RESIN: A Dockerlized Schema-Guided Cross-document Cross-lingual Cross-media Information Extraction and Event Tracking System

Haoyang Wen[1], Ying Lin[1], Tuan M. Lai[1], Xiaoman Pan[1], Sha Li[1], Xudong Lin[2]
Ben Zhou[3], Manling Li[1], Haoyu Wang[3], Hongming Zhang[3], Xiaodong Yu[3]
Alexander Dong[3], Zhenhailong Wang[1], Yi R. Fung[1], Piyush Mishra[4], Qing Lyu[3]
Dídac Surís[2], Brian Chen[2], Susan W. Brown[4], Martha Palmer[4], Chris Callison-Burch[3]
Carl Vondrick[2], Jiawei Han[1], Dan Roth[3], Shih-Fu Chang[2], Heng Ji[1]
[1] University of Illinois at Urbana-Champaign [2] Columbia University
[3] University of Pennsylvania [4] University of Colorado, Boulder
hengji@illinois.edu,sc250@columbia.edu,danroth@seas.upenn.edu

Abstract

We present a new information extraction system that can automatically construct temporal event graphs from a collection of news documents from multiple sources, multiple languages (English and Spanish for our experiment), and multiple data modalities (speech, text, image and video). The system advances state-of-the-art from two aspects: (1) extending from sentence-level event extraction to cross-document cross-lingual cross-media event extraction, coreference resolution and temporal event tracking; (2) using human curated event schema library to match and enhance the extraction output. We have made the dockerlized system publicly available for research purpose at GitHub[1], with a demo video[2].

1 Introduction

Event extraction and tracking technologies can help us understand real-world events described in the overwhelming amount of news data, and how they are inter-connected. These techniques have already been proven helpful in various application domains, including news analysis (Glavaš and Štajner, 2013; Glavaš et al., 2014; Choubey et al., 2020), aiding natural disaster relief efforts (Panem et al., 2014; Zhang et al., 2018; Medina Maza et al., 2020), financial analysis (Ding et al., 2014, 2016; Yang et al., 2018; Jacobs et al., 2018; Ein-Dor et al., 2019; Özbayoglu et al., 2020) and healthcare monitoring (Raghavan et al., 2012; Jagannatha and Yu, 2016; Klassen et al., 2016; Jeblee and Hirst, 2018).

However, it's much more difficult to remember event-related information compared to entity-related information. For example, most people in the United States will be able to answer the question "Which city is Columbia University located in?", but very few people can give a complete answer to "Who died from COVID-19?". Progress in natural language understanding and computer vision has helped automate some parts of event understanding but the current, *first-generation*, automated event understanding is overly simplistic since most methods focus on sentence-level sequence labeling for event extraction. Existing methods for complex event understanding also lack of incorporating knowledge in the form of a repository of abstracted event schemas (complex event templates), understanding the progress of time via temporal event tracking, using background knowledge, and performing global inference and enhancement.

To address these limitations, in this paper we will demonstrate a new end-to-end open-source dockerlized research system to extract temporally ordered events from a collection of news documents from multiple sources, multiple languages (English and Spanish for our experiment), and multiple data modalities (speech, text, image and video). Our system consists of a pipeline of components that involve schema-guided complex entity, relation and event extraction, coreference resolution, temporal event tracking and cross-media grounding. Event schemas encode knowledge of stereotypical structures of events and their connections. Our end-to-end system has been dockerlized and made publicly available for research purpose.

2 Approach

2.1 Overview

The architecture of our system is illustrated in Figure 1. Our system extracts information from multilingual multimedia document clusters. Each document cluster contains documents about a specific

[1]Github: https://github.com/RESIN-KAIROS/RESIN-pipeline-public
[2]Video: http://blender.cs.illinois.edu/software/resin/resin.mp4

