIE. USM takes label schema and text as input and directly outputs the target structure through the *Structuring* and *Conceptualizing* operations.

Very recently, Lu et al. (2022) proposed the concept of universal information extraction (UIE), which aims to resolve multiple IE tasks using one universal model. To this end, they proposed a sequence-to-sequence generation model, which takes flattened schema and text as input, and directly generates diversified target information structures. Unfortunately, all associations between information pieces and schemas are implicitly formulated due to the black-box nature of sequence-to-sequence models (Alvarez-Melis and Jaakkola 2017). Consequently, it is difficult to identify what kind of abilities and knowledge are learned to transfer across different tasks and schemas. Therefore we have no way of diagnosing under what circumstances such transfer learning across tasks or schemas would fail. For the above reasons, it is necessary to explicitly model and learn transferable knowledge to obtain effective, robust, and explainable transferability.

We find that, as shown in Figure 1, even with diversified tasks and extraction targets, all IE tasks can be fundamentally decoupled into the following two critical operations: 1) ***Structuring***, which proposes label-agnostic basic substructures of the target structure from the text. For example, proposing the utterance structure "Monet" for entity mention and "born in" for event mention, the associated pair structure ("Monet", "Paris") for relation mention, and ("born in", "Paris") for event argument mention. 2) ***Conceptualiz-***

[*] Equally contribution.
[†] Corresponding authors.

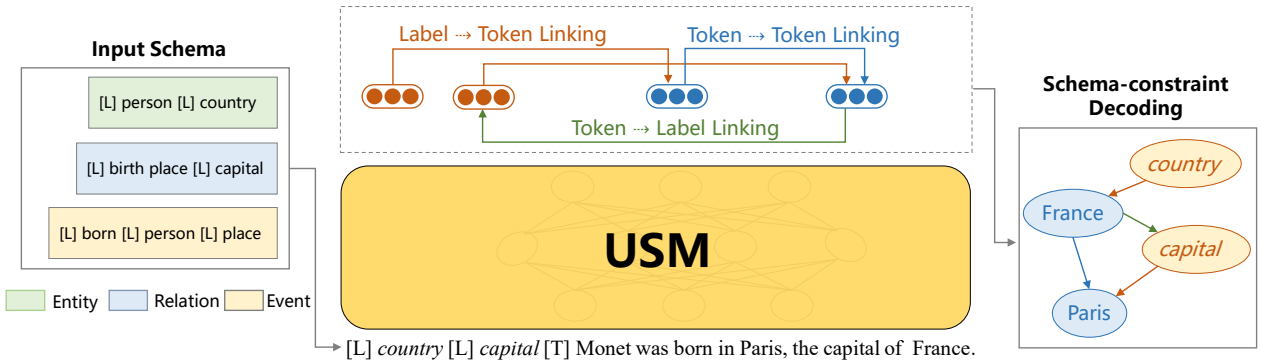

Figure 2: The overall framework of Unified Semantic Matching.

ing, which generalizes utterance and paired substructures to corresponding target semantic concepts. More importantly, these two operations can be explicitly reformulated using a semantic matching paradigm when given a target extraction schema. Specifically, structuring operations can be viewed as building specific kinds of semantic associations between utterances in the input text, while conceptualizing operations can be regarded as matching between target semantic labels and the given utterances or substructures. Consequently, if we universally transform information extraction into combinations of a series of structuring and conceptualizing, reformulate all these operations with the semantic matching between structures and schemas, and jointly learn all IE tasks under the same paradigm, we can easily conduct various kinds of IE tasks with one universal architecture and share knowledge across different tasks and schemas.

Unfortunately, directly conducting semantic matching between structures and schemas is impractical for universal information extraction. First, sentences have many substructures, resulting in a large number of potential matching candidates and a large scale of matching, which makes the computational efficiency of the model unacceptable. Second, the schema of IE is structural and hard to match with the plain text. In this paper, we propose directed token linking for universal IE. The main idea is to transform the structuring and conceptualizing into a series of directed token linking operations, which can be reverted to semantic matching between utterances and schema.

Based on the above observation, we propose USM, a unified semantic matching framework for universal information extraction (UIE), which decomposes structures and verbalizes label types for sharing structuring and conceptualizing abilities. Specifically, we design a set of directed token linking operations (token-token linking, label-token linking, and token-label linking) to decouple task-specific IE tasks into two extraction abilities. To learn the common extraction abilities, we pre-train USM by leveraging heterogeneous supervision from linguistic resources. Compared to previous works, USM is a new transferable, controllable, efficient end-to-end framework for UIE, which jointly encodes extraction schema and input text, uniformly extracts substructures, and controllably decodes target structures on demand.

We conduct experiments on four main IE tasks under the supervised, multi-task, and zero/few-shot transfer settings. The proposed USM framework achieves state-of-the-art results in all settings and solves massive tasks using a single multi-task model. Under the zero/few-shot transfer settings, USM shows a strong cross-type transfer ability due to the shared structuring and conceptualizing obtained by pre-training.

In summary, the main contributions of this paper are:

1. We propose an end-to-end framework for universal information extraction – USM, which can jointly model schema and text, uniformly extract substructures, and controllably generate the target structure on demand.
2. We design three unified token linking operations to decouple various IE tasks, sharing extraction capabilities across different target structures and semantic schemas and achieving "one model for solving all tasks" by multi-task learning.
3. We pre-train a universal foundation model with large-scale heterogeneous supervisions, which can benefit future research on IE.

Unified Semantic Matching via Directed Token Linking

Information extraction is structuring the text's information and elevating it into specific semantic categories. As shown in Figure 2, USM takes the arbitrary extraction label schema l and the raw text t as input and directly outputs the structure according to the given schema. For example, given the text "Monet was born in Paris, the capital of France", USM needs to extract ("France", *capital*, "Paris") for the relation type *capital* and (*person*, "Monet")/(*country*, "France") for the entity type *person* and *country*. The main challenges here are: 1) how to unifiedly extract heterogeneous structures using the shared structuring ability; 2) how to uniformly represent different extraction tasks under diversified label schemas to share the common conceptualizing ability.

