RESIN-11: Schema-guided Event Prediction for 11 Newsworthy Scenarios

Xinya Du[1], Zixuan Zhang[1], Sha Li[1], Ziqi Wang[1], Pengfei Yu[1], Hongwei Wang[1], Tuan Manh Lai[1], Xudong Lin[2], Iris Liu[3], Ben Zhou[4], Haoyang Wen[1], Manling Li[1], Darryl Hannan[5], Jie Lei[5], Hyounghun Kim[5], Rotem Dror[4], Haoyu Wang[4], Michael Regan[3], Qi Zeng[1], Qing Lyu[4], Charles Yu[1], Carl Edwards[1], Xiaomeng Jin[1], Yizhu Jiao[1], Ghazaleh Kazeminejad[3], Zhenhailong Wang[1], Chris Callison-Burch[4], Carl Vondrick[2], Mohit Bansal[5], Dan Roth[4], Jiawei Han[1], Shih-Fu Chang[2], Martha Palmer[3], Heng Ji[1]

[1] University of Illinois Urbana-Champaign [2] Columbia University [3] University of Colorado, Boulder
[4] University of Pennsylvania [5] University of North Carolina at Chapel Hill
{xinyadu2,zixuan11,shal2,hengji}@illinois.edu

Abstract

We introduce RESIN-11, a new schema-guided event extraction and prediction system that can be applied to a large variety of newsworthy scenarios. The framework consists of two parts: (1) an open-domain end-to-end multimedia multilingual information extraction system with weak-supervision and zero-shot learning-based techniques. (2) a schema matching and schema-guided event prediction system based on our curated schema library. We build a demo website[1] based on our dockerized system and schema library publicly available for installation[2]. We also include a video demonstrating the system.[3]

1 Introduction

If the evening news discusses a migration of people due to drought and the infrastructure cannot handle the population influx, what will happen next? While annotated datasets have fueled progress in machine intelligence, the knowledge required for event forecasting is vast and potentially ambiguous. For example, to predict that a rebellion is likely in the future, models need to integrate background events (drought), abstractions (strained infrastructure causes unrest), and event schemas (structure and duration of rebellion). Existing link prediction (Zhang and Chen, 2018; Wang et al., 2018; Lei et al., 2019) or knowledge graph completion methods (Zhang et al., 2019; Goel et al., 2020; Wang et al., 2021a) cannot meet this goal because the event instance graphs extracted from news are often incomplete and noisy. Recent work (Li et al., 2020b, 2021a) proposes to leverage complex event schemas (stereotypical evolution pattern of complex events) for event prediction. However, these methods are often limited to a few scenarios due to the lack of high-quality, open-domain information extraction systems to construct event instance graphs needed for schema induction.

To tackle these challenges, in this paper we use the event schemas discovered at scale to guide the learning of predictive models. We first identify 11 newsworthy scenarios, and construct comprehensive hierarchical schemas through the combination of automatic schema induction and manual curation. Then we develop an open-domain end-to-end multimedia multilingual information extraction system that can extract entities, relations, events, and temporal orders for all of these scenarios, based on a series of weak supervision based methods, including few-shot learning and lifelong learning. Compared with previous event tracking systems (Wen et al., 2021) that conduct graph matching on a linearized sequence, we propose a new schema matching algorithm that directly operates on graphs. We also proposed a event prediction model trained with self-supervision to predict possible missing events to form a coherent story. Our contributions include:

- We induce and curate hierarchical schemas for 11 scenarios, capturing a wide coverage of newsworthy events.
- We extend our multi-lingual multi-media information extraction techniques (Wen et al., 2021) to handle the open-domain extraction setting.
- We develop a new schema matching and prediction algorithm that is capable of recovering missing events and predicting events that will likely to happen in the future.

[1]Demo: `http://18.221.187.153:11000`
[2]Github: `https://github.com/RESIN-KAIROS/RESIN-11`
[3]Video: `https://screencast-o-matic.com/watch/c3nlhnVbeyg`

Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: System Demonstrations, pages 54 - 63

2 Methodology

Overview Figure 1 illustrates the overall architecture of our framework. It includes three parts: (1) schema library construction (Section 2.1); (2) open-domain multimedia multilingual Information Extraction (IE) system (Section 2.2-2.3) ; (3) schema matching and prediction component (Section 2.4).

We first perform schema induction and curation for 11 identified newsworthy scenarios. Specifically, we use GPT-3 generated results as an outline, enrich the schema with the help of the WikiData ontology and expand the steps for better coverage.

On the IE side, we assume our input consists of multilingual multimedia document clusters about a specific scenario (e.g., disease outbreak). Each document cluster contains documents about a specific complex event scenario (e.g., COVID-19 pandemic). Our textual IE pipeline takes documents and transcribed speech as input and extracts entity, relation and event mentions (Section 2.2). In order to extend IE to open-domain, we have adopted weak supervision and zero-shot transfer learning techniques. Then we perform cross-document cross-lingual entity and event coreference resolution, and link them to WikiData. The extracted events are then ordered by temporal relation extraction. Our visual pipeline extracts events and arguments from visual signals (i.e., images and videos), and link the extracted knowledge elements to our extracted graph using cross-media event coreference resolution (Section 2.3). Finally, our system automatically selects the schema from a schema library that best matches the extracted IE graph and new events are predicted (Section 2.4).