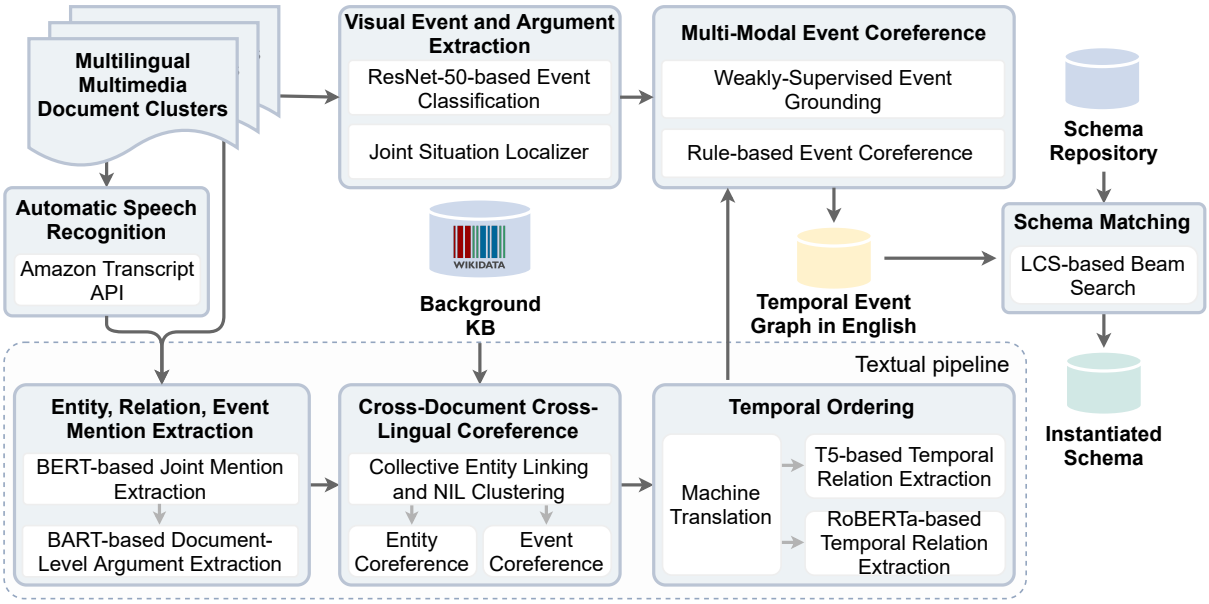

Figure 1: The architecture of RESIN schema-guided information extraction and temporal event tracking system.

complex event. Our textual pipeline takes input from texts and transcribed speeches. It first extracts entity, relation and event mentions (Section 2.2-2.3) and then perform cross-document cross-lingual entity and event coreference (Section 2.4). The extracted events are then ordered by temporal relation extraction (Section 2.5). Our visual pipeline takes images and videos as input and extracts events and arguments from visual signals and ground the extracted knowledge elements onto our extracted graph via cross-media event coreference resolution (Section 2.6). Finally, our system selects the schema from a schema repository that best matches the extracted IE graph and merges these two graphs (Section 2.7). Our system can extract 24 types of entities, 46 types of relations and 67 types of events as defined in the DARPA KAIROS[3] ontology.

2.2 Joint Entity, Relation and Event Mention Extraction and Linking from Speech and Text

For speech input, we apply the Amazon Transcribe API [4] for converting English and Spanish speech to text. When the language is not specified, it is automatically detected from the audio signal. It returns the transcription with starting and ending times for each detected words, as well as potential alternative transcriptions.

Then from the speech recognition results and text input, we extract entity, relation, and event mentions using OneIE (Lin et al., 2020), a state-of-the-art 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. We use a pre-trained BERT encoder (Devlin et al., 2018) to obtain contextualized word representations for the input sentence. Next, we adopt separate conditional random field-based taggers to identify entity mention and event trigger spans from the sentence. We represent each span, or node in the information graph, by averaging vectors of words in the span. After that, we calculate the label scores for each node or edge using separate task-specific feed forward networks. In order to capture the interactions among knowledge elements, we incorporate schema-guided global features when decoding information graphs. For a candidate graph G, we define a global feature vector $\boldsymbol{f} = \{f_1(G), ..., f_M(G)\}$, where $f_i(\cdot)$ is a function that evaluates whether G matches a specific global feature. We compute the global feature score as $\boldsymbol{uf}$, where $\boldsymbol{u}$ is a learnable weight vector. Finally, we use a beam search-based decoder to generate the information graph with the highest global score. After we extract these mentions, we apply a syntactic parser (Honnibal et al., 2020) to extend mention head words to their extents. Then

[3] https://www.darpa.mil/program/knowledge-directed-artificial-intelligence-\reasoning-over-schemas

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

we apply a cross-lingual entity linker (Pan et al., 2017) to link entity mentions to WikiData (Vrandečić and Krötzsch, 2014)[5].