In this section, we describe how to end-to-end extract the information structures from the text using USM. Specifically, as shown in Figure 3, USM first verbalizes all label schemas (Levy et al. 2017; Li et al. 2020; Lu et al. 2022) and learns the schema-text joint embedding to build a shared label text semantic space. Then we describe three basic token

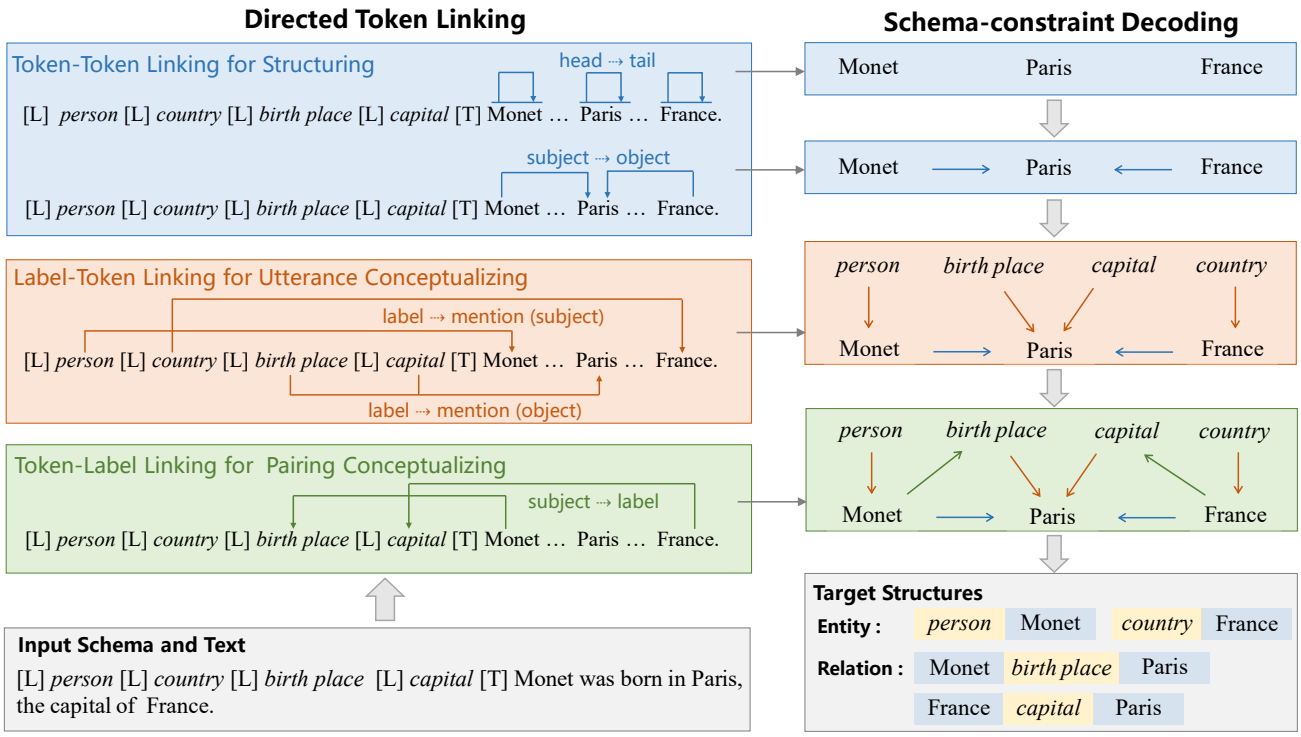

Figure 3: Illustrations of Directed Token Linking. Token-Token Linking structures utterance and association pair substructures from the text, Label-Token Linking conceptualizes the utterance, and Token-Label Linking conceptualizes the association pair. In practice, we employ different label symbols "`[L]`" for utterance conceptualizing: "`[LM]`" for the label of single mention, such as entity types and event trigger types; "`[LP]`" for the predicate of association pair, such as relation types and event argument types.

linking operations and how to structure and conceptualize information from text using these three operations. Finally, we introduce how to decode the final results using schema-constraint decoding.

Schema-Text Joint Embedding

To capture the interaction between label schema and text, USM first learns the joint contextualized embeddings of schema labels and text tokens. Concretely, USM first verbalizes the extraction schema s as token sequence $l = \{l_1, l_2, ..., l_{|l|}\}$ following the structural schema instructor (Lu et al. 2022), then concatenates schema sequence l and text tokens $t = \{t_1, t_2, ..., t_{|t|}\}$ as input, and finally computes the joint label-text embeddings $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_{|l|+|t|}]$ as follow:

$$\mathbf{H} = \text{Encoder}(l_1, l_2, ..., l_{|l|}, t_1, t_2, ..., t_{|t|}, \mathbf{M}) \quad (1)$$

where $\text{Encoder}(\cdot)$ is a transformer encoder, and $\mathbf{M} \in \mathbb{R}^{(|l|+|t|)\times(|l|+|t|)}$ is the mask matrix that determines whether a pair of tokens can be attended to each other.

Token-Token Linking for Structuring

After obtaining the joint label-text embeddings $\mathbf{H} = [\mathbf{h}^l_1, ..., \mathbf{h}^l_{|l|}, \mathbf{h}^t_1, ..., \mathbf{h}^t_{|t|}]$, USM structures all valid substructures using Token-Token Linking (TTL) operations:

1. **Utterance**: a continuous token sequence in the input text, e.g., entity mention "Monet" or event trigger "born in". We extract a single utterance with inner span head-to-tail (H2T) linking, as shown in Figure 3. For example, to extract the span "Monet" and "born in" as valid substructures, USM utilizes H2T to link "Monet" to itself and link "born" to "in".
2. **Association pair**: a basic related pair unit extracted from the text, e.g., relation subject-object pair ("Monet", "Paris") or event trigger-argument ("born in", "Paris"). We extract span pairs with head-to-head (H2H) and tail-to-tail (T2T) linking operations. For example, to extract the subject-object pair "Monet" and "Paris" as a valid substructure, USM links "Monet" and "Paris" using H2H as well as links "Monet" and "Paris" using T2T.

For the above three token-to-token linking (H2T, H2H, T2T) operations, USM respectively calculates the token-to-token linking score $\mathbf{s}_{\text{TTL}}(t_i, t_j)$ over all valid token pair candidates $\langle t_i, t_j \rangle$. For each token pair $\langle t_i, t_j \rangle$, the linking score $\mathbf{s}_{\text{TTL}}(t_i, t_j)$ is calculated as:

$$\mathbf{s}_{\text{TTL}}(t_i, t_j) = \text{FFNN}^l_{\text{TTL}}(\mathbf{h}^t_i)^T \mathbf{R}_{j-i} \text{FFNN}^r_{\text{TTL}}(\mathbf{h}^t_j) \quad (2)$$

where $\text{FFNN}^{l/r}$ are feed-forward layers with output size d. $\mathbf{R}_{j-i} \in \mathbb{R}^{d\times d}$ is the rotary position embedding (Su et al. 2021, 2022) that can effectively inject relative position information into the valid structure mentioned above.