2.1 Schema Induction and Curation

For our schema library creation, we first start with creating schemas with a zero-shot approach utilizing GPT-3. Given a scenario for which we want to create a schema, we generate multiple texts that discuss the topically-related complex events using the OpenAI GPT-3 API[4] with the Davinci-instruct-beta-v3 model. We use three prompting methods to generate documents of diverse genres such as news articles, Wikihow-style documents, and step-by-step event description. One example of such a prompt and the generated output is shown in Figure 2. Then we identify the events mentioned in the texts and link the events to WikiData Qnodes. We use generated documents instead of real documents because we observe that generated documents are generally cleaner and contain a higher percentage of events that can be linked. For instance, comparing the generated text and the crawled news articles for IED attacks, the generated text contains 0.13 events per token and the news articles only contain 0.06 events per token. From there we add arguments to the events, and identify the temporal and hierarchical relations between the events effectively converting each text into a graph structure (Figure 3).

Note that the automatic induced schema is often noisy and has limited coverage. Some of such mistakes come from the GPT3 generation, e.g., in the generated output in Figure 2, it omits how the disease was discovered. Other mistakes root from incorrect prediction of temporal relations, such as the “Kill” → “Come (Attack)” edge in Figure 3.

To improve coverage, human curators further check Wikipedia and news articles on the related topics and add more events. Three other crucial aspects of human curation are (1) entity coreference resolution, (2) temporal ordering and (3) hierarchical structure construction. Entity coreference chains in schemas often involve implicit entities, such as the “area where the sick live” in step 2. This location entity is futher coreferential with the “contaminated areas” entity mentioned later in step 4. The generated output is a list of steps, which can be converted to a linear ordering of events. However, some events can happen concurrently such as the “Educate” event in step 4 and all other events. In addition, some events have strong semantic coherence involving the same set of entities and thus can be grouped together. For example, this chain of “Identify-Quarantine-Disinfect” can be seen as a medical response to one batch of infections. We refer to this as a sub-schema and this medical response sub-schema can be repeated with a different set of patients and medical agents. An example of the human curated schema is shown in Figure 4.

In the curation process, we use a web-based graphical interface (Mishra et al., 2021) to help visualize and assess schemas.

2.2 Open Domain IE from Speech and Text

We first convert speech data into text using the Amazon Transcribe API[5]. When the language is not specified, it is automatically detected from the audio signal. It returns the transcription with start-

[4] https://openai.com/blog/openai-api/

[5] https://aws.amazon.com/transcribe/

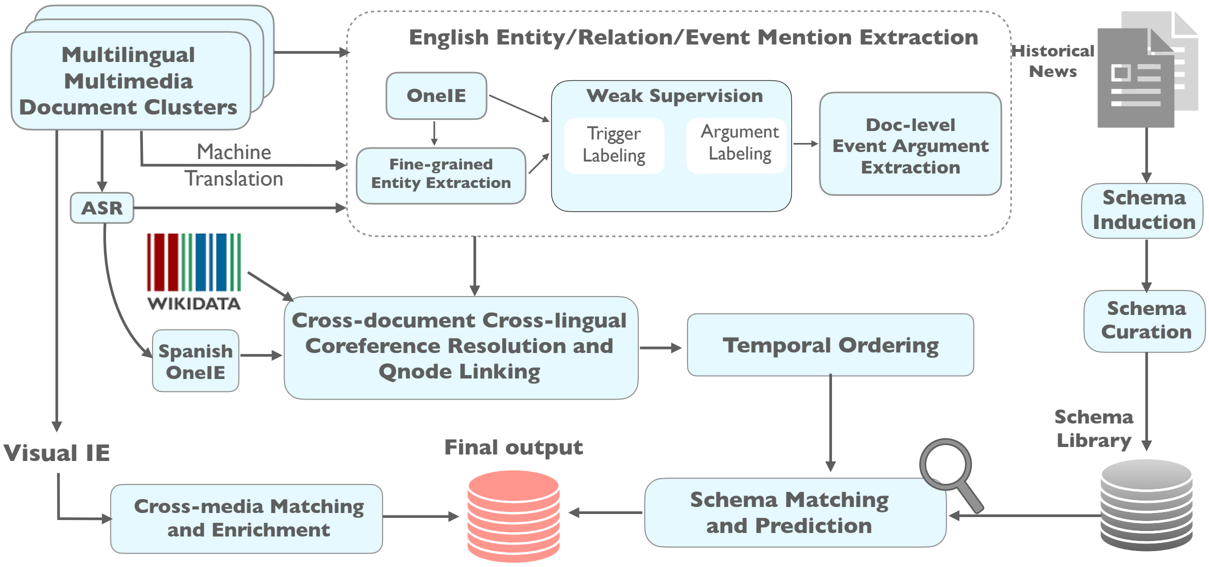

Figure 1: The architecture of RESIN-11 framework: including (1) open-domain multimedia multilingual information extraction; and (2) schema matching and prediction.