2.3 Document-level Event Argument Extraction

The previous module can only operate on the sentence level. In particular, event arguments can often be found in neighboring sentences. To make up for this, we further develop a document-level event argument extraction model and use the union of the extracted arguments from both models as the final output. We formulate the argument extraction problem as *conditional text generation*. Our model can easily handle the case of missing arguments and multiple arguments in the same role without the need of tuning thresholds and can extract all arguments in a single pass. 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*. The desired output is a filled template with the arguments.

Our model is based on BART (Lewis et al., 2020), which is an encoder-decoder language model. To utilize the encoder-decoder LM for argument extraction, we construct an input sequence of $\langle$s$\rangle$ template $\langle s\rangle\langle/s\rangle$document $\langle/s\rangle$. All argument names (arg1, arg2 etc.) in the template are replaced by a special placeholder token $\langle$arg$\rangle$. This model is trained in an end-to-end fashion by directly optimizing the generation probability.

To align the extracted arguments back to the document, we adopt a simple postprocessing procedure and find the matching text span closest to the corresponding event trigger.

2.4 Cross-document Cross-lingual Entity and Event Coreference Resolution

After extracting all mentions of entities and events, we apply our cross-document cross-lingual entity coreference resolution model, which is an extension of the e2e-coref model (Lee et al., 2017). We use the multilingual XLM-RoBERTa (XLM-R) Transformer model (Conneau et al., 2020) so that our coreference resolution model can handle non-English data. Second, we port the e2e-coref model to the *cross-document* setting. Given N input documents, we create $\frac{N(N-1)}{2}$ pairs of documents and treat each pair as a single "mega-document". We apply our model to each mega-document and, at the end, aggregate the predictions across all mega-documents to extract the coreference clusters. Finally, we also apply a simple heuristic rule that prevents two entity mentions from being merged together if they are linked to different entities with high confidence.

Our event coreference resolution is similar to entity coreference resolution, while incorporating additional symbolic features such as the event type information. If the input documents are all about one specific complex event, we apply some schema-guided heuristic rules to further refine the predictions of the neural event coreference resolution model. For example, in a bombing schema, there is typically only one bombing event. Therefore, in a document cluster, if there are two event mentions of type *bombing* and they have several arguments in common, these two mentions will be considered as coreferential.

2.5 Cross-document Temporal Event Ordering

Based on the event coreference resolution component described above, we group all mentions into clusters. Next we aim to order events along a timeline. We follow Zhou et al. (2020) to design a component for temporal event ordering. Specifically, we further pre-train a T5 model (Raffel et al., 2020) with distant temporal ordering supervision signals. These signals are acquired through two set of syntactic patterns: 1) before/after keywords in text and 2) explicit date and time mentions. We take such a pre-trained temporal T5 model and fine-tune it on MATRES (Ning et al., 2018b) and use it as the system for temporal event ordering. We perform pair-wise temporal relation classification for all event mention pairs in a documents.

We further train an alternative model from fine-tuning RoBERTa (Liu et al., 2019) on MATRES (Ning et al., 2018b). We only consider event mention pairs which are within neighboring sentences, or can be connected by shared arguments.

Besides model prediction, we also learn high confident patterns from the schema repository. We consider temporal relations that appear very frequently as our prior knowledge. For each given document cluster, we apply these patterns as high-precision patterns before two statistical temporal ordering models separately. The schema matching algorithm will select the best matching from two

[5] `https://www.wikidata.org/`

graphs as the final instantiated schema results.

Because the annotation for non-English data can be expensive and time-consuming, the temporal event tracking component has only been trained on English input. To extend the temporal event tracking capability to cross-lingual setting, we apply Google Cloud neural machine translation [6] to translate Spanish documents into English and apply the FastAlign algorithm (Dyer et al., 2013) to obtain word alignment.

2.6 Cross-media Information Grounding and Fusion

Visual event and argument role extraction: Our goal is to extract visual events along with their argument roles from visual data, i.e., images and videos. In order to train event extractor from visual data, we have collected a new dataset called **Video M2E2** which contains 1,500 video-article pairs by searching over YouTube news channels using 18 event primitives related to visual concepts as search keywords. We have extensively annotated the the videos and sampled key frames for annotating bounding boxes of argument roles.