Label-Token Linking for Utterance Conceptualizing

Given label token embeddings $\mathbf{h}^l_1, ..., \mathbf{h}^l_{|l|}$ and text token embeddings $\mathbf{h}^t_1, ..., \mathbf{h}^t_{|t|}$, USM conceptualizes valid utterance structures with label-token linking (LTL) operations. The output of LTL is a pair of label name and text mention, e,g., (*person*, "Monet"), (*country*, "France"), and (*born*, "born in"). There are two types of utterance conceptualizing: the first one is the type of mention, which indicates assigning the label types to every single mention, such as entity type *person* for entity mention "Monet"; the second one is the predicate of object, which assigns the predicate type to each object candidate, such as relation type *birth place* for "Paris" and event argument type *place* for "Paris".

We conceptualize the type of mention and the predicate of object with the same label-to-token linking operation, thus enabling the two label semantics to reinforce each other. Following the head-tail span extraction style, we name each substructure with label-to-head (L2H) and label-to-tail (L2T) linking operations. For the pair of label name *birth place* and text span *Paris*, USM links the head of the label *birth* with the head of text span "Paris" and links the tail of label *place* with the tail of text span "Paris".

For the above two label-to-token linking (L2H, L2T) operations, USM respectively calculates the label-to-token linking score $\mathbf{s}_{\text{LTL}}(l_i, t_j)$ over all valid label and text token pair candidates $\langle l_i, t_j \rangle$:

$$\mathbf{s}_{\text{LTL}}(l_i, t_j) = \text{FFNN}^{\text{label}}_{\text{LTL}}(\mathbf{h}^l_i)^T \mathbf{R}_{j-i} \text{FFNN}^{\text{text}}_{\text{LTL}}(\mathbf{h}^t_j) \quad (3)$$

Token-Label Linking for Pairing Conceptualizing

To conceptualize the association pair, USM links the subject of the association pair to the label name using Token-Label Linking (TLL). Precisely, TLL operation links the subject of triplet and the predicate type with head-to-label (H2L) and tail-to-label (T2L) operations. For instance, TLL links the head of text span "Monet" and the head of the label *birth* with H2L and links the tail of text span "Monet" and the tail of the label *place* with T2L following the head-tail span extraction style. For the above two token-label linking (H2L, T2L) operations, the linking score $\mathbf{s}_{\text{TLL}}(t_i, l_j)$ is computed as:

$$\mathbf{s}_{\text{TLL}}(t_i, l_j) = \text{FFNN}^{\text{text}}_{\text{TLL}}(\mathbf{h}^l_i)^T \mathbf{R}_{j-i} \text{FFNN}^{\text{label}}_{\text{TLL}}(\mathbf{h}^t_j) \quad (4)$$

Schema-constraint Decoding for Structure Composing

USM decodes the final structures using a schema-constraint decoding algorithm, given substructures extracted by unified token linking operations. During the decoding stage, we separate types for different tasks according to the schema definition. For instance, in the joint entity and relation extraction task, we uniformly encode entity types and relation types as labels to utilize the common structuring and conceptualizing ability but compose the final result by separating the entity or relation types from input types.

As shown in Figure 3, USM 1) first decodes mentions and subject-object unit extracted by token-token linking operation: {"Monet", "Paris", "France", ("Monet", "Paris"), ("France", "Paris")}; 2) and then decodes label-mention pairs by label-token linking operation: {(*person*, "Monet"), (*country*, "France"), (*birth place*, "Paris"), (*capital*, "Paris")}; 3) and finally decodes label-association pairs using token-label linking operation: ("Monet", *birth place*), ("France", *capital*). The above three token linking operations do not affect each other; hence the extraction operations are fully non-autoregressive and highly parallel.

Finally, we separate the entity types *country* and *person*, relation types *birth place*, and *capital* from input types according to the schema definition. Based on the result from token-label linking ("Monet", *birth place*), ("France", *capital*), we can consistently obtain the full structure ("Monet", *birth place*, "Paris") and ("France", *capital*, "Paris").

Learning from Heterogeneous Supervision

This section introduces how to leverage heterogeneous supervised resources to learn the common structuring and conceptualizing abilities for unified token linking. Specifically, with the help of verbalized label representation and unified token linking, we unify heterogeneous supervision signals into <text, token pairs> for pre-training. We first pre-train the USM on the heterogeneous resources, which contain three different supervised signals, including task annotation signals (e.g., IE datasets), distant signals (e.g., distant supervision datasets), and indirect signals (e.g., question answering datasets), then adopt the pre-trained USM model to specific downstream information extraction tasks.

Pre-training

USM uniformly encodes label schema and text in the shared semantic representation and employs unified token linking to structure and conceptualize information from text. To help USM to learn the common structuring and conceptualizing abilities, we collect three different supervised signals from existing linguistic sources for the pre-training of USM:

$\mathcal{D}_{\text{task}}$ is the task annotation dataset, where each instance has a gold annotation for information extraction. We use Ontonotes (Pradhan et al. 2013), widely used in the field of information extraction as gold annotation, which contains 18 entity types. $\mathcal{D}_{\text{task}}$ is used as in-task supervision signals to learn task-specific structuring and conceptualizing abilities.

$\mathcal{D}_{\text{distant}}$ is the distant supervision dataset, where each instance is aligned by text and knowledge base. Distant supervision is a common practice to obtain large-scale training data for information extraction (Mintz et al. 2009; Riedel et al. 2013). We employ NYT (Riedel et al. 2013) and Rebel (Huguet Cabot and Navigli 2021) as our distant supervision datasets, which are obtained by aligning text with Freebase and Wikidata, respectively. Rebel dataset has a large label schema, and all verbalized schemas are too long to be concatenated with input text and fed to the pre-trained transformer encoder. We sample negative label schema to construct meta schema (Lu et al. 2022) as label schema for pre-training.