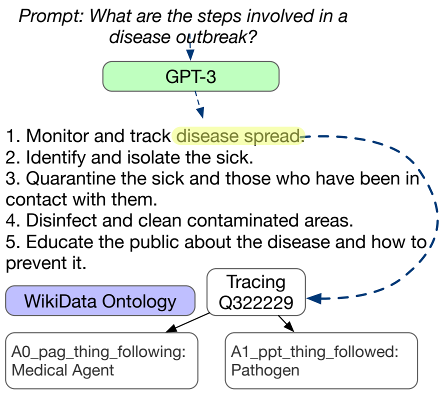

Figure 2: An illustration of the schema curation process. The steps are actual output from GPT-3.

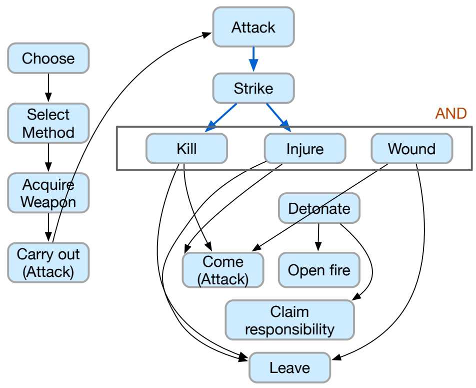

Figure 3: An automatically induced schema for the terrorist attack scenario with model predicted temporal order, hierarchical relations and logical relations. Arguments are omitted for clarity.

ing and ending times for each detected words, as well as potential alternative transcriptions.

To achieve wide coverage of event types from 11 scenarios, our information extraction system consists of 3 components: (1) the supervised joint entity, relation and event extraction model OneIE (Lin et al., 2020); (2) weakly supervised keyword-guided event extraction; and (3) zero-shot generation-based argument extraction (Li et al., 2021b).

OneIE is a joint neural model for sentence-level information extraction. Given a sentence, the goal of this module is to extract an information graph $G = (V, E)$, where V is the node set containing entity mentions and event triggers and E is the edge set containing entity relations and event-argument links. In order to capture the interactions among knowledge elements, we incorporate schema-guided global features when decoding information graphs. After we extract these mentions, we apply a syntactic parser (Honnibal et al., 2020) to extend mention head words to their extents. Based on OneIE relations output, we perform rule-based relation enrichment to obtain fine-grained relation subsubtypes. We collect keywords for various fine-grained types, then we construct rules by checking keywords in Shortest Dependency Paths (SDP) between two relation entities.

To extract emergent event types for which we do not have large-scale annotation, we employ a keyword-based event detection system. Specifically, we select a list of keywords for each new

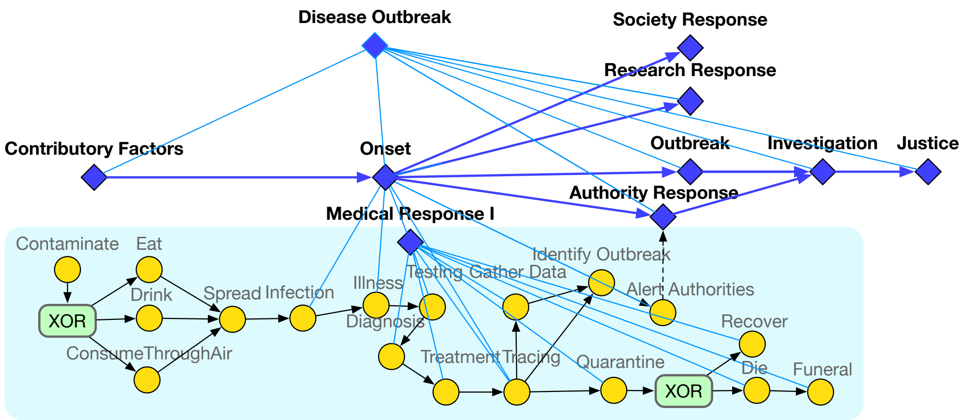

Figure 4: The curated schema for the disease outbreak scenario. Blue diamond shapes represent sub-schemas and yellow circles represent primitive events. Black arrows between primitive events represent temporal order, light blue lines between the primitive events and the sub-schema node represent event-subevent hierarchical relationship. Here we only show the primitive events under the Onset sub-schema.

event type, and search for the occurrences of these keywords in a text corpus. We compute keyword representations by averaging the contextualized representations from BERT (Devlin et al., 2019) of keyword occurrences, and cluster keyword representations for the same event type to get a set of cluster representations for each event type. For event trigger detection, we first compute BERT representations of all the tokens in a sentence, and consider a token as an event trigger if its cosine similarity with some cluster representations of an event type is larger than a threshold. We tuned the threshold using a few example event mentions.