Our Visual Event and Argument Role Extraction system consists of an event classification model (ResNet-50 (He et al., 2016)) and an argument role extraction model (JSL (Marasović et al., 2020)). To extract the events and associated argument roles, we leverage a public dataset called Situation with Groundings (SWiG) (Marasović et al., 2020) to pretrain our system. SWiG is designed for event and argument understanding in images with object groundings but has a different ontology. We mapped the event types, argument role types and entity names in SWiG to our ontology (covering 12 event sub-types) so that our model is able to extract event information from both images and videos. For videos, we sample frames at a frame rate of 1 frame per second and process them as individual images. In this way, we have a unified model for both image and video inputs.

Multimodal event coreference: We further extended the previous visual event extraction model to find coreference links between visual and text events. For the video frames with detected events, we apply a weakly-supervised grounding model (Akbari et al., 2019) to find sentences and video frames that have high frame-to-sentence similarity, representing the sentence content similar to the video frame content. We apply a rule-based approach to determine if a visual event mention and a textual event mention are coreferential: (1) Their event types match; (2) No contradiction in the entity types for the same argument role across different modalities. (3) The video frame and sentence have a high semantic similarity score. Based on this pipeline, we are able to add visual provenance of events into the event graph. Moreover, we are able to add visual-only arguments to the event graph, which makes the event graph more informative.

2.7 Schema Matching

Once we have acquired a large-scale schema repository by schema induction methods (Li et al., 2020b), we can view it as providing a scaffolding that we can instantiate with incoming data to construct temporal event graphs. Based on each document cluster, we need to find the most accurate schema from the schema repository. We further design a schema matching algorithm that can align our extracted event, entities and relations to a schema.

We first perform topological sort for events based on temporal relations for both IE graph and schema graph so that we can get linearized event sequences in chronological order. Then for each pair of IE graph and schema graph, we apply the longest common subsequence (LCS) method to find the best matching. Our schema matching considers coreference and relations, which will break the optimal substructure when only considering event sequences. We extend the algorithm by replacing the best results for subproblems with a beam of candidates with ranking from a scoring metric that considers matched events, arguments and relations. The candidates consist of matched event pairs, and then we greedily match their arguments and relations for scoring. We merge the best matched IE graph and schema graph to form the final instantiated schema.

3 Experiments

3.1 Data

We have conducted evaluations including schema matching and schema-guided information extraction.