$\mathcal{D}_{\text{indirect}}$ is the indirect supervision dataset, where each instance is derived from other related NLP tasks (Wang, Ning, and Roth 2020; Chen et al. 2022). We utilize reading comprehension datasets from MRQA (Fisch et al. 2019) as our

Dataset	Metric	UIE	Task-specific SOTA Methods		$\text{USM}_{\text{Roberta}}$	USM	$\text{USM}_{\text{Unify}}$
ACE04	Entity F1	86.89	(Lou, Yang, and Tu 2022)	**87.90**	87.79	87.62	87.34
ACE05-Ent	Entity F1	85.78	(Lou, Yang, and Tu 2022)	86.91	86.98	**87.14**	-
CoNLL03	Entity F1	92.99	(Wang et al. 2021b)	**93.21**	92.76	**93.16**	92.97
ACE05-Rel	Relation Strict F1	66.06	(Yan et al. 2021)	66.80	66.54	**67.88**	-
CoNLL04	Relation Strict F1	75.00	(Huguet Cabot and Navigli 2021)	75.40	75.86	**78.84**	77.12
NYT	Relation Boundary F1	93.54	(Huguet Cabot and Navigli 2021)	93.40	93.96	**94.07**	94.01
SciERC	Relation Strict F1	36.53	(Yan et al. 2021)	**38.40**	37.05	37.36	37.42
ACE05-Evt	Event Trigger F1	73.36	(Wang et al. 2022b)	**73.60**	71.68	72.41	72.31
ACE05-Evt	Event Argument F1	54.79	(Wang et al. 2022b)	55.10	55.37	**55.83**	53.57
CASIE	Event Trigger F1	69.33	(Lu et al. 2021)	68.98	70.77	**71.73**	71.56
CASIE	Event Argument F1	61.30	(Lu et al. 2021)	60.37	63.05	**63.26**	63.00
14-res	Sentiment Triplet F1	74.52	(Lu et al. 2022)	74.52	76.35	**77.26**	77.29
14-lap	Sentiment Triplet F1	63.88	(Lu et al. 2022)	63.88	65.46	**65.51**	66.60
15-res	Sentiment Triplet F1	67.15	(Lu et al. 2022)	67.15	68.80	**69.86**	-
16-res	Sentiment Triplet F1	75.07	(Lu et al. 2022)	75.07	76.73	**78.25**	-
AVE-unify	-	71.10	-	71.34	71.83	**72.46**	72.11
AVE-total	-	71.75	-	72.05	72.61	**73.35**	-

Table 1: Overall results of USM on different datasets. AVE-unify indicates the average performance of non-overlapped datasets (except ACE05-Rel/Evt and 15/16-res), and AVE-total indicates the average performance of all datasets.

indirect supervision datasets: HotpotQA (Yang et al. 2018), Natural Questions (Kwiatkowski et al. 2019), NewsQA (Trischler et al. 2017), SQuAD (Rajpurkar et al. 2016) and TriviaQA (Joshi et al. 2017). Compared with limited entity types in $\mathcal{D}_{\text{task}}$ and relation types $\mathcal{D}_{\text{distant}}$, diversified question expressions can provide richer label semantic information for learning conceptualizing. For each (question, context, answer) instance in $\mathcal{D}_{\text{indirect}}$, we take the question as label schema, the context as input text, and the answer as mention. It captures structuring and conceptualizing ability in the pre-training stage by learning token-token and label-token linking operations.

Learning Function

For pre-training, fine-tuning and multi-task learning, we unify all datasets as $\{(x_i, y_i)\}$, where x_i is text and y_i is linking annotation of each token linking pair (TTM, LTM, TLM). We use the same learning function for all settings with the homogenized data format.

The main challenge of USM learning is the sparsity of linked token pairs. The linked ratio only occupies less than 1% of all valid token pair candidates. To overcome the extreme sparsity of linking instances, we optimize class imbalance loss (Su et al. 2022) for each instance as follows:

$$\mathcal{L} = \sum_{m \in \mathcal{M}} \log \left(1 + \sum_{(i,j) \in m^+} e^{-s_m(i,j)} \right) + \log \left(1 + \sum_{(i,j) \in m^-} e^{s_m(i,j)} \right) \tag{5}$$

where $\mathcal{M}$ denotes linking types of USM, m^+ indicates the linked pairs, m^- indicates the non-linked pairs, and $s_m(i,j)$ is the predicate linking score for the linking operation m.

Experiments

This section conducts massive experiments under supervised settings and transfer settings to demonstrate the effectiveness of the proposed unified semantic matching framework.

Experiments on Supervised Settings

We conduct supervised experiments on extensive information extraction tasks, including 4 tasks and 13 datasets (entity extraction, relation extraction, event extraction, sentiment extraction) and their combinations (e.g., joint entity-relation extraction). The used datasets includes ACE04 (Mitchell et al. 2005), ACE05 (Walker et al. 2006); CoNLL03 (Tjong Kim Sang and De Meulder 2003), CoNLL04 (Roth and Yih 2004), SciERC (Luan et al. 2018), NYT (Riedel, Yao, and McCallum 2010), CASIE (Satyapanich, Ferraro, and Finin 2020), SemEval-14/15/16 (Pontiki et al. 2014, 2015, 2016). We employ the same end-to-end settings and evaluation metrics as Lu et al. (2022).

We compare the proposed USM framework with the task-specific state-of-the-art methods and the unified structure generation method – UIE (Lu et al. 2022). For our approach, we show three different settings:

- USM is the pre-trained model which learned unified token linking ability from heterogeneous supervision;
- $\text{USM}_{\text{Roberta}}$ is the initial model of the pre-trained USM, which employs RoBERTa-Large (Liu et al. 2019) as the pre-trained transformer encoder;
- $\text{USM}_{\text{Unify}}$ is initialized by the pre-trained USM and conducts multi-task learning with all datasets but ignores overlapped datasets: ACE05-Ent/Rel and 15/16-res.

For the $\text{USM}_{\text{Roberta}}$ and USM settings, we fine-tune them on each specific task separately. We run each experiment with three seeds and report their average performance.

