

A Scalable Cross-Domain Event Extraction System via a Unified Generative Training Framework

Siting Liang^{1,2}, Omar Adjali¹, Omair Shahzad Bhatti¹ and Daniel Sonntag^{1,2}

¹German Research Center for Artificial Intelligence

²Carl von Ossietzky University of Oldenburg

{siting.liang, omar.adjali, omair_shahzad.bhatti, daniel.sonntag}@dfki.de

Abstract

Event extraction is fundamental to information extraction. Prior approaches often separate event detection and argument extraction or depend on dataset-specific designs, limiting scalability and cross-domain generalization. We propose a unified generative sequence-to-sequence framework that performs event extraction subtasks jointly and supports both pipeline and end-to-end configurations. We fine-tune pretrained language models on multiple event datasets across diverse domains, enabling a single model to retain domain-specific semantics while generalizing over large and evolving label spaces. We demonstrate these capabilities through a web-based application tailored for researchers and practitioners. The platform supports document upload, schema-aware event extraction, visualization of triggers and arguments, and comparison of different extraction configurations across domains.

1 Introduction

Event extraction (EE) is a fundamental information extraction task that converts unstructured text into structured event representations, which serve as a critical foundation for knowledge base construction [Yang *et al.*, 2019]. It is typically formulated in two stages: Event Detection (ED), which identifies triggers and classifies their types, and Event Argument Extraction (EAE), which identifies arguments and assigns them semantic roles, as illustrated in Figure 1. Many prior methods design dedicated models or learning objectives for individual subtasks, often organized in pipelined architectures [Zhang and Ji, 2021; Hsu *et al.*, 2023a]. These approaches are often tied to specific annotation schemas and domains, making them difficult to scale or generalize across datasets with different event definitions and label spaces and requiring extensive re-engineering and retraining for new domains. Related cross-domain information extraction studies have shown that pre-trained language models and unified semantic abstractions can support transfer across heterogeneous domain-specific label spaces [Liang *et al.*, 2023].

More recently, unified information extraction frameworks and large language models have explored general-purpose EE

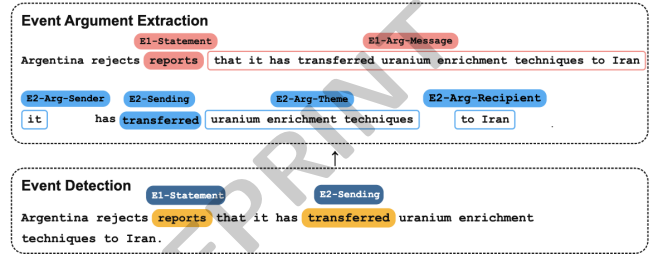


Figure 1: (1) Event Detection identifies event triggers and assigns event types, followed by (2) Event Argument Extraction, which extracts arguments and labels them with semantic roles.

by modeling the task in an end-to-end (E2E), generative manner [Li *et al.*, 2021; Hsu *et al.*, 2022; Hsu *et al.*, 2023b; Wang *et al.*, 2023; Adjali *et al.*, 2026]. However, these methods often use prompts to encode entire event schemas and are typically evaluated on datasets with relatively small label sets [Sainz *et al.*, 2024; Gui *et al.*, 2024]. Furthermore, their high computational requirements make scaling to large or complex event taxonomies impractical. Efforts have also sought to address scalability by constructing large ontology datasets [Li *et al.*, 2023; Ebner *et al.*, 2020]. These datasets primarily rely on weak supervision signals without domain-specific adaptation, which limits their ability to capture domain-specific context. This underscores the need for a unified EE system that also operates robustly across domains, datasets, and large-scale annotation schemas with minimal re-design demands.

In this work, we present an event extraction system based on a cross-domain unified generative framework using T5 architecture [Raffel *et al.*, 2020]. The framework formulates event extraction as a sequence-to-sequence task, enabling a single model to handle ED and EAE in a unified manner. Furthermore, the framework flexibly supports both pipeline and E2E modes. Based on this framework, we train a cross-domain backbone model on six datasets spanning multiple domains jointly, enabling the model to learn representations that generalize across domains while retaining domain-specific semantics. We present a unified, cross-domain EE system implemented as a web-based application designed for practitioners to demonstrate the effectiveness and efficiency of the model trained jointly across diverse EE datasets.