[6] https://cloud.google.com/translate/docs/advanced/translating-text-v3

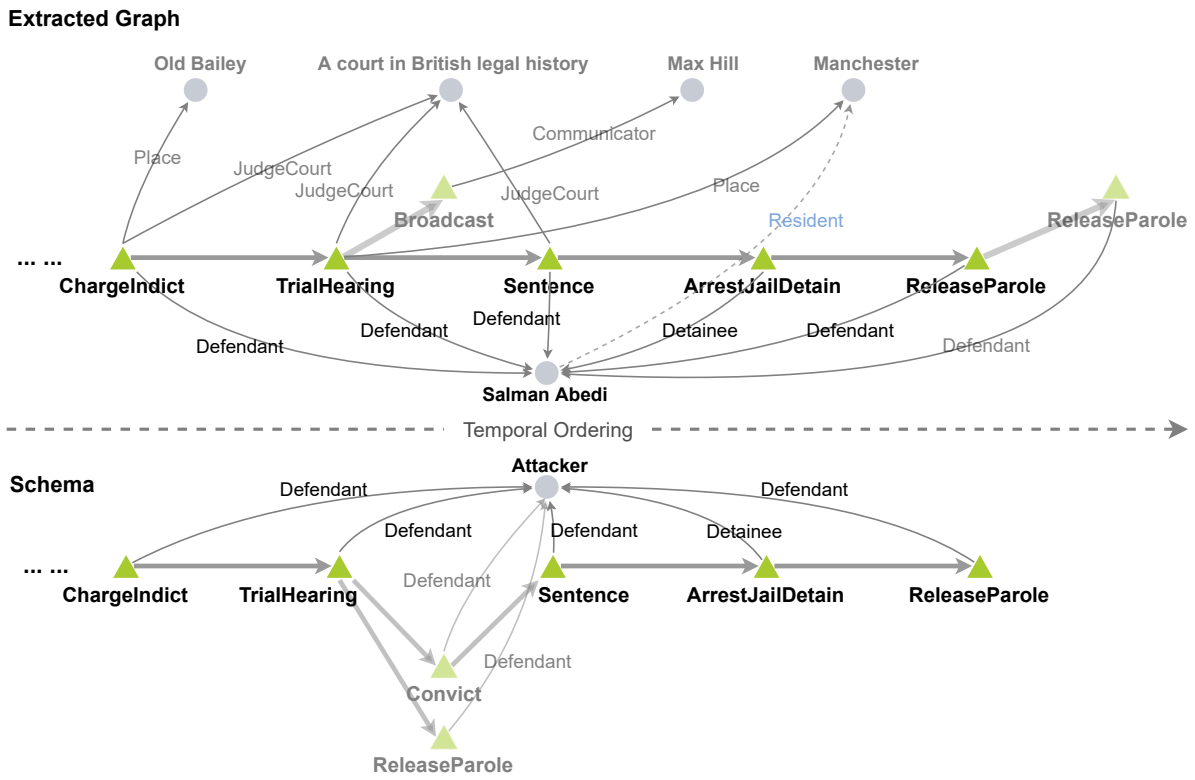

Figure 2: The visualization of schema matching results from extracted graph and schema. The unmatched portions for both extracted graph and schema are blurred.

3.2 Quantitative Performance

Schema Induction. To induce schemas, we collect Wikipedia articles describing complex events related to *improvised explosive device (IED)*, and extract event graphs by applying our IE system. The data statistics are shown in Table 1. We induce schemas by applying the path language model (Li et al., 2020b) over event paths in the training data, and merge top ranked paths into schema graphs for human curation. The statistics of the human curated schema repository are shown in Table 2.

Split	#Docs	#Events	#Arguments	#Relations
Train	5,247	41,672	136,894	122,846
Dev	575	4,661	15,404	13,320
Test	577	5,089	16,721	14,054

Table 1: Data statistics of IED Schema Learning Corpus.

Schema-guided Information Extraction. The performance of each component is shown in Table 3. We evaluate the end-to-end performance of our system on a complex event corpus (LDC2020E39), which contains multi-lingual multi-media document clusters. The data statistics are shown in Table 4. We train our mention extraction component on ACE 2005 (Walker et al., 2006) and ERE (Song et al., 2015); document-level argument exraction on ACE 2005 (Walker et al., 2006) and RAMS (Ebner et al., 2020); coreference component on ACE 2005 (Walker et al., 2006), EDL 2016[7], EDL 2017[8], OntoNotes (Pradhan et al., 2012), ERE (Song et al., 2015), CoNLL 2002 (Tjong Kim Sang, 2002), DCEP (Dias, 2016) and SemEval 2010 (Recasens et al., 2010); temporal ordering component on MATRES (Ning et al., 2018b); visual event and argument extraction on

Schema	#Steps	#Arguments	#Temporal Links
Disease Outbreak	20	94	20
Disaster Relief	15	85	15
Medical Treatment	8	37	8
Search and Rescue	11	50	10
General Attack	21	89	27
General IED	33	144	43
Roadside IED	28	123	36
Car IED	34	148	45
Drone Strikes IED	32	142	48
Backpack IED	31	138	40

Table 2: Data statistics of the induced schema library.

[7]LDC2017E03
[8]LDC2017E52

Video M2E2 and SWiG (Marasović et al., 2020). The statistics of our output are shown in Table 5.

Component			Benchmark	Metric	Score
Mention Extraction	En	Trigger	ACE+ERE	F_1	64.1
		Argument	ACE+ERE	F_1	49.7
		Relation	ACE+ERE	F_1	49.5
	Es	Trigger	ACE+ERE	F_1	63.4
		Argument	ACE+ERE	F_1	46.0
		Relation	ACE+ERE	F_1	46.6
Document-level Argument Extraction			ACE	F_1	66.7
			RAMS	F_1	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
Temporal Ordering		RoBERTa	MATRES	F1	78.8
		T5	MATRES-b	Acc.	89.6
Visual Event Extraction			Video M2E2	Acc.	70.0

Table 3: Performance (%) of each component. MATRES-b refers to MATRES binary classification that only considers BEFORE and AFTER relations.

Category	Complex Events	Documents	Images	Videos
#	11	139	1,213	31

Table 4: Data statistics for schema matching corpus (LDC2020E39).

Category	Extracted Events	Schema Steps	Instantiated Steps
#	3,180	1,738	958

Table 5: Results of schema matching.

3.3 Qualitative Analysis

Figure 2 illustrates a subset of examples for the best matched results from our end-to-end system. We can see that our system can extract events, entities and relations and align them well with the selected schema. The final instantiated schema is the hybrid of two graphs from merging the matched elements.