Table 1 shows the overall performance of USM and other baselines on the 13 datasets, where AVE-unify indicates the average performance of non-overlapped datasets, and AVE-total indicates the average performance of all datasets. We

	Movie	Restaurant	Social	AI	Literature	Music	Politics	Science	Ave
Performance on Unseen Label Subset of $\mathcal{D}_t$ and $\mathcal{D}_i$									
#Unseen/#All	12/12	7/8	7/10	10/14	8/12	9/13	5/9	13/17	-
$\mathcal{D}_{\text{task}}$	25.07	2.50	22.54	10.82	50.74	44.11	9.75	13.98	22.44
$\mathcal{D}_{\text{task}}$ + $\mathcal{D}_{\text{indirect}}$	37.73	14.73	29.34	28.18	56.00	44.93	36.10	44.09	36.39
Performance on Unseen Label Subset of Pre-training Dataset									
#Unseen/#All	10/12	7/8	6/10	8/14	7/12	8/13	4/9	12/17	-
$\mathcal{D}_{\text{task}}$	32.10	2.50	1.64	10.68	52.42	45.93	11.16	14.12	21.32
$\mathcal{D}_{\text{task}}$ + $\mathcal{D}_{\text{indirect}}$	39.76	14.73	20.62	24.12	56.24	44.21	32.92	44.25	34.61
$\mathcal{D}_{\text{task}}$ + $\mathcal{D}_{\text{distant}}$	35.35	21.10	40.64	27.57	56.97	49.29	43.72	44.05	39.84
$\mathcal{D}_{\text{task}}$ + $\mathcal{D}_{\text{distant}}$ + $\mathcal{D}_{\text{indirect}}$	42.11	26.01	44.37	34.91	65.69	60.07	56.65	55.26	48.13
Δ	10.01	23.51	42.73	24.23	13.27	14.14	45.49	41.14	26.82

Table 2: Performance of Zero-shot transfer settings on unseen entity label subset with different supervision signals. Unseen indicates label types that do not appear in the pre-training dataset. **Δ** indicates the improvement of pre-training using extra supervision signals ($\mathcal{D}_{\text{distant}}$ and $\mathcal{D}_{\text{indirect}}$).

	CoNLL04	Model Size
GPT-3	18.10	175B
DEEPSTRUCT	25.80	10B
USM	**25.95**	356M

Table 3: Performance of Zero-shot transfer settings on relation extraction. * GPT-3 175B indicates formulating the extraction task as a question answering problem through prompting, and DEEPSTRUCT 10B is a pre-trained language model for structure prediction (Wang et al. 2022a)

can observe that: 1) *By verbalizing labels and modeling all IE tasks as unified token linking, USM provides a novel and effective framework for IE.* USM achieves state-of-the-art performance and outperforms the strong task-specific methods by 1.30 in AVE-total. Even without pre-training, $\text{USM}_{\text{Roberta}}$ also shows strong performance, which indicates the strong portability and generalization ability of unified token linking. 2) *Heterogeneous supervision provides a better foundation for structuring and conceptualizing information extraction.* Compared to the initial model $\text{USM}_{\text{Roberta}}$ and the pre-trained model USM, the heterogeneous pre-training achieved an average 0.74 improvement across all datasets. 3) *By homogenizing diversified label schemas and heterogeneous target structures into the unified token sequence, USM_{Unify} can solve massive IE tasks with a single multi-task model.* $\text{USM}_{\text{Unify}}$ outperforms task-specific state-of-the-art methods with different model architectures and encoder backbones in average, providing an efficient solution for application and deployment.

Experiments on Zero-shot Transfer Settings

We conduct zero-shot cross-type transfer experiments on 9 datasets across various domains to verify the transferable conceptualization learned by USM. In this setting, we directly employ the pre-trained USM to conduct extraction on new datasets.

	Model	1-Shot	5-Shot	10-Shot	AVE-S
Entity CoNLL03	UIE-Large*	57.53	75.32	79.12	70.66
	$\text{USM}_{\text{Roberta}}$	9.69	40.66	62.87	37.74
	$\text{USM}_{\text{Symbolic}}$	60.56	81.87	83.87	75.43
	USM	**71.11**	**83.25**	**84.58**	**79.65**
Relation CoNLL04	UIE-Large*	34.88	51.64	58.98	48.50
	$\text{USM}_{\text{Roberta}}$	0.00	12.81	31.02	14.61
	$\text{USM}_{\text{Symbolic}}$	13.45	48.31	58.91	40.22
	USM	**36.17**	**53.20**	**60.99**	**50.12**
Event Trigger ACE05-Evt	UIE-Large*	**42.37**	53.07	54.35	49.93
	$\text{USM}_{\text{Roberta}}$	26.39	47.10	51.46	41.65
	$\text{USM}_{\text{Symbolic}}$	1.97	30.77	52.30	28.35
	USM	40.86	**55.61**	**58.79**	**51.75**
Event Argument ACE05-Evt	UIE-Large*	14.56	31.20	35.19	26.98
	$\text{USM}_{\text{Roberta}}$	6.47	27.00	34.20	22.56
	$\text{USM}_{\text{Symbolic}}$	0.08	13.71	33.52	15.77
	USM	**19.01**	**36.69**	**42.48**	**32.73**
Sentiment 16res	UIE-Large*	23.04	42.67	53.28	39.66
	$\text{USM}_{\text{Roberta}}$	2.68	35.71	48.56	28.98
	$\text{USM}_{\text{Symbolic}}$	20.08	41.25	50.90	37.41
	USM	**30.81**	**52.06**	**58.29**	**47.05**

Table 4: Few-shot results on end-to-end IE tasks. For a fair comparison, we conduct text-structure pre-training from T5-v1.1-large using the same pre-training corpus of USM, refer to UIE-Large*.

For entity extraction, the cross-type extraction datasets include Movie (MIT-Movie), Restaurant (MIT-Restaurant) (Liu et al. 2013), Social (WNUT-16) (Strauss et al. 2016), and AI/Literature/Music/Politics/Science from CrossNER (Liu et al. 2021). We investigate the effect of different supervised signals in the zero-shot entity extraction setting. $\mathcal{D}_{task}$ indicates we first train USM on the common entity extraction dataset – Ontonotes, then directly conduct extraction on the new types, which emulates the most common label transfer method used in real-world scenarios. To be consistent with the real scenario, we select the best checkpoint according to the F1 score on the dev set of $\mathcal{D}_{task}$.

For zero-shot relation extraction, we compare USM with

the following strong baselines:

- GPT-3 175B (Brown et al. 2020) is a large-scale, generative pre-trained model, which can extract entity and relation by formulating the task as a question answering problem through prompting (Wang et al. 2022a).
- DEEPSTRUCT 10B is a structured prediction model pre-trained on six large-scale entity, relation, and triple datasets (Wang et al. 2022a).