After identifying the event triggers, we further employ a document-level event argument extraction model (Li et al., 2021b) to improve the recall of event argument role labeling. This model formulates the argument extraction problem as conditional text generation. The condition consists of the original document and a blank event template. For example, the template for `Transportation` event type is *arg1 transported arg2 in arg3 from arg4 place to arg5 place*. To apply this model in a zero-shot setting, we create new templates for the emerging event types and use them as input.

For entity linking over Wikidata, we directly use the EPGEL system proposed in Lai et al. (2022). For cross-document cross-lingual coreference resolution, we follow the approach of (Wen et al., 2021). After the coreference resolution/entity linking stage, we conduct temporal ordering for all of the extracted events. First we provide two independent temporal ordering results from two learning-based pairwise event order classification systems (Zhou et al., 2021; Wen and Ji, 2021). To make the prediction consistent and valid over each document cluster, we use a greedy algorithm that selects conflict-free predicted temporal relations to the final instance graph sequentially based on their confidence scores. Similar to Wen et al. (2021), these two results will be used for schema matching and event prediction and only the best prediction will be used in the final output.

2.3 Cross-media Info Grounding and Fusion

Visual event and argument role extraction Our goal is to extract structured visual events and entities. Specifically, given an image or a video segment, the desired output are its event type and the associated argument roles. Due to the expensive cost of event annotation for images and videos, it is not feasible to perform annotation for each new type. Unlike existing systems leveraging supervised training (Chen et al., 2021a), we propose an open-domain framework to enable the visual event extraction for a broader spectrum of event types.

Our proposed system is composed of two complementary models. The first model is a supervised model based on a large-scale image dataset, Situation with Groundings (SWiG) (Pratt et al., 2020). We manually define the mapping that covers 16 event types and use the model pretrained on the SWiG dataset to extract event and argument roles. The second model is an unsupervised model by leveraging large-scale vision-language pretrained model (Li et al., 2022; Radford et al., 2021). We conduct further pretraining on an event-rich corpus (Li et al., 2022) by adding an additional pretraining task of event structure alignment between two modalities. In detail, we extract event structures from captions and utilize them as the supervision for image event extraction. The pretraining corpus comprises multiple scenarios, providing support for the extraction of events for a wide range of scenarios. To process images and videos in a unified manner, we follow Wen et al. (2021) to sample frames at a frame rate of 1 frame per second from videos and process these key-frames as individual images.

Cross-media event coreference resolution To augment the text event graph, we leverage a weakly-supervised coreference resolution model (Chen et al., 2021a) that is trained based on the alignment between video frames and speech texts on a large collection of news videos to predict the relevance between a textual event and the extracted visual event. Once the relevance is higher than a threshold, we leverage a rule-based approach to decide whether the visual event mention and a textual event mention are coreferential: (1) Matched event types; (2) No contradiction of entity types for the same argument role in different modalities. This pipeline enables adding provenance of visual-only arguments into the event graph, which provides more comprehensive event understanding.

2.4 Schema Matching and Prediction

After obtaining a large-scale library of schema graphs for various scenarios, our goal is to instantiate the schema graphs with extracted events, and then use it for schema-guided event prediction.

Schema Matching To match the event nodes between the IE graphs and schema graphs, previous work (Wen et al., 2021) first linearizes graphs into event sequences and then conducts event matching using longest common subsequences (LCS). However, such a sequence-based matching algorithm cannot well capture some global dependencies in a graph point of view, and the performance largely depends on the quality of event temporal ordering results. Also, the LCS based matching algorithm can only handle the cases where the events in schema graph and IE graph use an identical ontology (i.e., the same category of event types), which is however not applicable for open-domain settings since the names of events could be diverse and multifarious.

To tackle these problems, we propose a new schema matching algorithm that directly operates on each pair of instance graph I and schema graph S. We formulate schema matching as an integer programming problem, where we can use an assignment matrix $X \in \mathbb{R}^{|I|\times|S|}$ to represent the matching results. To enable matching between events with different names, we compute the pairwise Synsets similarities from WordNet[6] and store it into a node similarity matrix A. For each event e_i in a given instance graph I, we obtain the set of all reachable event nodes $\mathcal{R}_I(i) = \{e \mid P_I(e_i, e) = 1, e \in I\}$, where $P_I(e_i, e)$ denotes whether there exists a path from e_i to e in the instance graph I. Similarly, we can also obtain the reachable event sets $\mathcal{R}_S$ and P_S for the schema graph S. For edge similarity between each pair of events e_i and e_j, instead of strictly judging whether they are temporal neighbors (i.e., whether there exists an event-event temporal link between e_i and e_j) , we only use the reachability as temporal constraints (i.e., whether $e_j \in \mathcal{R}_I(i)$) to mitigate the high dependence on the quality of event temporal ordering results. Specifically, we aim to find the optimal solution X_{opt}

$$X_{opt} = \arg\max_X \sum_{i,j} A_{i,j} X_{i,j} - c \cdot \mathcal{Q}(X), \qquad (1)$$

where c is a hyper-parameter and $\mathcal{Q}(X)$ denotes the penalty term for the violation of temporal constraints. The penalty term $\mathcal{Q}(X)$ is defined as the total number of event pairs that violates the temporal constraints.