4 Related Work

Text Information Extraction. Existing end-to-end Information Extraction (IE) systems (Wadden et al., 2019; Li et al., 2020a; Lin et al., 2020; Li et al., 2019) mainly focus on extracting entities, events and entity relations from individual sentences. In contrast, we extract and infer arguments over the global document context. Furthermore, our IE system is guided by a schema repository. The extracted graph will be used to instantiate a schema graph, which can be applied to predict future events.

Multimedia Information Extraction. Previous multimedia IE systems (Li et al., 2020a; Yazici et al., 2018) only include cross-media entity coreference resolution by grounding the extracted visual entities to text. We are the first to perform cross-media joint event extraction and coreference resolution to obtain the coreferential events from text, images and videos.

Coreference Resolution. Previous neural models for event coreference resolution use non-contextual (Nguyen et al., 2016; Choubey et al., 2020; Huang et al., 2019) or contextual word representations (Lu et al., 2020; Yu et al., 2020). We incorporate a wide range of symbolic features (Chen and Ji, 2009; Chen et al., 2009; Sammons et al., 2015; Lu and Ng, 2016, 2017; Duncan et al., 2017), such as event attributes and types, into our event coreference resolution module using a context-dependent gate mechanism.

Temporal Event Ordering. Temporal relations between events are extracted for neighbor events in one sentence (Ning et al., 2017, 2018a, 2019; Han et al., 2019), ignoring the temporal dependencies between events across sentences. We perform document-level event ordering and propagate temporal attributes through shared arguments. Furthermore, we take advantage of the schema repository knowledge by using the frequent temporal order between event types to guide the ordering between events.

5 Conclusions and Future Work

We demonstrate a state-of-the-art schema-guided cross-document cross-lingual cross-media information extraction and event tracking system. This system is made publicly available to enable users to effectively harness rich information from a variety of sources, languages and modalities. In the future, we plan to develop more advanced graph neural networks based method for schema matching and schema-guided event prediction.

6 Broad Impact

Our goal in developing Cross-document Cross-lingual Cross-media information extraction and event tracking systems is to advance the state of the art and enhance the field's ability to fully understand real-world events from multiple sources, languages and modalities. We believe that to make real progress in event-centric Natural Language Understanding, we should not focus only on datasets, but to also ground our work in real-world applications. The application we focus on is navigating news, and the examples shown here and in the paper demonstrate the potential use in news understanding.

For our demo, the distinction between beneficial use and harmful use depends, in part, on the data. Proper use of the technology requires that input documents/images are legally and ethically obtained. We are particularly excited about the potential use of the technologies in applications of broad societal impact, such as disaster monitoring and emergency response. Training and assessment data is often biased in ways that limit system accuracy on less well represented populations and in new domains. The performance of our system components as reported in the experiment section is based on the specific benchmark datasets, which could be affected by such data biases. Thus questions concerning generalizability and fairness should be carefully considered.

A general approach to ensure proper, rather than malicious, application of dual-use technology should: incorporate ethics considerations as the first-order principles in every step of the system design, maintain a high degree of transparency and interpretability of data, algorithms, models, and functionality throughout the system. We intend to make our software available as open source and shared docker containers for public verification and auditing, and explore countermeasures to protect vulnerable groups.

References

Hassan Akbari, Svebor Karaman, Surabhi Bhargava, Brian Chen, Carl Vondrick, and Shih-Fu Chang. 2019. Multi-level multimodal common semantic space for image-phrase grounding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 12476–12486.

Zheng Chen and Heng Ji. 2009. Graph-based event coreference resolution. In *Proceedings of the 2009 Workshop on Graph-based Methods for Natural Language Processing (TextGraphs-4)*, pages 54–57.

Zheng Chen, Heng Ji, and Robert M Haralick. 2009. A pairwise event coreference model, feature impact and evaluation for event coreference resolution. In *Proceedings of the workshop on events in emerging text types*, pages 17–22.

Prafulla Kumar Choubey, Aaron Lee, Ruihong Huang, and Lu Wang. 2020. Discourse as a function of event: Profiling discourse structure in news articles around the main event. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5374–5386, Online. Association for Computational Linguistics.

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

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

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

Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2014. Using structured events to predict stock price movement: An empirical investigation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1415–1425, Doha, Qatar. Association for Computational Linguistics.

Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2016. Knowledge-driven event embedding for stock prediction. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2133–2142, Osaka, Japan. The COLING 2016 Organizing Committee.