Table 2 shows the entity extraction performance on the unseen label subset, in which types are not appearing in the pre-training dataset. And Table 3 shows the performance of zero-shot relation extraction on CoNLL04. From Table 2 and Table 3, we can see that: 1) *USM has a strong zero-shot transferability across labels.* USM shows good migration performance on Movie, Literature, and Music domains even when learning from $\mathcal{D}_{\text{task}}$ with limited entity types. For relation extraction, USM (356M) outperforms the strong zero-shot baseline GPT-3 (175B) and DEEPSTRUCTURE (10B) with a smaller model size. 2) *Heterogeneous supervision boosts USM with unified label semantics and outperforms the task annotation baseline by a large margin.* Compared to the task annotation baseline ($\mathcal{D}_{\text{task}}$), USM significantly and consistently improves the performance on all datasets.

Experiments on Few-shot Transfer Settings

To further investigate the effects of verbalized label semantics, we conduct few-shot transfer experiments on four IE tasks and compare USM with the following baselines:

- **UIE-large*** is the pre-trained sequence-to-structure model for effective low-resource IE tasks, which injects label semantics by generating labels and words in structured extraction language synchronously and guiding the generation with a structural schema instructor.
- **USM$_{\text{Roberta}}$** is the initial model of USM, which directly use Roberta-large as the pre-trained encoder;
- **USM$_{\text{Symbolic}}$** replaces the names of labels with symbolic representation (meaning-less labels, e.g., label1, label2, ...) during the fine-tuning stage of USM, which is used to verify the effect of verbalized label semantics.

For few-shot transfer experiments, we follow the data splits and settings with the previous work (Lu et al. 2022) and repeat each experiment 10 times to avoid the influence of random sampling (Huang et al. 2021). Table 4 shows the performance on 4 IE tasks under the few-shot settings, where AVE-S is the average performance of 1/5/10-shot experiments. We can see that: 1) *By modeling IE tasks via unified semantic matching, USM exceeds the few-shot state-of-the-art UIE-large 5.11 on average.* Although UIE also adopts verbalized label representation, this structure generation method needs to learn to generate the novel schema word in the target structure during transfer learning. In contrast, USM only needs to learn to match them, providing a better inductive bias and leading to a much smaller decoding search space. The pre-trained unified token linking ability boosts the USM in all settings. 2) *It is crucial to verbalize label schemas rather than meaningless symbols, especially for complex extraction tasks.* USM$_{\text{Symbolic}}$, which uses symbolic labels instead of verbalized labels, drastically reduces performance on all tasks. For tasks with more semantic types, such as event extraction with 33 types, the performance drops significantly, even lower than that of USM$_{\text{Roberta}}$ initialized directly with Roberta-large.

Related Work

In the past decade, due to powerful representation ability, deep learning methods (Bengio et al. 2003; Collobert et al. 2011) have made amazing achievements in information extraction tasks. Most of these methods decompose extraction into multiple sub-tasks and follow the classical neural classifier method (Krizhevsky, Sutskever, and Hinton 2012) to model each sub-task, such as entity extraction, relation classification, event trigger detection, event argument classification, etc. And several architectures are proposed to model the extraction, such as sequence tagging (Lample et al. 2016; Zheng et al. 2017), span classification (Sohrab and Miwa 2018; Song et al. 2019; Wadden et al. 2019), table filling (Gupta, Schütze, and Andrassy 2016; Wang and Lu 2020), question answering (Levy et al. 2017; Li et al. 2020), and token pair (Wang et al. 2020; Yu et al. 2021).

Recently, to solve various IE tasks with a single architecture, UIE employs unified structure generation, models the various IE tasks with structured extraction language, and pre-trains the ability of structure generation using distant text-structure supervision (Lu et al. 2022). Unlike the generation-based approach, we model universal information extraction as unified token linking, which reduces the search space during decoding and leads to better generalization performance. Beyond distant supervision, we further introduce indirect supervision from related NLP tasks to learn the unified token linking ability.

Similar to USM in this paper, matching-based IE approaches aim to verbalize the label schema and structure candidate to achieve better generalization (Liu et al. 2022). Such methods usually use pre-extracted syntactic structures (Wang et al. 2021a) and semantic structures (Huang et al. 2018) as candidate structures, then model the extraction as text entailment (Obamuyide and Vlachos 2018; Sainz et al. 2021; Lyu et al. 2021; Sainz et al. 2022) and semantic structure mapping (Chen and Li 2021; Dong, Pan, and Luo 2021). Different from the pre-extraction and matching style, this paper decouples various IE tasks to unified token linking operations and designs a one-pass end-to-end information extraction framework for modeling all tasks.

Conclusion

In this paper, we propose a unified semantic matching framework – USM, which jointly encodes extraction schema and input text, uniformly extracts substructures in parallel, and controllably decodes target structures on demand. Experimental results show that USM achieves state-of-the-art performance under the supervised experiments and shows strong generalization ability under zero/few-shot transfer settings, which verifies USM is a novel, transferable, controllable, and efficient framework. For future work, we want to extend USM to NLU tasks, e.g., text classification, and investigate more indirect supervision signals for IE, e.g., text entailment.

Acknowledgments

We sincerely thank the reviewers for their insightful comments and valuable suggestions. This work is supported by the National Key Research and Development Program of China (No.2020AAA0109400) and the Natural Science Foundation of China (No.62122077, 61876223, and 62106251). Hongyu Lin is sponsored by CCF-Baidu Open Fund.

References

Alvarez-Melis, D.; and Jaakkola, T. 2017. A causal framework for explaining the predictions of black-box sequence-to-sequence models. In *Proc. of EMNLP*.

Andersen, P. M.; Hayes, P. J.; Weinstein, S. P.; Huettner, A. K.; Schmandt, L. M.; and Nirenburg, I. B. 1992. Automatic Extraction of Facts from Press Releases to Generate News Stories. In *Proc. of ANLP*.