Unified Framework and Multi-Datasets Training

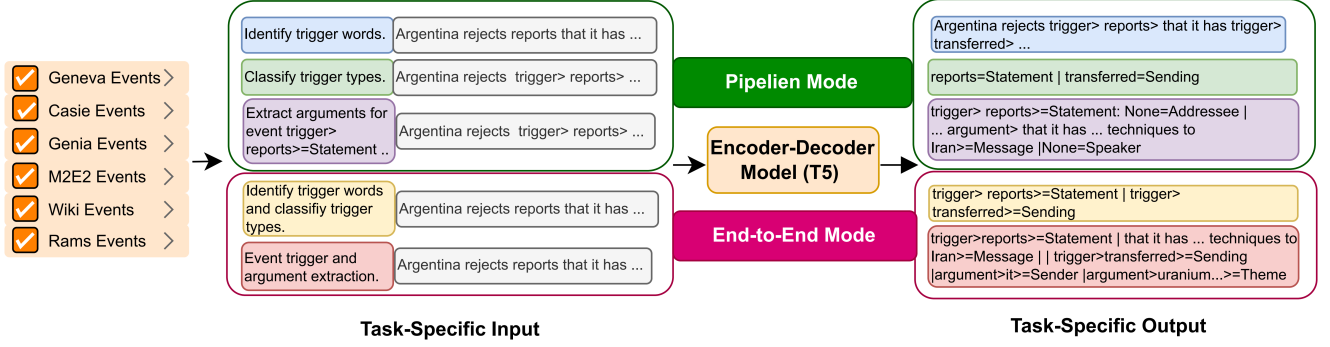


Figure 2: We train a unified sequence-to-sequence model based on T5 across six diverse domains, using all available datasets. Each dataset is mapped to a task-specific prompt and a corresponding output pattern to standardize event extraction. This formulation allows a single model to learn cross-domain event representations and supports both pipeline and E2E modes.

2 Technology and Innovation

We present a cross-domain unified generative framework for event extraction, designed to balance scalability, flexibility, and domain adaptation. The training framework is displayed in Figure 2 and the main technical contributions of our system are summarized as follows.

Unified generative modeling of event extraction tasks. Our framework formulates all event extraction tasks as a unified sequence-to-sequence task, handling trigger identification, event type classification and argument extraction simultaneously. This approach makes the system flexible, as it defines specific task codes and can handle changes to the input text without requiring modifications to the core architecture. Unifying modeling reduces the need for separate models, allowing consistent training across datasets while maintaining competitive performance in low-resource scenarios.

Supporting both pipeline and E2E modes. Our generative framework supports E2E and pipeline configurations in inference time, see Figure 3. Switching between modes requires using distinct input and target patterns for each mode for the same tasks as displayed in Figure 2. The pipeline design provides explicit control over each extraction stage, supporting interpretability and stage-specific evaluation in inference time. In pipeline mode, the model detects event triggers in the input text. Each detected trigger is then classified to predict its event type based on the selected dataset schema, and finally, argument extraction is performed based on the predicted event type and its associated argument roles from the selected dataset schema. By contrast, E2E jointly performs trigger identification and event type classification in a single prediction step, simplifying inference and reducing error propagation, albeit with less fine-grained control than the pipeline workflow. A fully E2E formulation that incorporates argument extraction is also feasible, though more challenging due to the increased complexity of the output space.

Handling cross-domain and evolving label spaces. To further enhance the generative framework, we extend the task-only conditioning paradigm by adding both a task

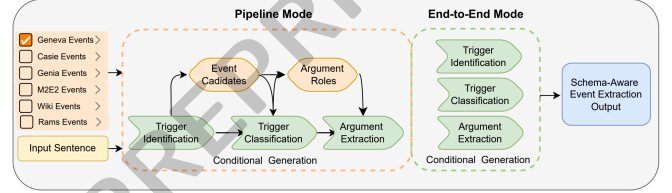


Figure 3: Pipeline and E2E event extraction workflows in inference time.

identifier and a domain identifier to the beginning of the input sequence. This yields the unified input format: $\{dataset_name\} + \{task_prefix\} + \{text\}$. The dataset name serves as a domain indicator and schema guide. This design allows the model to be trained jointly on multiple event extraction datasets, learning patterns that generalize across domains and accommodating diverse datasets and evolving schemas with minimal additional engineering while preserving domain-specific semantics. The six datasets used in our system follow the preprocessing of Huang [2024] and are summarized in Table 1. After fine-tuning, the cross-domain model can handle a wide range of event types simply by specifying the corresponding domain name in the input sequence, enabling the model predicting to the relevant domain.

3 Application Features

We present a web-based event extraction application based on the cross-domain unified generative framework. It allows JSON uploads and document selection with previews. Currently, it supports six event schemas (refer to Table 1) and provides domain-specific models trained on a single dataset and optimized for their respective domains, as well as a cross-domain model trained on all available datasets to generalize across domains. All models are fine-tuned on T5-base pre-trained model with approximately 220 million parameters. The application’s main features are as follows.