Schema-guided Event Prediction After schema matching, an instance graph I is mapped to a subgraph of the schema graph, i.e., $I' \subseteq S$. The next step is to determine whether a candidate event node $e \in S \backslash I'$ is a missing node for I'. Specifically, we aim to learn a function $f(e, I') : S \times 2^S \mapsto [0, 1]$, which outputs the probability that event node e is missing for subgraph I'. We consider two factors when designing the function $f(e, I')$: (1) Neighbors of e and I'. We use a graph neural network (GNN) to aggregate neighbor information and learn embedding vectors for nodes in S, then aggregate embeddings of nodes in I' to obtain the embedding of I'. The embeddings of e and I' are concatenated, followed by an MLP to output the predicted probability. (2) Paths. We identify all paths that connect e and each node in I' in the schema graph S, then aggregate the paths together to obtain the bag-of-path feature for the pair of (e, I'). The bag-of-path feature is fed into another MLP to output

[6] https://www.nltk.org/howto/wordnet.html

Scenario	# Episodes	# Events	# Ents	# Rels
Business Change	18	81	24	54
Civil Unrest	6	34	18	24
Disease Outbreak	19	102	27	93
Election	8	35	14	33
International Conflict	17	95	56	50
Kidnapping	9	66	15	56
Mass Shooting	8	37	13	31
Sports Events	4	17	14	19
Terrorist Attacks	8	36	11	26
Disaster/Manmade Disaster	8	38	10	29
Disaster/Natural Disaster	4	23	8	18
IED/General Attack	19	52	40	22
IED/General IED	10	48	18	39
IED/Drone Strikes	10	50	19	43
IED/Backpack IED	10	49	18	40
IED/Roadside IED	10	48	19	39
IED/Car IED	10	50	19	43

Table 1: Statistics of our schema library.

the predicted probability. Finally, the outputs of the above two modules are averaged as the final prediction.

3 Schema and Experiments

The overall statistics of our schema library are presented in Table 1. The performance of each component is shown in Table 2. We evaluate the performance of our full system on a complex event corpus (LDC2022E02), which contains multi-lingual multi-media document clusters. We train our mention extraction component on ACE05 (Walker et al., 2006) and ERE (Song et al., 2015); document-level argument extraction on ACE05 and RAMS (Ebner et al., 2020); coreference component on ACE05, EDL 2016 (LDC2017E03), EDL 2017 (LDC2017E52), OntoNotes (Pradhan et al., 2012), ERE, CoNLL 2002 (Tjong Kim Sang, 2002), DCEP (Dias, 2016) and SemEval10 (Recasens et al., 2010); temporal ordering component on MATRES (Ning et al., 2018); weakly supervised event extraction on ACE05; schema matching and prediction on LDC2022E03; visual event and argument extraction on M2E2 (Li et al., 2020a) and cross-media event coreference on Video M2E2 (Chen et al., 2021a). For coreference resolution, similar to previous work (Wen et al., 2021), we use the CoNLL metric.

4 Related Work

Event Schema Induction and Curation Event schemas, or otherwise known as scripts, are structures that represent typical event progressions (Schank and Abelson, 1975). Prior to this work, there has been some effort in creating schema databases through crowdsourcing (Regneri et al., 2010; Modi et al., 2016; Wanzare et al., 2016; Sakaguchi et al., 2021). The key characteristics that separate our schema library from exiting resources include (1) focus on diverse newsworthy scenarios instead of everyday events; (2) highly structured multi-level schema organization.

In addition to schema resources, there has also been work on automating the schema induction process, through the use of probabilistic graphical models (Chambers, 2013; Cheung et al., 2013; Nguyen et al., 2015; Weber et al., 2018) and event-based language models (Pichotta and Mooney, 2016; Modi and Titov, 2014; Rudinger et al., 2015; Li et al., 2020b, 2021a). We hope that our schema library can serve as a resource for the development of better automatic schema induction methods.

Weakly-Supervised Event Extraction Due to the high cost of annotating event instances, low resource event extraction has received much attention in recent years. There are a variety of settings explored, including zero-shot transfer learning (Lyu et al., 2021; Huang et al., 2018), cross-lingual trans-

Component			Benchmark	Metric	Score
Weakly-supervised IE		Trigger	ACE	F1	63.3
		Argument	ACE	F1	41.5
Mention Extraction	En	Trigger	ACE+ERE	F1	64.1
		Argument	ACE+ERE	F1	49.7
		Relation	ACE+ERE	F1	49.5
	Es	Trigger	ACE+ERE	F1	63.4
		Argument	ACE+ERE	F1	46.0
		Relation	ACE+ERE	F1	46.6
Document-level Argument Extraction			ACE	F1	66.7
			RAMS	F1	48.6
Coreference Resolution	En	Entity	OntoNotes	CoNLL	92.4
		Event	ACE	CoNLL	84.8
	Es	Entity	SemEval 2010	CoNLL	67.6
		Event	ERE-ES	CoNLL	81.0
Wikidata QNode Linking			TACKBP-2010	Acc.	90.9
Temporal Ordering		RoBERTa	MATRES	F1	81.7
		T5	MATRES-b	Acc.	89.6
Visual Event Extraction			M2E2	F1	52.7
Cross-media Event Coreference			Video M2E2	F1	51.5