Bengio, Y.; Ducharme, R.; Vincent, P.; and Janvin, C. 2003. A Neural Probabilistic Language Model. *J. Mach. Learn. Res.*

Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; Agarwal, S.; Herbert-Voss, A.; Krueger, G.; Henighan, T.; Child, R.; Ramesh, A.; Ziegler, D.; Wu, J.; Winter, C.; Hesse, C.; Chen, M.; Sigler, E.; Litwin, M.; Gray, S.; Chess, B.; Clark, J.; Berner, C.; McCandlish, S.; Radford, A.; Sutskever, I.; and Amodei, D. 2020. Language Models are Few-Shot Learners. In *Proc. of NeurIPS*.

Chen, C.-Y.; and Li, C.-T. 2021. ZS-BERT: Towards Zero-Shot Relation Extraction with Attribute Representation Learning. In *Proc. of NAACL*.

Chen, M.; Huang, L.; Li, M.; Zhou, B.; Ji, H.; and Roth, D. 2022. New Frontiers of Information Extraction. In *Proc. of NAACL*.

Collobert, R.; Weston, J.; Bottou, L.; Karlen, M.; Kavukcuoglu, K.; and Kuksa, P. 2011. Natural Language Processing (Almost) from Scratch. *J. Mach. Learn. Res.*

Dong, M.; Pan, C.; and Luo, Z. 2021. MapRE: An Effective Semantic Mapping Approach for Low-resource Relation Extraction. In *Proc. of EMNLP*.

Fisch, A.; Talmor, A.; Jia, R.; Seo, M.; Choi, E.; and Chen, D. 2019. MRQA 2019 Shared Task: Evaluating Generalization in Reading Comprehension. In *Proc. of MRQA*.

Grishman, R. 2019. Twenty-five years of information extraction. *Natural Language Engineering*.

Gupta, P.; Schütze, H.; and Andrassy, B. 2016. Table Filling Multi-Task Recurrent Neural Network for Joint Entity and Relation Extraction. In *Proc. of COLING*.

Huang, J.; Li, C.; Subudhi, K.; Jose, D.; Balakrishnan, S.; Chen, W.; Peng, B.; Gao, J.; and Han, J. 2021. Few-Shot Named Entity Recognition: An Empirical Baseline Study. In *Proc. of EMNLP*.

Huang, L.; Ji, H.; Cho, K.; Dagan, I.; Riedel, S.; and Voss, C. 2018. Zero-Shot Transfer Learning for Event Extraction. In *Proc. of ACL*.

Huguet Cabot, P.-L.; and Navigli, R. 2021. REBEL: Relation Extraction By End-to-end Language generation. In *Proc. of EMNLP Findings*.

Joshi, M.; Choi, E.; Weld, D.; and Zettlemoyer, L. 2017. TriviaQA: A Large Scale Distantly Supervised Challenge Dataset for Reading Comprehension. In *Proc. of ACL*.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Proc. of NeurIPS*.

Kwiatkowski, T.; Palomaki, J.; Redfield, O.; Collins, M.; Parikh, A.; Alberti, C.; Epstein, D.; Polosukhin, I.; Devlin, J.; Lee, K.; Toutanova, K.; Jones, L.; Kelcey, M.; Chang, M.-W.; Dai, A. M.; Uszkoreit, J.; Le, Q.; and Petrov, S. 2019. Natural Questions: A Benchmark for Question Answering Research. *Transactions of the Association for Computational Linguistics*.

Lample, G.; Ballesteros, M.; Subramanian, S.; Kawakami, K.; and Dyer, C. 2016. Neural Architectures for Named Entity Recognition. In *Proc. of NAACL*.

Levy, O.; Seo, M.; Choi, E.; and Zettlemoyer, L. 2017. Zero-Shot Relation Extraction via Reading Comprehension. In *Proc. of CoNLL*.

Li, X.; Feng, J.; Meng, Y.; Han, Q.; Wu, F.; and Li, J. 2020. A Unified MRC Framework for Named Entity Recognition. In *Proc. of ACL*.

Liu, F.; Lin, H.; Han, X.; Cao, B.; and Sun, L. 2022. Pretraining to Match for Unified Low-shot Relation Extraction. In *Proc. of ACL*.

Liu, J.; Pasupat, P.; Cyphers, S.; and Glass, J. 2013. Asgard: A portable architecture for multilingual dialogue systems. In *Proc. of ICASSP*.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *CoRR*.

Liu, Z.; Xu, Y.; Yu, T.; Dai, W.; Ji, Z.; Cahyawijaya, S.; Madotto, A.; and Fung, P. 2021. CrossNER: Evaluating Cross-Domain Named Entity Recognition. *Proc. of AAAI*.

Lou, C.; Yang, S.; and Tu, K. 2022. Nested Named Entity Recognition as Latent Lexicalized Constituency Parsing. In *Proc. of ACL*.

Lu, Y.; Lin, H.; Xu, J.; Han, X.; Tang, J.; Li, A.; Sun, L.; Liao, M.; and Chen, S. 2021. Text2Event: Controllable Sequence-to-Structure Generation for End-to-end Event Extraction. In *Proc. of ACL*.

Lu, Y.; Liu, Q.; Dai, D.; Xiao, X.; Lin, H.; Han, X.; Sun, L.; and Wu, H. 2022. Unified Structure Generation for Universal Information Extraction. In *Proc. of ACL*.

Luan, Y.; He, L.; Ostendorf, M.; and Hajishirzi, H. 2018. Multi-Task Identification of Entities, Relations, and Coreference for Scientific Knowledge Graph Construction. In *Proc. of EMNLP*.

Lyu, Q.; Zhang, H.; Sulem, E.; and Roth, D. 2021. Zero-shot Event Extraction via Transfer Learning: Challenges and Insights. In *Proc. of ACL*.

Mintz, M.; Bills, S.; Snow, R.; and Jurafsky, D. 2009. Distant supervision for relation extraction without labeled data. In *Proc. of ACL*.

Mitchell, A.; Strassel, S.; Huang, S.; and Zakhary, R. 2005. ACE 2004 Multilingual Training Corpus. In *LDC*.

Obamuyide, A.; and Vlachos, A. 2018. Zero-shot Relation Classification as Textual Entailment. In *Proc. of FEVER*.

Pontiki, M.; Galanis, D.; Papageorgiou, H.; Androutsopoulos, I.; Manandhar, S.; AL-Smadi, M.; Al-Ayyoub, M.; Zhao, Y.; Qin, B.; De Clercq, O.; Hoste, V.; Apidianaki, M.; Tannier, X.; Loukachevitch, N.; Kotelnikov, E.; Bel, N.; Jiménez-Zafra, S. M.; and Eryiğit, G. 2016. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In *Proc. of SemEval*.