Chase Duncan, Liang-Wei Chan, Haoruo Peng, Hao Wu, Shyam Upadhyay, Nitish Gupta, Chen-Tse Tsai, Mark Sammons, and Dan Roth. 2017. Ui ccg tackbp2017 submissions: Entity discovery and linking, and event nugget detection and co-reference. In *TAC*.

Chris Dyer, Victor Chahuneau, and Noah A Smith. 2013. A simple, fast, and effective reparameterization of ibm model 2. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 644–648.

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*.

Liat Ein-Dor, Ariel Gera, Orith Toledo-Ronen, Alon Halfon, Benjamin Sznajder, Lena Dankin, Yonatan Bilu, Yoav Katz, and Noam Slonim. 2019. Financial event extraction using Wikipedia-based weak supervision. In *Proceedings of the Second Workshop on Economics and Natural Language Processing*, pages 10–15, Hong Kong. Association for Computational Linguistics.

Goran Glavaš, Jan Šnajder, Marie-Francine Moens, and Parisa Kordjamshidi. 2014. HiEve: A corpus for extracting event hierarchies from news stories. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3678–3683, Reykjavik, Iceland. European Language Resources Association (ELRA).

Goran Glavaš and Sanja Štajner. 2013. Event-centered simplification of news stories. In *Proceedings of the Student Research Workshop associated with RANLP 2013*, pages 71–78, Hissar, Bulgaria. INCOMA Ltd. Shoumen, BULGARIA.

Rujun Han, Qiang Ning, and Nanyun Peng. 2019. Joint event and temporal relation extraction with shared representations and structured prediction. 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 2019, Hong Kong, China, November 3-7, 2019*, pages 434–444. Association for Computational Linguistics.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

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

Yin Jou Huang, Jing Lu, Sadao Kurohashi, and Vincent Ng. 2019. Improving event coreference resolution by learning argument compatibility from unlabeled data. 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 785–795.

Gilles Jacobs, Els Lefever, and Véronique Hoste. 2018. Economic event detection in company-specific news text. In *Proceedings of the First Workshop on Economics and Natural Language Processing*, pages 1–10, Melbourne, Australia. Association for Computational Linguistics.

Abhyuday N Jagannatha and Hong Yu. 2016. Bidirectional RNN for medical event detection in electronic health records. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 473–482, San Diego, California. Association for Computational Linguistics.

Serena Jeblee and Graeme Hirst. 2018. Listwise temporal ordering of events in clinical notes. In *Proceedings of the Ninth International Workshop on Health Text Mining and Information Analysis*, pages 177–182, Brussels, Belgium. Association for Computational Linguistics.

Prescott Klassen, Fei Xia, and Meliha Yetisgen. 2016. Annotating and detecting medical events in clinical notes. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3417–3421, Portorož, Slovenia. European Language Resources Association (ELRA).

Kenton Lee, Luheng He, Mike Lewis, and Luke Zettlemoyer. 2017. End-to-end neural coreference resolution. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 188–197, Copenhagen, Denmark. Association for Computational Linguistics.

M. Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, A. Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2020. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *ArXiv*, abs/1910.13461.

Manling Li, Ying Lin, Joseph Hoover, Spencer Whitehead, Clare Voss, Morteza Dehghani, and Heng Ji. 2019. Multilingual entity, relation, event and human value extraction. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pages 110–115.

Manling Li, Alireza Zareian, Ying Lin, Xiaoman Pan, Spencer Whitehead, Brian Chen, Bo Wu, Heng Ji, Shih-Fu Chang, Clare Voss, Daniel Napierski, and Marjorie Freedman. 2020a. GAIA: A fine-grained multimedia knowledge extraction system. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 77–86, 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 *Proc. The 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP2020)*.

Ying Lin, Heng Ji, Fei Huang, and Lingfei Wu. 2020. A joint end-to-end neural model for information ex-

traction with global features. In *Proc. The 58th Annual Meeting of the Association for Computational Linguistics (ACL2020)*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Jing Lu and Vincent Ng. 2016. Event coreference resolution with multi-pass sieves. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3996–4003.

Jing Lu and Vincent Ng. 2017. Joint learning for event coreference resolution. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 90–101.

Yaojie Lu, Hongyu Lin, Jialong Tang, Xianpei Han, and Le Sun. 2020. End-to-end neural event coreference resolution. *arXiv preprint arXiv:2009.08153*.