	Benchmark	Metric	Score
Schema Matching	LDC2022E03	Recall	63.2
Schema Prediction	WikiEvents	F1	45.5

Table 2: Performance (%) of each component our open-domain multimedia multilingual IE system (upper) and schema matching and prediction component (bottom).

fer (Subburathinam et al., 2019), inducing event types (Huang et al., 2016; Wang et al., 2021b), keyword-based supervision (Zhang et al., 2021) and few-shot learning (Peng et al., 2016; Lai et al., 2020; Shen et al., 2021; Cong et al., 2021; Chen et al., 2021b).

Schema-Guided Event Prediction Our schema-guided event prediction model is also related to link prediction (Zhang and Chen, 2018; Wang et al., 2018; Lei et al., 2019) and graph completion (Zhang et al., 2019; Goel et al., 2020; Wang et al., 2021a) methods. The advantages of our schema-guided method are that: (1) Our method is specially designed for multiple small event graphs, rather than a single large graph as studied in previous work. Therefore, using event schema enables us to model the common pattern of instance event graphs. (2) An event schema can be seen as a pool of inter-connected candidate events, which provides new event nodes that can be added into an incomplete instance event graph. However, existing work can only predict missing links rather than missing nodes.

5 Conclusions and Future Work

We build an open-domain schema-guided event prediction system that is capable of extracting and predicting structured information regarding events from various scenarios. In the future, we plan to further improve both the extraction quality and portability to cover even more scenarios, and use the automatic zero-shot schema induction algorithm to iteratively extend our curated schemas. The hierarchy structure of our event schemas can also be further utilized to improve future event prediction.

6 Broader Impact

The goal of this project is to advance the state-of-the-art schema-guided information extraction and event prediction from real-world multi-modal news sources. We believe that grounding our work in real-world applications will help us make progress in event-centric natural language understanding. However, this work is not void of possible improper use that may have adverse social impacts.

One major distinction between beneficial use and harmful use depends on the data sources. Proper use of the technology requires that input sources are legally and ethically obtained. As an instance of beneficial use, our demo may contribute to disease outbreak monitoring and disaster emergency response, which is included in our chosen scenarios. Besides, we should also be aware of the possible biases that may exists in the datasets. Our system components, as well as pretrained language models that we use, are trained and evaluated on specific benchmark datasets, which could be affected by such biases. For example, as is observed in Abid et al. (2021), the text generated by GPT-3 might include undesired social biases. Our careful human curators' effort involved in the schema library building can mitigate this issue.

Generally, increasing transparency and explainability of models can help prevent social harm, such as over-estimation of the model ability. We plan to make our software fully open source for public audition and verification. We are also open to explore countermeasures to prevent unintended consequences.

The event prediction part of our model is able to forecast the future trend of the current complex event, which enables us to better understand the structure and semantics of complex events. Moreover, it is particularly helpful for us to analyze and predict the public opinion.

Acknowledgement

We thank the anonymous reviewers helpful suggestions. This research is based upon work supported by U.S. DARPA KAIROS Program No. FA8750-19-2-1004, U.S. DARPA AIDA Program No. FA8750-18-2-0014 and LORELEI Program No. HR0011-15-C-0115. 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 DARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

References

Abubakar Abid, Maheen Farooqi, and James Zou. 2021. Large language models associate muslims with violence. *Nature Machine Intelligence*, 3(6):461–463.

Nathanael Chambers. 2013. Event schema induction with a probabilistic entity-driven model. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP2013)*, volume 13, pages 1797–1807.

Brian Chen, Xudong Lin, Christopher Thomas, Manling Li, Shoya Yoshida, Lovish Chum, Heng Ji, and

Shih-Fu Chang. 2021a. Joint multimedia event extraction from video and article. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 74–88, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Jiawei Chen, Hongyu Lin, Xianpei Han, and Le Sun. 2021b. Honey or poison? solving the trigger curse in few-shot event detection via causal intervention. *EMNLP*.

Jackie Chi Kit Cheung, Hoifung Poon, and Lucy Vanderwende. 2013. Probabilistic frame induction. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 837–846, Atlanta, Georgia. Association for Computational Linguistics.

Xin Cong, Shiyao Cui, Bowen Yu, Tingwen Liu, Wang Yubin, and Bin Wang. 2021. Few-Shot Event Detection with Prototypical Amortized Conditional Random Field. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 28–40, Online. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Francisco Dias. 2016. Multilingual Automated Text Anonymization. Msc dissertation, Instituto Superior Técnico, Lisbon, Portugal, May.

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*.