Pontiki, M.; Galanis, D.; Papageorgiou, H.; Manandhar, S.; and Androutsopoulos, I. 2015. SemEval-2015 Task 12: Aspect Based Sentiment Analysis. In *Proc. of SemEval*.

Pontiki, M.; Galanis, D.; Pavlopoulos, J.; Papageorgiou, H.; Androutsopoulos, I.; and Manandhar, S. 2014. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In *Proc. of SemEval*.

Pradhan, S.; Moschitti, A.; Xue, N.; Ng, H. T.; Björkelund, A.; Uryupina, O.; Zhang, Y.; and Zhong, Z. 2013. Towards Robust Linguistic Analysis using OntoNotes. In *Proc. of CoNLL*.

Rajpurkar, P.; Zhang, J.; Lopyrev, K.; and Liang, P. 2016. SQuAD: 100,000+ Questions for Machine Comprehension of Text. In *Proc. of EMNLP*.

Riedel, S.; Yao, L.; and McCallum, A. 2010. Modeling Relations and Their Mentions without Labeled Text. In *Machine Learning and Knowledge Discovery in Databases*.

Riedel, S.; Yao, L.; McCallum, A.; and Marlin, B. M. 2013. Relation Extraction with Matrix Factorization and Universal Schemas. In *Proc. of NAACL*.

Roth, D.; and Yih, W.-t. 2004. A Linear Programming Formulation for Global Inference in Natural Language Tasks. In *Proc. of CoNLL*.

Sainz, O.; Gonzalez-Dios, I.; Lopez de Lacalle, O.; Min, B.; and Agirre, E. 2022. Textual Entailment for Event Argument Extraction: Zero- and Few-Shot with Multi-Source Learning. In *Proc. of ACL Findings*.

Sainz, O.; Lopez de Lacalle, O.; Labaka, G.; Barrena, A.; and Agirre, E. 2021. Label Verbalization and Entailment for Effective Zero and Few-Shot Relation Extraction. In *Proc. of EMNLP*.

Satyapanich, T.; Ferraro, F.; and Finin, T. 2020. CASIE: Extracting Cybersecurity Event Information from Text. In *Proc. of AAAI*.

Sohrab, M. G.; and Miwa, M. 2018. Deep Exhaustive Model for Nested Named Entity Recognition. In *Proc. of EMNLP*.

Song, L.; Zhang, Y.; Gildea, D.; Yu, M.; Wang, Z.; and Su, J. 2019. Leveraging Dependency Forest for Neural Medical Relation Extraction. In *Proc. of EMNLP-IJCNLP*.

Strauss, B.; Toma, B.; Ritter, A.; de Marneffe, M.-C.; and Xu, W. 2016. Results of the WNUT16 Named Entity Recognition Shared Task. In *Proc. of WNUT*.

Su, J.; Lu, Y.; Pan, S.; Murta, A.; Wen, B.; and Liu, Y. 2021. RoFormer: Enhanced Transformer with Rotary Position Embedding. In *CoRR*.

Su, J.; Murtadha, A.; Pan, S.; Hou, J.; Sun, J.; Huang, W.; Wen, B.; and Liu, Y. 2022. Global Pointer: Novel Efficient Span-based Approach for Named Entity Recognition. In *CoRR*.

Tjong Kim Sang, E. F.; and De Meulder, F. 2003. Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition. In *Proc. of CoNLL*.

Trischler, A.; Wang, T.; Yuan, X.; Harris, J.; Sordoni, A.; Bachman, P.; and Suleman, K. 2017. NewsQA: A Machine Comprehension Dataset. In *Proc. of RepL4NLP*.

Wadden, D.; Wennberg, U.; Luan, Y.; and Hajishirzi, H. 2019. Entity, Relation, and Event Extraction with Contextualized Span Representations. In *Proc. of EMNLP*.

Walker, C.; Strassel, S.; Medero, J.; and Maeda, K. 2006. ACE 2005 Multilingual Training Corpus. In *LDC*.

Wang, C.; Liu, X.; Chen, Z.; Hong, H.; Tang, J.; and Song, D. 2021a. Zero-Shot Information Extraction as a Unified Text-to-Triple Translation. In *Proc. of EMNLP*.

Wang, C.; Liu, X.; Chen, Z.; Hong, H.; Tang, J.; and Song, D. 2022a. DeepStruct: Pretraining of Language Models for Structure Prediction. In *Proc. of ACL Findings*.

Wang, J.; and Lu, W. 2020. Two are Better than One: Joint Entity and Relation Extraction with Table-Sequence Encoders. In *Proc. of EMNLP*.

Wang, K.; Ning, Q.; and Roth, D. 2020. Learnability with Indirect Supervision Signals. In *Proc. of NeurIPS*.

Wang, S.; Yu, M.; Chang, S.; Sun, L.; and Huang, L. 2022b. Query and Extract: Refining Event Extraction as Type-oriented Binary Decoding. In *Proc. of ACL Findings*.

Wang, X.; Jiang, Y.; Bach, N.; Wang, T.; Huang, Z.; Huang, F.; and Tu, K. 2021b. Improving Named Entity Recognition by External Context Retrieving and Cooperative Learning. In *Proc. of ACL*.

Wang, Y.; Yu, B.; Zhang, Y.; Liu, T.; Zhu, H.; and Sun, L. 2020. TPLinker: Single-stage Joint Extraction of Entities and Relations Through Token Pair Linking. In *Proc. of COLING*.

Yan, Z.; Zhang, C.; Fu, J.; Zhang, Q.; and Wei, Z. 2021. A Partition Filter Network for Joint Entity and Relation Extraction. In *Proc. of EMNLP*.

Yang, Z.; Qi, P.; Zhang, S.; Bengio, Y.; Cohen, W.; Salakhutdinov, R.; and Manning, C. D. 2018. HotpotQA: A Dataset for Diverse, Explainable Multi-hop Question Answering. In *Proc. of EMNLP*.

Yu, B.; Wang, Y.; Liu, T.; Zhu, H.; Sun, L.; and Wang, B. 2021. Maximal Clique Based Non-Autoregressive Open Information Extraction. In *Proc. of EMNLP*.

Zheng, S.; Wang, F.; Bao, H.; Hao, Y.; Zhou, P.; and Xu, B. 2017. Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme. In *Proc. of ACL*.