Ana Marasović, Chandra Bhagavatula, Jae Sung Park, Ronan Le Bras, Noah A Smith, and Yejin Choi. 2020. Natural language rationales with full-stack visual reasoning: From pixels to semantic frames to commonsense graphs. *arXiv preprint arXiv:2010.07526*.

Salvador Medina Maza, Evangelia Spiliopoulou, Eduard Hovy, and Alexander Hauptmann. 2020. Event-related bias removal for real-time disaster events. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 3858–3868, Online. Association for Computational Linguistics.

Thien Huu Nguyen, Adam Meyers, and Ralph Grishman. 2016. New york university 2016 system for kbp event nugget: A deep learning approach. In *TAC*.

Qiang Ning, Zhili Feng, and Dan Roth. 2017. A structured learning approach to temporal relation extraction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1027–1037, Copenhagen, Denmark. Association for Computational Linguistics.

Qiang Ning, Zhili Feng, Hao Wu, and Dan Roth. 2018a. Joint reasoning for temporal and causal relations. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pages 2278–2288. Association for Computational Linguistics.

Qiang Ning, Sanjay Subramanian, and Dan Roth. 2019. An improved neural baseline for temporal relation 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 2019, Hong Kong, China, November 3-7, 2019*, pages 6202–6208. Association for Computational Linguistics.

Qiang Ning, H. Wu, and D. Roth. 2018b. A multi-axis annotation scheme for event temporal relations. *ArXiv*, abs/1804.07828.

Ahmet Murat Özbayoglu, Mehmet Ugur Gudelek, and Omer Berat Sezer. 2020. Deep learning for financial applications : A survey. *Appl. Soft Comput.*, 93:106384.

Xiaoman Pan, Boliang Zhang, Jonathan May, Joel Nothman, Kevin Knight, and Heng Ji. 2017. Cross-lingual name tagging and linking for 282 languages. In *Proc. the 55th Annual Meeting of the Association for Computational Linguistics (ACL2017)*.

Sandeep Panem, Manish Gupta, and Vasudeva Varma. 2014. Structured information extraction from natural disaster events on twitter. In *Proceedings of the 5th International Workshop on Web-scale Knowledge Representation Retrieval & Reasoning, Web-KR@CIKM 2014, Shanghai, China, November 3, 2014*, pages 1–8. ACM.

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.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, M. Matena, Yanqi Zhou, W. 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.

Preethi Raghavan, Eric Fosler-Lussier, and Albert Lai. 2012. Temporal classification of medical events. In *BioNLP: Proceedings of the 2012 Workshop on Biomedical Natural Language Processing*, pages 29–37, Montréal, Canada. Association for Computational Linguistics.

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.

Mark Sammons, Haoruo Peng, Yangqiu Song, Shyam Upadhyay, Chen-Tse Tsai, Pavankumar Reddy, Subhro Roy, and Dan Roth. 2015. Illinois ccg tac 2015 event nugget, entity discovery and linking, and slot filler validation systems. In *TAC*.

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.

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)*.

Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledge base. *Communications of the ACM*, 57(10).

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-IJCNLP2019)*.

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

Hang Yang, Yubo Chen, Kang Liu, Yang Xiao, and Jun Zhao. 2018. DCFEE: A document-level Chinese financial event extraction system based on automatically labeled training data. In *Proceedings of ACL 2018, System Demonstrations*, pages 50–55, Melbourne, Australia. Association for Computational Linguistics.

Adnan Yazici, Murat Koyuncu, Turgay Yilmaz, Saeid Sattari, Mustafa Sert, and Elvan Gulen. 2018. An intelligent multimedia information system for multimodal content extraction and querying. *Multimedia Tools and Applications*, 77(2):2225–2260.

Xiaodong Yu, Wenpeng Yin, and Dan Roth. 2020. Paired representation learning for event and entity coreference. *arXiv preprint arXiv:2010.12808*.

Boliang Zhang, Ying Lin, Xiaoman Pan, Di Lu, Jonathan May, Kevin Knight, and Heng Ji. 2018. Elisa-edl: A cross-lingual entity extraction, linking and localization system. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pages 41–45.

Ben Zhou, Kyle Richardson, Qiang Ning, Tushar Khot, A. Sabharwal, and D. Roth. 2020. Temporal reasoning on implicit events from distant supervision. *ArXiv*, abs/2010.12753.