Rishab Goel, Seyed Mehran Kazemi, Marcus Brubaker, and Pascal Poupart. 2020. Diachronic embedding for temporal knowledge graph completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3988–3995.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python. *Zenodo*.

Lifu Huang, Taylor Cassidy, Xiaocheng Feng, Heng Ji, Clare R. Voss, Jiawei Han, and Avirup Sil. 2016. Liberal event extraction and event schema induction. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 258–268, Berlin, Germany. Association for Computational Linguistics.

Lifu Huang, Heng Ji, Kyunghyun Cho, Ido Dagan, Sebastian Riedel, and Clare Voss. 2018. Zero-shot transfer learning for event extraction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2160–2170, Melbourne, Australia. Association for Computational Linguistics.

Tuan Manh Lai, Heng Ji, and ChengXiang Zhai. 2022. Improving candidate retrieval with entity profile generation for wikidata entity linking. *arXiv preprint arXiv:2202.13404*.

Viet Dac Lai, Thien Huu Nguyen, and Franck Dernoncourt. 2020. Extensively matching for few-shot learning event detection. In *Proceedings of the First Joint Workshop on Narrative Understanding, Storylines, and Events*, pages 38–45, Online. Association for Computational Linguistics.

Kai Lei, Meng Qin, Bo Bai, Gong Zhang, and Min Yang. 2019. Gcn-gan: A non-linear temporal link prediction model for weighted dynamic networks. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pages 388–396. IEEE.

Manling Li, Sha Li, Zhenhailong Wang, Lifu Huang, Kyunghyun Cho, Heng Ji, Jiawei Han, and Clare Voss. 2021a. The future is not one-dimensional: Complex event schema induction by graph modeling for event prediction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 5203–5215.

Manling Li, Ruochen Xu, Shuohang Wang, Luowei Zhou, Xudong Lin, Chenguang Zhu, Michael Zeng, Heng Ji, and Shih-Fu Chang. 2022. Clip-event: Connecting text and images with event structures. *arXiv preprint arXiv:2201.05078*.

Manling Li, Alireza Zareian, Qi Zeng, Spencer Whitehead, Di Lu, Heng Ji, and Shih-Fu Chang. 2020a. Cross-media structured common space for multimedia event extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2557–2568, Online. Association for Computational Linguistics.

Manling Li, Qi Zeng, Ying Lin, Kyunghyun Cho, Heng Ji, Jonathan May, Nathanael Chambers, and Clare Voss. 2020b. Connecting the dots: Event graph schema induction with path language modeling. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 684–695.

Sha Li, Heng Ji, and Jiawei Han. 2021b. Document-level event argument extraction by conditional generation. In *Proc. The 2021 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT2021)*.

Ying Lin, Heng Ji, Fei Huang, and Lingfei Wu. 2020. A joint end-to-end neural model for information extraction with global features. In *Proc. The 58th Annual*

Meeting of the Association for Computational Linguistics (ACL2020).

Qing Lyu, Hongming Zhang, Elior Sulem, and Dan Roth. 2021. Zero-shot event extraction via transfer learning: Challenges and insights. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 322–332, Online. Association for Computational Linguistics.

Piyush Mishra, Akanksha Malhotra, Susan Windisch Brown, Martha Palmer, and Ghazaleh Kazeminejad. 2021. A graphical interface for curating schemas. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations*, pages 159–166, Online. Association for Computational Linguistics.

Ashutosh Modi, Tatjana Anikina, Simon Ostermann, and Manfred Pinkal. 2016. InScript: Narrative texts annotated with script information. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3485–3493, Portorož, Slovenia. European Language Resources Association (ELRA).

Ashutosh Modi and Ivan Titov. 2014. Inducing neural models of script knowledge. In *Proceedings of the Eighteenth Conference on Computational Natural Language Learning*, pages 49–57.

Kiem-Hieu Nguyen, Xavier Tannier, Olivier Ferret, and Romaric Besançon. 2015. Generative event schema induction with entity disambiguation. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 188–197.

Qiang Ning, Hao Wu, and Dan Roth. 2018. A multi-axis annotation scheme for event temporal relations. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1318–1328, Melbourne, Australia. Association for Computational Linguistics.

Haoruo Peng, Yangqiu Song, and Dan Roth. 2016. Event detection and co-reference with minimal supervision. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 392–402, Austin, Texas. Association for Computational Linguistics.

Karl Pichotta and Raymond J Mooney. 2016. Learning statistical scripts with lstm recurrent neural networks. In *Thirtieth AAAI Conference on Artificial Intelligence*.

Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina, and Yuchen Zhang. 2012. Conll-2012 shared task: Modeling multilingual unrestricted coreference in ontonotes. In *Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning - Proceedings of the Shared Task: Modeling Multilingual Unrestricted Coreference in OntoNotes, EMNLP-CoNLL 2012, July 13, 2012, Jeju Island, Korea*, pages 1–40. ACL.

Sarah Pratt, Mark Yatskar, Luca Weihs, Ali Farhadi, and Aniruddha Kembhavi. 2020. Grounded situation recognition. In *European Conference on Computer Vision*, pages 314–332. Springer.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pages 8748–8763. PMLR.