Two-stage Event Extraction. Our event extraction system uses a two-stage workflow as displayed in Figure 4. First, se-

lect the dataset (event schema) to define labels and roles, e.g., *geneva*. The model performs event detection in two modes: pipeline and E2E. In pipeline mode, trigger spans are first identified and then classified into event types. In E2E mode, both steps are performed jointly in one pass. The outputs of the two modes are then merged into a consolidated trigger set in the first stage. Stage 2 then extracts arguments based on the event detection results of each mode obtained in the first stage. Model selection is configurable in both stages. This flexibility is motivated by practical observations that domain-specific models often achieve stronger trigger identification performance in high-resource domains, while cross-domain models tend to generalize better in low-resource settings, helping to mitigate error propagation in pipeline-based extraction.

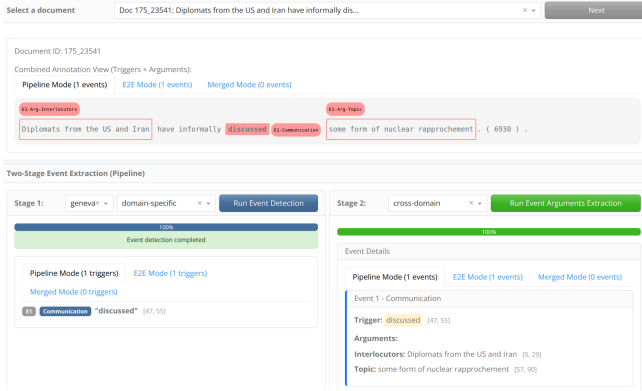


Figure 4: Two-stage event extraction interface. Stage 1 performs event detection using pipeline and E2E modes and merges the detected triggers. Stage 2 extracts arguments conditioned on the detected events and event-type-specific argument roles. Text highlighting supports result inspection.

Results visualization comparison for all configurations. In two-stage extraction mode, the visualization design is central: precise offsets, aligned highlights, and event cards make cross-mode differences explicit, so users can scrutinize Stage 1 outputs, decide which detection mode to trust, and see how that choice propagates into Stage 2 arguments. Beyond the two-stage workflow, the system supports full E2E extraction

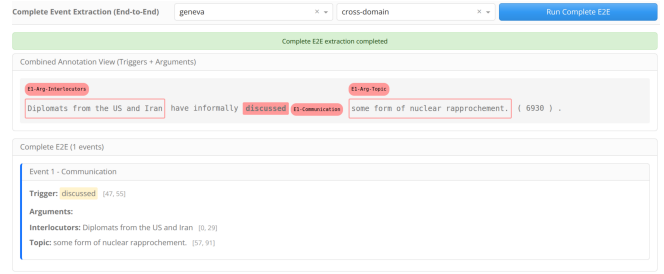


Figure 5: View of complete E2E event extraction, where the model runs all event extraction subtasks in one step.

in a single step for efficient large-scale processing, and the interface presents all modes (5) in a consistent, comparable layout. Results can be examined per document and saved for side-by-side comparison, enabling informed configuration selection for each document.

Schema-aware multi-dataset annotation. The backend cross-domain fine-tuned model allows applying multiple event schemas to the same document while keeping each dataset’s labels and structure separate, see Figure 6. All event extraction results are saved with schema-aware labels that preserve dataset provenance across stage 1/2 and complete E2E outputs. The interface enables repeated annotation of the same document under different event schemas, producing a hierarchical, reusable JSON artifact suitable for downstream analysis and comparison.

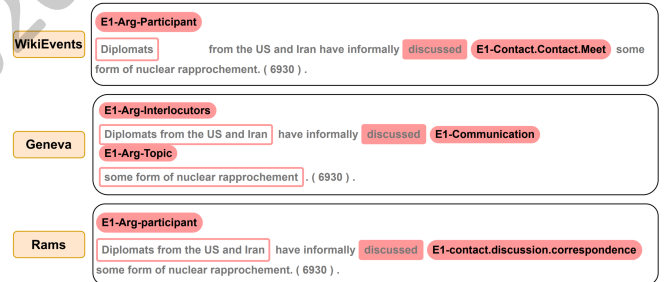


Figure 6: Event extraction on the same document with different dataset schemas.

Dataset	#Types	Core Argument Roles
CASIE (Cyber)	5	Attacker, Target, Malware, Vulnerability, Compromised Data, Instrument
GENEVA (Generic)	115	Protester, Target, Place, Time, Reason, Organizer
GENIA 2013 (Bio)	13	Theme, Agent, Site, Resulting Entity
M2E2 (Geo)	8	Location, Victims, Casualties, Time
RAMS (Newswire)	139	Participant, Location, Time, Topic, Outcome
WikiEvents (Wiki)	50	Entity, Organization, Position, Place, Time, Reason

Table 1: Representative event types and their argument roles.