Marta Recasens, Lluís Màrquez, Emili Sapena, M. Antònia Martí, Mariona Taulé, Véronique Hoste, Massimo Poesio, and Yannick Versley. 2010. Semeval-2010 task 1: Coreference resolution in multiple languages. In *Proceedings of the 5th International Workshop on Semantic Evaluation, SemEval@ACL 2010, Uppsala University, Uppsala, Sweden, July 15-16, 2010*, pages 1–8. The Association for Computer Linguistics.

Michaela Regneri, Alexander Koller, and Manfred Pinkal. 2010. Learning script knowledge with web experiments. In *ACL*.

Rachel Rudinger, Pushpendre Rastogi, Francis Ferraro, and Benjamin Van Durme. 2015. Script induction as language modeling. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1681–1686.

Keisuke Sakaguchi, Chandra Bhagavatula, Ronan Le Bras, Niket Tandon, Peter Clark, and Yejin Choi. 2021. proScript: Partially ordered scripts generation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2138–2149, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Roger C Schank and Robert P Abelson. 1975. Scripts, plans, and knowledge. In *IJCAI*, volume 75, pages 151–157.

Shirong Shen, Tongtong Wu, Guilin Qi, Yuan-Fang Li, Gholamreza Haffari, and Sheng Bi. 2021. Adaptive knowledge-enhanced Bayesian meta-learning for few-shot event detection. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2417–2429, Online. Association for Computational Linguistics.

Zhiyi Song, Ann Bies, Stephanie Strassel, Tom Riese, Justin Mott, Joe Ellis, Jonathan Wright, Seth Kulick, Neville Ryant, and Xiaoyi Ma. 2015. From light to rich ere: annotation of entities, relations, and events. In *Proceedings of the the 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 89–98.

Ananya Subburathinam, Di Lu, Heng Ji, Jonathan May, Shih-Fu Chang, Avirup Sil, and Clare Voss. 2019. Cross-lingual structure transfer for relation and event extraction. 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 313–325, Hong Kong, China. Association for Computational Linguistics.

Erik F. Tjong Kim Sang. 2002. Introduction to the CoNLL-2002 shared task: Language-independent named entity recognition. In *COLING-02: The 6th Conference on Natural Language Learning 2002 (CoNLL-2002)*.

Christopher Walker, Stephanie Strassel, Julie Medero, and Kazuaki Maeda. 2006. Ace 2005 multilingual training corpus. *Linguistic Data Consortium, Philadelphia*, 57.

Hongwei Wang, Hongyu Ren, and Jure Leskovec. 2021a. Relational message passing for knowledge graph completion. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 1697–1707.

Hongwei Wang, Fuzheng Zhang, Min Hou, Xing Xie, Minyi Guo, and Qi Liu. 2018. Shine: Signed heterogeneous information network embedding for sentiment link prediction. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 592–600.

Ziqi Wang, Xiaozhi Wang, Xu Han, Yankai Lin, Lei Hou, Zhiyuan Liu, Peng Li, Juanzi Li, and Jie Zhou. 2021b. CLEVE: Contrastive Pre-training for Event Extraction. 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 6283–6297, Online. Association for Computational Linguistics.

Lilian D. A. Wanzare, Alessandra Zarcone, Stefan Thater, and Manfred Pinkal. 2016. A crowdsourced database of event sequence descriptions for the acquisition of high-quality script knowledge. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3494–3501, Portorož, Slovenia. European Language Resources Association (ELRA).

Noah Weber, Leena Shekhar, Niranjan Balasubramanian, and Nathanael Chambers. 2018. Hierarchical quantized representations for script generation. *ACL*.

Haoyang Wen and Heng Ji. 2021. Utilizing relative event time to enhance event-event temporal relation extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 10431–10437, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Haoyang Wen, Ying Lin, Tuan Lai, Xiaoman Pan, Sha Li, Xudong Lin, Ben Zhou, Manling Li, Haoyu Wang, Hongming Zhang, et al. 2021. Resin: A dockerized schema-guided cross-document cross-lingual cross-media information extraction and event tracking system. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstrations*, pages 133–143.

Hongming Zhang, Haoyu Wang, and Dan Roth. 2021. Zero-shot Label-aware Event Trigger and Argument Classification. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1331–1340, Online. Association for Computational Linguistics.

Muhan Zhang and Yixin Chen. 2018. Link prediction based on graph neural networks. *Advances in Neural Information Processing Systems*, 31:5165–5175.

Shuai Zhang, Yi Tay, Lina Yao, and Qi Liu. 2019. Quaternion knowledge graph embeddings. In *Advances in Neural Information Processing Systems*, pages 2735–2745.

Ben Zhou, Kyle Richardson, Qiang Ning, Tushar Khot, Ashish Sabharwal, and Dan Roth. 2021. Temporal reasoning on implicit events from distant supervision. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1361–1371, Online. Association for Computational Linguistics.