4 Conclusion

In this work, we propose a unified generative approach for cross-domain event extraction. We demonstrate the system across multiple datasets spanning diverse domains and show robust performance across datasets, extraction configurations, and ontology sizes. We release an open-source application and video at: Youtube and GitHub that deliver a complete workflow from document ingestion to structured event outputs, supports both pipeline and E2E extraction, and exposes domain-specific and cross-domain models. The design prioritizes practical deployment, flexible adaptation to large label spaces, and reproducible analysis without extensive human redesign or dataset-specific customization, making it a scalable solution for real-world event extraction applications.

Ethical statement

During the preparation of this work, we used GitHub Copilot to assist with the development of the code. After using this AI tool, we carefully reviewed and validated all the generated code. We take full responsibility for the accuracy, integrity and final content of this publication.

References

- [Adjali *et al.*, 2026] Omar Adjali, Siting Liang, Omair Shahzad Bhatti, and Daniel Sonntag. Aligning instruction-tuned llms for event extraction with multi-objective reinforcement learning. In *Advances in Information Retrieval*, pages 586–595, Cham, 2026. Springer Nature Switzerland.
- [Ebner *et al.*, 2020] Seth Ebner, Patrick Xia, Ryan Culkin, Kyle Rawlins, and Benjamin Van Durme. Multi-sentence argument linking. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8057–8077, 2020.
- [Gui *et al.*, 2024] Honghao Gui, Lin Yuan, Hongbin Ye, Ningyu Zhang, Mengshu Sun, Lei Liang, and Huajun Chen. Iepile: Unearthing large scale schema-conditioned information extraction corpus. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 127–146, 2024.
- [Hsu *et al.*, 2022] I-Hung Hsu, Kuan-Hao Huang, Elizabeth Boschee, Scott Miller, Prem Natarajan, Kai-Wei Chang, and Nanyun Peng. DEGREE: A data-efficient generation-based event extraction model. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, 2022.
- [Hsu *et al.*, 2023a] I-Hung Hsu, Kuan-Hao Huang, Shuning Zhang, Wenxin Cheng, Prem Natarajan, Kai-Wei Chang, and Nanyun Peng. TAGPRIME: A unified framework for relational structure extraction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*, 2023.
- [Hsu *et al.*, 2023b] I-Hung Hsu, Zhiyu Xie, Kuan-Hao Huang, Prem Natarajan, and Nanyun Peng. AMPERE: amr-aware prefix for generation-based event argument extraction model. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*, 2023.
- [Huang *et al.*, 2024] Kuan-Hao Huang, I-Hung Hsu, Tanmay Parekh, Zhiyu Xie, Zixuan Zhang, Prem Natarajan, Kai-Wei Chang, Nanyun Peng, and Heng Ji. Textee: Benchmark, reevaluation, reflections, and future challenges in event extraction. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 12804–12825, 2024.
- [Li *et al.*, 2021] Sha Li, Heng Ji, and Jiawei Han. Document-level event argument extraction by conditional generation. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 2021.
- [Li *et al.*, 2023] Sha Li, Qiusi Zhan, Kathryn Conger, Martha Palmer, Heng Ji, and Jiawei Han. GLEN: general-purpose event detection for thousands of types. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023.
- [Liang *et al.*, 2023] Siting Liang, Mareike Hartmann, and Daniel Sonntag. Cross-domain German medical named entity recognition using a pre-trained language model and unified medical semantic types. In Tristan Naumann, Asma Ben Abacha, Steven Bethard, Kirk Roberts, and Anna Rumshisky, editors, *Proceedings of the 5th Clinical Natural Language Processing Workshop*, pages 259–271, Toronto, Canada, July 2023. Association for Computational Linguistics.
- [Raffel *et al.*, 2020] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67, 2020.
- [Sainz *et al.*, 2024] Oscar Sainz, Iker García-Ferrero, Rodrigo Agerri, Oier Lacalle, German Rigau, and Eneko Agirre. Gollie: Annotation guidelines improve zero-shot information-extraction. In *International Conference on Learning Representations*, volume 2024, pages 47083–47107, 2024.
- [Wang *et al.*, 2023] Zhenhailong Wang, Can Xu, Shizhu He, Kang Liu, and Jun Zhao. Instructuie: Multi-task instruction tuning for unified information extraction. In *Proceedings of ACL*, 2023.
- [Yang *et al.*, 2019] Sen Yang, Dawei Feng, Linbo Qiao, Zhigang Kan, and Dongsheng Li. Exploring pre-trained language models for event extraction and generation. In Anna Korhonen, David Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5284–5294, Florence, Italy, July 2019. Association for Computational Linguistics.
- [Zhang and Ji, 2021] Zixuan Zhang and Heng Ji. Abstract meaning representation guided graph encoding and decoding for joint information extraction. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, (NAACL-HLT)*, 2021.